Feature-specific imaging

Mark A. Neifeld and Premchandra Shankar

We analyze the performance of feature-specific imaging systems. We study incoherent optical systems that directly measure linear projects of the optical irradiance distribution. Direct feature measurement exploits the multiplex advantage, and for small numbers of projections can provide higher feature-fidelity than those systems that postprocess a conventional image. We examine feature-specific imaging using Wavelet, Karhunen–Loeve (KL), Hadamard, and independent-component features, quantifying feature fidelity in Gaussian-, shot-, and quantization-noise environments. An example of feature-specific imaging based on KL projections is analyzed and demonstrates that within a high-noise environment it is possible to improve image fidelity via direct feature measurement. A candidate optical system is presented and a preliminary implementational study is undertaken.

OCIS codes: 110.0110, 100.3010, 100.5010, 110.6980, 200.4740.

1. Introduction

Imaging is an important component of many applications in both the government and commercial sectors. Defense-related imaging applications are concerned with a broad range of tasks from navigation to threat assessment and targeting; while commercial applications include such diverse tasks as parts inspection, medical imaging, and security/surveillance. It is important to realize that nearly all of these important applications include, as part of the overall system goal, some quantitative diagnostic process. As an example we can consider an astronomical imaging system for which the goal might be to answer the question: Does the imager field-of-view contain any previously unknown celestial objects and if so, what are their characteristics? Another example can be found within a target recognition environment for which the overall system goal may be to label the objects in a scene as either target or nontarget objects. These two examples serve to illustrate a common disconnect between imaging system design and the diagnostic task for which the image measurements will be used. Namely, nearly every imaging system attempts to form, either directly or as an intermediate result, a visually appealing representation of the scene; however, without any additional specifications, it is not clear that such a representation is necessary to achieve the goal. It is not uncommon therefore to find imager hardware performance quantified by using a metric (e.g., visual quality) that bears little relevance to the eventual use of the measured data.

The above discussion suggests a need for task-specific imaging system design strategies. Task-specific image quality metrics along with their associated hardware/software co-design procedures would facilitate the manipulation of both optical and algorithmic degrees of freedom toward the optimal solution to a specific imaging system goal. These observations have been made by several authors before us and have given rise to important nontraditional imaging systems and system design metrics.[1] Task-specific ideal-observer performance metrics have had an important impact on the techniques used to manipulate various forms of medical imagery, while information theoretic metrics have been applied to optical-imaging systems for applications ranging from storage to astronomy.[2–4] The joint design of imager hardware and software has been demonstrated through the use of cubic phase masks for extended depth of field, aberration correction, etc.[5] Recent demonstrations of novel tomographic, holographic, and interferometric imaging systems provide additional platforms upon which these novel solutions might be tested.[6–8]

In this paper we consider a form of task-specific imaging for which overall system performance is determined by the fidelity of linear object features. The next section presents our analysis of such a

The authors are with the Department of Electrical and Computer Engineering, Optical Sciences Center, University of Arizona, Tucson, Arizona 85721.

Received 30 October 2002; revised manuscript received 21 February 2003.

0003-6935/03/173379-11$15.00/0

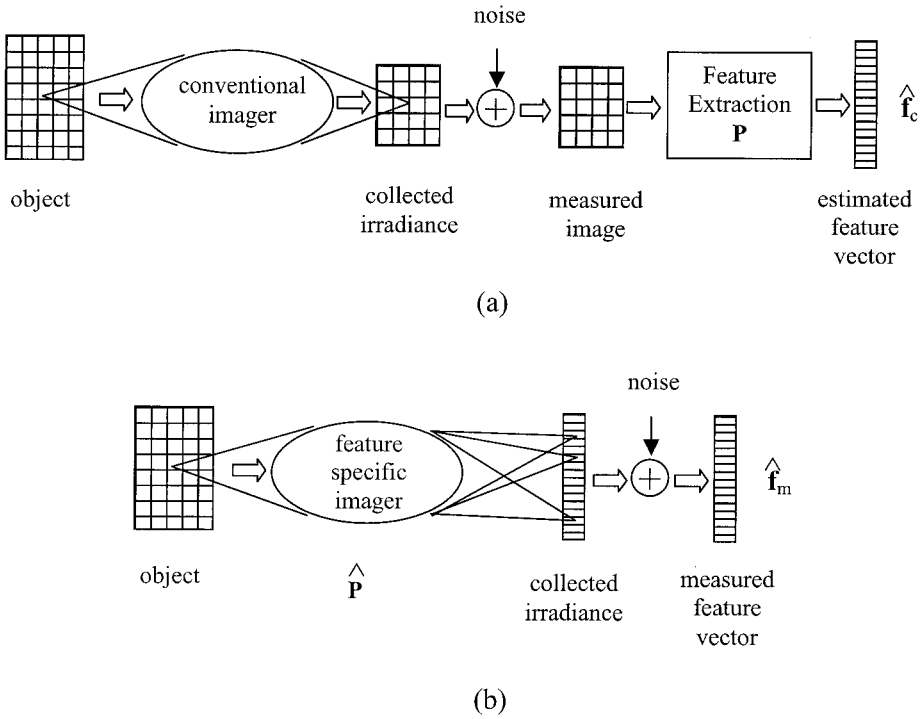


Fig. 1. Schematic diagram of (a) a conventional imager, (b) a feature-specific imager.

feature-specific imaging system, and we elucidate the important role that photon count plays in defining the comparison with conventional imaging. This analysis demonstrates that the direct measurement of linear features can be fundamentally superior to their extraction (via postprocessing) from a conventional image. The novelty of our approach resides in our attention to feature fidelity subject to a fixed photon budget, and results in the conclusion that fidelity can be improved via direct feature measurement. Section 3 quantifies the feature fidelity obtained from a feature-specific imager for various types of noise, while Section 4 presents a design example of feature-specific imaging for the case of Karhunen–Loeve (KL) features and demonstrates some important tradeoffs in the context of hardware/software co-design. Section 5 presents a candidate optical architecture for a photon-efficient feature-specific imager along with a sensitivity analysis of this candidate implementation.

2. Feature-Specific Imaging Framework

In this section we will present a framework for the design and analysis of task-specific imagers. Our approach will be suitable for those problems that can be described in terms of linear object features. This defines a very broad class of tasks, each of which will have a preferred set of features. For example, if the imager is required to transmit or store its imagery, then features might be selected to optimize compression efficiency. We will consider wavelet features in this regard.[9] If a coding/decoding problem underlies the deployment of the imager, then Hadamard features may be required.[10] Alternately, if image mean-squared error (MSE) is the relevant task-specific metric, then KL features (i.e., principal components) may be best, while for the extraction of maximum image information the so-called independent component features are optimal.[11,12] Although we will describe feature-specific imaging in a general way, examples of feature-specific imaging will be limited to the four types of features mentioned above.

Figure 1 presents schematic representations of the two optical systems that we will compare. We will limit our study to incoherent optical systems, and we will seek to measure linear projections of the object irradiance distribution. Objects are assumed to be spatially discrete without loss of generality. Figure 1(a) depicts a conventional imager that measures the object irradiance on a two-dimensional detector array. The object is treated as a set of K disjoint object blocks, each of which is assumed to contain $\sqrt{N} \times \sqrt{N}$ pixels. Features will be computed independently on different object blocks so that each block may be treated as a vector $\mathbf{g}_i$ with dimension N, where $i = 1, 2, \ldots K$. The use of object blocks (instead of full objects) allows us to consider both local and global features within a common framework, and because each block is treated independently, we can often drop the subscript without producing any confusion. The measurement of each object block $\mathbf{g}$ is corrupted by noise to yield the estimate $\hat{\mathbf{g}}$. We first consider the case of additive white Gaussian noise (AWGN) so that $\hat{\mathbf{g}} = \mathbf{g} + \mathbf{n}$, where $\mathbf{n}$ represents a zero-mean AWGN process with variance σ^2. Shot noise and quantization noise are considered in the next section. In Fig. 1(a) the linear object features are computed as a postprocessing step by using the

$M \times N$ projection matrix $\mathbf{P}$ to yield the feature estimates $\hat{\mathbf{f}}_c \in \Re^M$. We quantify the feature fidelity by use of the MSE criterion as $\mathrm{MSE}_c = \langle|\mathbf{f} - \hat{\mathbf{f}}_c|^2\rangle/M$, where $\mathbf{f}$ is the desired (noise free) feature vector and $\langle\ \rangle$ represents statistical expectation. Assuming that the noise variables are independent and identically distributed (iid) and further assuming that the rows of the projection matrix are normalized, we find that $\mathrm{MSE}_c = \sigma^2$. This simple result demonstrates that the fidelity of features computed via postprocessing a conventional image is limited by the detector noise alone.

Figure 1(b) depicts the operation of a feature-specific imager. Instead of measuring the object irradiance distribution in preparation for later electronic feature computation, this optical system is responsible for computing the desired projections and measuring them directly. It is simple to imagine an optical system based on transparencies that will accomplish this goal.[13,14] More photon-efficient solutions based on diffractive and/or reflective optics can also be envisioned. In Section 5 we will present a novel polarization-based optical solution. Regardless of the specific hardware realization, however, we can consider the detected signal in our feature-specific imager to be a corrupted version of the desired feature vector. In the case of AWGN we can write the M-dimensional measurement vector as $\mathbf{m} = \hat{\mathbf{P}}\mathbf{g} + \mathbf{n}$, where $\hat{\mathbf{P}}$ is the projection matrix that is realized by the optical hardware. There are many reasons why $\mathbf{P} \neq \hat{\mathbf{P}}$ (e.g., imperfect optical components, finite optical bandwidth, misalignment, etc.) and most of these will depend upon the specifics of the hardware implementation, however, there are two aspects of the optics that are unavoidable. The first unavoidable aspect of the optical solution derives from the conservation of energy. In particular, no column in the $\hat{\mathbf{P}}$ matrix can sum (in absolute value) to greater than one. This constraint ensures that for a given object irradiance distribution and measurement time, the same number of photons are used in both the conventional and the feature-specific measurements. The mathematical implication of this photon-count constraint is a scaling of the measurement so that $\hat{\mathbf{P}} = \mathbf{P}/C$, where C is the maximum absolute column sum of $\mathbf{P}$. The actual feature estimate is thus obtained by multiplying the measurement $\mathbf{m}$ by the constant C to produce $\hat{\mathbf{f}}_m = \mathbf{P}\mathbf{g} + C\mathbf{n}$. We see that the photon-count constraint has resulted in an effective noise scaling for the direct feature measurement.

The second unavoidable aspect of an optical feature-specific imaging solution is the need for dual-rail signaling. Because negative quantities cannot be represented directly in an incoherent optical system, two complementary arms will be required to realize a projection matrix with bipolar quantities: One arm uses the positive values of $\mathbf{P}$ to generate the measurement $\mathbf{m}^+ = \mathrm{k}\mathbf{P}^+\mathbf{g}/C_+ + \mathbf{n}$ and the other arm uses the negative values to compute $\mathbf{m}^- = (1 - k)\mathbf{P}^-\mathbf{g}/C_- + \mathbf{n}$, where k is the power-splitting ratio of the dual-rail system, the ijth element of $\mathbf{P}^+$ is given by ${P_{ij}}^+ = P_{ij}$ if $P_{ij} > 0$ and ${P_{ij}}^+ = 0$ otherwise, and C_+ is the maximum column sum of $\mathbf{P}^+$. $\mathbf{P}^-$ and C_- are defined in an analogous way. In the case of a dual-rail implementation, the final feature estimate is produced by subtracting the scaled outputs of the two arms: $\hat{\mathbf{f}}_m = C_+\mathbf{m}^+/k - C_-\mathbf{m}^-/(1 - k)$. We can once again use the MSE criterion to quantify the performance of this feature-specific imager. We define $\mathrm{MSE}_m = \langle|\mathbf{f} - \hat{\mathbf{f}}_m|^2\rangle/M$ and in the case of AWGN, this can be reduced to $\mathrm{MSE}_m = \{[C_+/k]^2 + [C_-/(1 - k)]^2\}\sigma^2$. Because in general $C_+ \neq C_-$ we find that $k = 0.5$ is often not the optimal splitting ratio for minimizing MSE in a dual-rail feature-specific imager. Throughout the remainder of this paper all results will use the optimal (projection-matrix–dependent) splitting ratio.

In contrast with a conventional imager, the above discussion of feature-specific imaging indicates that the feature MSE depends on both the measurement noise and the characteristics of the projection matrix. In the next section we will see that because C_+ and C_- depend on the number of rows in $\mathbf{P}$, in cases for which relatively few features are required $\{[C_+/k]^2 + [C_-/(1 - k)]^2\} < 1$ and a multiplex advantage can be obtained from the feature-specific imager.[15]

3. Feature Fidelity Results

The feature-specific imager depicted in Fig. 1(b) provides two potential advantages over the conventional imager shown in Fig. 1(a). First is the reduced detector count required to accomplish some overall system goal. If the task for which an imager has been deployed requires M features, then the feature-specific imager makes $2M$ measurements per block as compared with the N (often $\gg 2M$) measurements made by a conventional imager. This reduction in detector count can impact both the complexity and maximum frame rate of the resulting imager. Second is the multiplex advantage described in the previous section. The potential reduction in MSE that accompanies the use of a feature-specific imager derives from the increase in the number of photons incident on each detector relative to the (fixed) measurement noise. This increase in measurement signal-to-noise ratio (SNR) translates directly into a reduction in feature MSE. Because the MSE reduction depends on the projection matrix we will quantify it for four cases of interest.

The first projections that we will consider are the so-called Hadamard projections. A Hadamard matrix of order N is a $N \times N$ orthogonal matrix of positive and negative ones. Figure 2 shows two example Hadamard matricies of size 8×8. The black squares indicate $+1$ and the white squares indicate -1. The rows in the matrix can be ordered in one of two ways. Figure 2(a) represents the natural ordering dictated by the recursive use of a Hadamard matrix of size 2, i.e., [1 1; 1 −1]. Figure 2(b), in contrast, is generated by a sequential ordering for which the number of sign changes increases by one with each row. It is sometimes convenient to use the

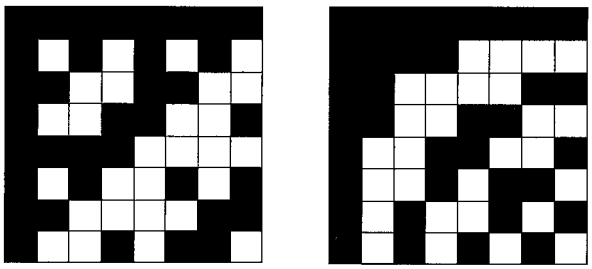


Fig. 2. Hadamard projection matricies of order 8 with: (a) natural ordering, (b) sequential ordering.

sequential ordering because the features measured by using this version of the Hadamard matrix are generated in order of increasing frequency. Hadamard matrices are important for many reasons. Although their use in multiplex measurement originates with the derivation of optimal weighing designs, they are also used in spectroscopy to achieve the multiplex advantage.[16,17] Hadamard projections have also been used in picture processing, coding, and transmission to achieve bandwidth reduction.[18]

We also consider feature-specific imaging by using wavelet projections. The wavelet transform is a multiresolution analysis tool that exploits the ability of certain classes of function to localize in both space and spatial frequency. Wavelets are widely used for image compression, signal/image analysis, and to facilitate the fast solution of certain linear systems.[9,19] In this work we will use two wavelet families: integer-valued Haar wavelets and real-valued Daubechies (DAUB4) wavelets. A wavelet decomposition can be expressed with a single unitary matrix transformation $\mathbf{P}$, where the rows of $\mathbf{P}$ are the wavelet projections arranged in order of increasing spatial frequency. Figures 3(a) and 3(b) show example 8×8 projection matricies for Haar and DAUB4 wavelets respectively.

The third linear features that we will consider here are KL features. KL features are optimum in the sense that they provide dimensionality reduction with the smallest possible MSE. This optimality results from their energy compaction property: A large percentage of energy in a zero-mean object can be concentrated into a small number of KL features. The KL transform also minimizes the correlation among features, providing information-optimal compression for Gaussian sources. For these reasons KL projections are often used for data compression, pattern recognition, and signal de-noising.[20–22] The KL projections are computed from training data and are given by the eigenvectors of the data covariance matrix. Figure 4(a) shows the ten faces that were used as training data for this work. Each face in this set was decomposed into $\sqrt{N} \times \sqrt{N}$ blocks and the resulting set of N-dimensional vectors was used to compute the data covariance matrix. The eigenvectors of the covariance matrix were then sorted in decreasing order of their corresponding eigenvalues. Figure 4(b) depicts the first 16 KL features for the training data shown in Fig. 4(a) for the case $N = 64$. Each eigenvector has been independently scaled into the range 0–255 to generate a gray-scale image for display purposes.

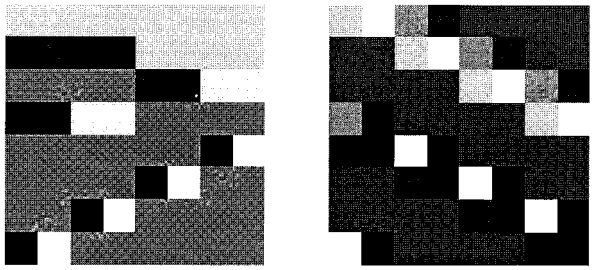
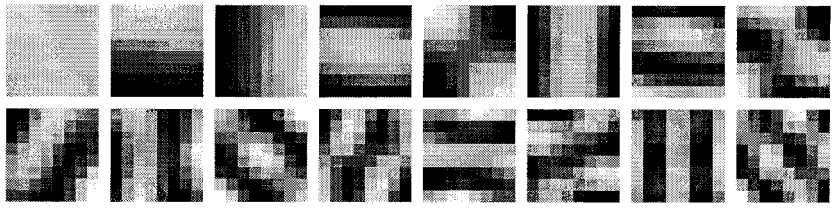


Fig. 3. Wavelet projection matricies of order 8: (a) Integer-valued Haar wavelets, (b) real-valued DAUB4 wavelets.

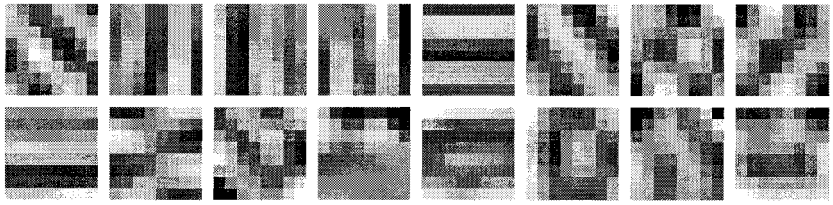
The last type of features that we will consider are the so-called independent components.[12] Independent component analysis (ICA) provides features that are statistically independent for any source distribution, converging to KL features for Gaussian sources. The ICA defines a linear transformation that is optimal in that it minimizes the mutual information among the transform coefficients. ICA has been used for various types of feature extraction as well as statistical applications such as blind source separation.[23,24] ICA features can be ordered in terms of their information content to produce a projection matrix $\mathbf{P}$. Using the training data shown in Fig. 4(a), we have computed ICA features using an algorithm described in Ref. 25. The first 16 ICA features are shown in Fig. 4(c) for the case of $N = 64$.

We know from Section 2 that the relative performance of feature-specific imaging can be written as $\eta = \mathrm{MSE}_m/\mathrm{MSE}_c = \{[C_+/k]^2 + [C_-/(1 - k)]^2\}$ and that the values of C_+ and C_- as well as the optimal value of k will depend on (a) the feature set of interest and (b) the number of features used to define the $\mathbf{P}$ matrix. Each of the feature sets described above has a natural ordering. For example, if $M < N$ KL features are desired then the M projections corresponding to the largest eigenvalues of the data covariance matrix are generally used. Using this natural ordering we can plot η as a function of the number of features M, for each of the four feature sets described above. We will use optimal power splitting into each arm of the dual-rail system throughout this exercise. The result is shown in Fig. 5(a) for the case $N = 64$. It is important to notice that all points for which $\eta < 1$ correspond to configurations for which feature-specific measurements produce higher feature fidelity than can be achieved via postprocessing a conventional image. From the data in Fig. 5(a) we see that $\eta \leq 1.0$ for $M \leq 2$ for ICA and KL features, $M \leq 3$ for Hadamard and Haar wavelet features, and $M \leq 4$ for DAUB4 wavelet features. From this data we conclude that for applications in which relatively

(a)



(b)



(c)

Fig. 4. (a) Ten faces used for training the KL and ICA features, (b) first 16 KL features computed using $N = 64$. Each 64-dimensional feature appears as an 8×8 image, (c) first 16 ICA features computed using $N = 64$. These features are rendered as in (b).

few features are required to meet the overall system goal, feature-specific multiplex imaging can be superior to conventional imaging. The diversity of feature sets included here gives us confidence that the general behavior represented by the data in Fig. 5(a) will persist for other feature sets as well.

The value of M for which each of these features sets achieves $\eta = 1.0$ represents a small fraction of the total $N = 64$ possible features. It is interesting to examine how the $\eta = 1.0$ condition depends upon the block size N. We will consider the example of KL features. In Fig. 5(b) we plot η versus the fractional number of KL features (M/N) for several values of N. From this data we see that the use of larger blocks will indeed increase the number of features that can be measured in a feature-specific imager while maintaining $\eta < 1.0$; however, the fraction of feature dimensions at which the $\eta = 1.0$ condition occurs is found to be approximately 5%, decreasing slightly with increasing N.

It is also possible to investigate feature-specific imaging under non-AWGN noise models. We have considered two such models: shot noise and quantization noise. In the case of shot-noise–limited imaging we do not expect an increase in the number of photons reaching a detector to provide any improvement in SNR. This is because the shot-noise-limited SNR is inherent in the irradiance collected by the instrument and does not depend on the energy incident on any particular detector. Thus we expect no multiplex advantage for feature-specific imaging in the presence of shot noise. Simulation results for this case are shown in Fig. 5(c), which uses the same feature sets ($N = 64$) as were used to generate the data in Fig. 5(a). Because $\eta = \mathrm{MSE}_m/\mathrm{MSE}_c$ cannot be written in a simple form as it was for AWGN, we use Monte Carlo simulation to generate the data in Fig. 5(c). From this data we see that $\eta \geq 1$ for all values of M confirming our intuition and demonstrating that feature-specific imaging provides no advan-

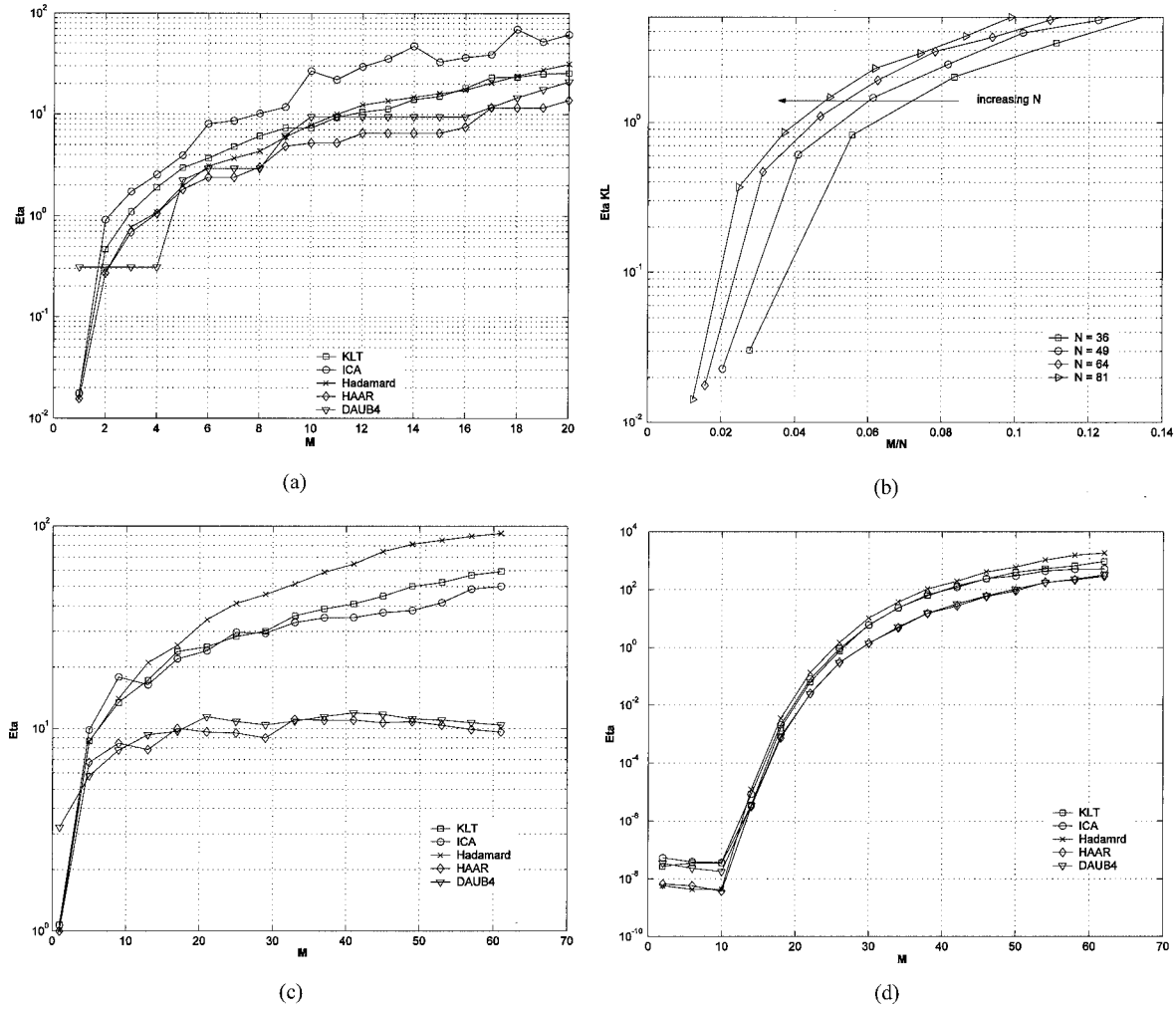
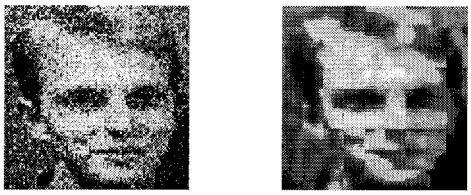


Fig. 5. (a) Relative feature fidelity (i.e., η) versus number of features (M) under AWGN noise model described in the text for $N = 64$. (b) η versus fractional number of KL features under AWGN model for $N = 36$, 49, 64, and 81. η versus M under (c) shot-noise, and (d) quantization noise models described in the text for $N = 64$.

tage under a shot–noise model. For the case of quantization noise the situation is quite different. Once again we investigate the behavior of η as a function of M; however, in this case we assume that 8-bit quantization limits the MSE of the conventional image measurement. Features are then computed from the measured image to generate some feature error (MSE_c). We consider bandwidth to be the constrained resource for this case, and we note that 8-bit quantization results in a data rate that scales like $8N$. This suggests that more precise analog to digital converters can be used in a feature-specific imager while maintaining the same data rate. In particular we assume that the word length used for each feature is given by $\min(24, 8N/M)$, where the minimum function is used to upper bound (e.g., to limit implementational complexity) the number of bits allocated to direct feature measurements. The results of this study are shown in Fig. 5(d). From this data we see that a large range of M values produce $\eta < 1$ suggesting that feature-specific imaging can be a powerful technique within a quantization-noise–limited environment such as the one described above.

4. Karhunen–Loeve Imaging

The preceding discussion has been quite general and has quantified the potential benefit of direct feature measurements in terms of feature fidelity. We have seen from this discussion that a trade-off exists between feature fidelity and the number of measured features, and that this tradeoff is governed by the photon-count constraint. In this section we will consider an example of feature-specific imaging in which we demonstrate the utility of direct feature measurements toward some overall imaging system goal. We select KL features for this example because (a) this feature set is familiar to a broad audience, (b) the system goal for which these features are optimal (i.e., minimum reconstruction MSE) is intuitive and can be represented visually, and (c) we find the somewhat surprising result that even when the system goal is to produce a "pretty picture" there are cases in which it



(a) (b)

Fig. 6. Reconstructed images obtained with (a) conventional imaging ($MSE_d = 500$), (b) feature-specific imaging ($MSE_{kl} = 154$). AWGN corrupts all measurements with $\sigma^2 = 500$ and the feature-specific imager used $M = 4$ and $N = 64$.

is preferable to measure features than to measure a conventional image.

We begin by considering a conventional imaging system that is subject to additive white Gaussian measurement noise with variance σ^2. Such a system produces the object estimate $\hat{\mathbf{g}}_d = \mathbf{g} + \mathbf{n}$, where $\mathbf{n}$ represents the AWGN. The pixelwise MSE associated with an image obtained via such a direct imaging system is simply $MSE_d = \langle|\mathbf{g} - \hat{\mathbf{g}}_d|^2\rangle/N = \sigma^2$. An example of a measured image with $\sigma = 22$ is given in Fig. 6(a). Alternately, a feature-specific imager might measure some number (M) of KL features $\{\hat{f}_i : i = 1, \ldots, M\}$ to produce the object estimate $\hat{\mathbf{g}}_{kl} = \sum_{i=1}^{M} \hat{f}_i \mathbf{p}^i$ where $\mathbf{p}^i$ is the ith row of $\mathbf{P}$ and is defined by the (normalized) ith largest eigenvector of the KL training-data covariance matrix. See figure 4b for an example of the first 16 eigenvectors computed using the training faces shown in figure 4a. It is important to note that the pixelwise MSE associated with $\hat{\mathbf{g}}_{kl}$, specifically $MSE_{kl} = \langle|\mathbf{g} - \hat{\mathbf{g}}_{kl}|^2\rangle/N$ contains two components. One of these arises from discarding (N–M) terms from the full KL expansion of the image. The other arises from the AWGN associated with the feature measurements. Including both of these components we can write MSE_{kl} as

$$MSE_{kl} = \frac{1}{N}\sum_{i=M+1}^{N} \lambda_i + \frac{M\sigma^2}{N}\left[\left(\frac{C_+}{k}\right)^2 + \left(\frac{C_-}{1-k}\right)^2\right], \tag{1}$$

where λ_i is the ith-largest eigenvalue of the KL training-data covariance matrix. From this expression it is clear that a tradeoff exists between these two sources of error. Increasing the number of features decreases the first term while it increases the second.

Figure 7(a) shows a plot of MSE_{kl} versus M for $\sigma = 22$ and several values of N. From this graph we can see that there is indeed an optimum number of features for each blocksize. This optimum choice of M balances the error arising from the truncated KL expansion against the error arising from the measurement process. We also notice that for all values of N considered here, the optimal value of MSE_{kl} is less than σ^2, which indicates that the feature-specific approach can be superior to conventional imaging for this example. Extracting the minimum MSE from each curve in Fig. 7(a) we can plot the optimal MSE versus N as shown in Fig. 7(b). This figure reflects a trade-off with the blocksize N. We can understand this trade-off in terms of a balance between image energy and feature fidelity. In particular we recall that increasing blocksize results in improved feature fidelity at constant M [see Fig. 5(b)], however, the fraction of image energy contained in that fixed number of features goes down with increasing N. The result shown in Fig. 7(b) suggests that $N = 121$ is optimal for this example. Figure 6(b) shows the reconstructed image associated with the optimal KL feature-specific imager for $\sigma = 22$. It uses $M = 6$ and $N = 121$ and achieves $MSE_{kl} = 132$. Because the relative performance of the feature-specific approach will be sensitive to the value of σ^2, we can repeat the analysis that gave rise to the graphs of Figs. 7(a) and 7(b) for many values of noise, extracting only the optimal performance. The result is shown in Figure 7c. This graph presents the MSE performance of both conventional (open squares) and KL feature-specific (open circles) imaging versus noise. From this figure we see that the KL feature-specific approach offers significant advantages in high-noise environments (i.e., $\sigma^2 > 74$). A convenient way to quantify the value of a KL imager is represented by the crossover point for these two curves. This is the value of noise above which KL imaging is superior to conventional imaging. From Fig. 7(c) we see that this crossover point is $\sigma_x = 8.5$.

The second term in Eq. (1) depends on the fidelity of the optical feature measurement process. In the previous section we discussed how this process is influenced by two physical constraints: (a) the positive-valued nature of optical irradiance resulted in the need for dual-rail signaling and (b) the conservation of energy required the scaling of the dual-rail projection matricies by their respective maximum column sums. Constraint (a) provided us with an additional degree of freedom in our optical-system design. We can optimize the fraction of collected photons that enter each arm of the dual-rail system via the parameter k. This optimization has in fact been performed for all data presented thus far. It is possible to carry this idea one step further, optimizing the fraction of photons that participate in each feature measurement. Consider scaling the ith row of the projection matrix $\mathbf{P}^+(\mathbf{P}^-)$ by a constant a_i (b_i). This results in a linear scaling of the corresponding feature measurements so that $\hat{f}_i^+ = f_i^+ + C_+(\mathbf{a})n/a_i$ and $\hat{f}_i^- = f_i^- + C_-(\mathbf{b})n/b_i$, where $\mathbf{a} = (a_1, a_2, \ldots, a_M)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_M)$ and n is the AWGN process. Note that the presence of these new feature scaling factors will change the values of the maximum column sums so that $C_+(\mathbf{a})$ and $C_-(\mathbf{b})$ are now written as explicit functions of $\mathbf{a}$ and $\mathbf{b}$. If we implement such a scaling operation within a KL imaging system,

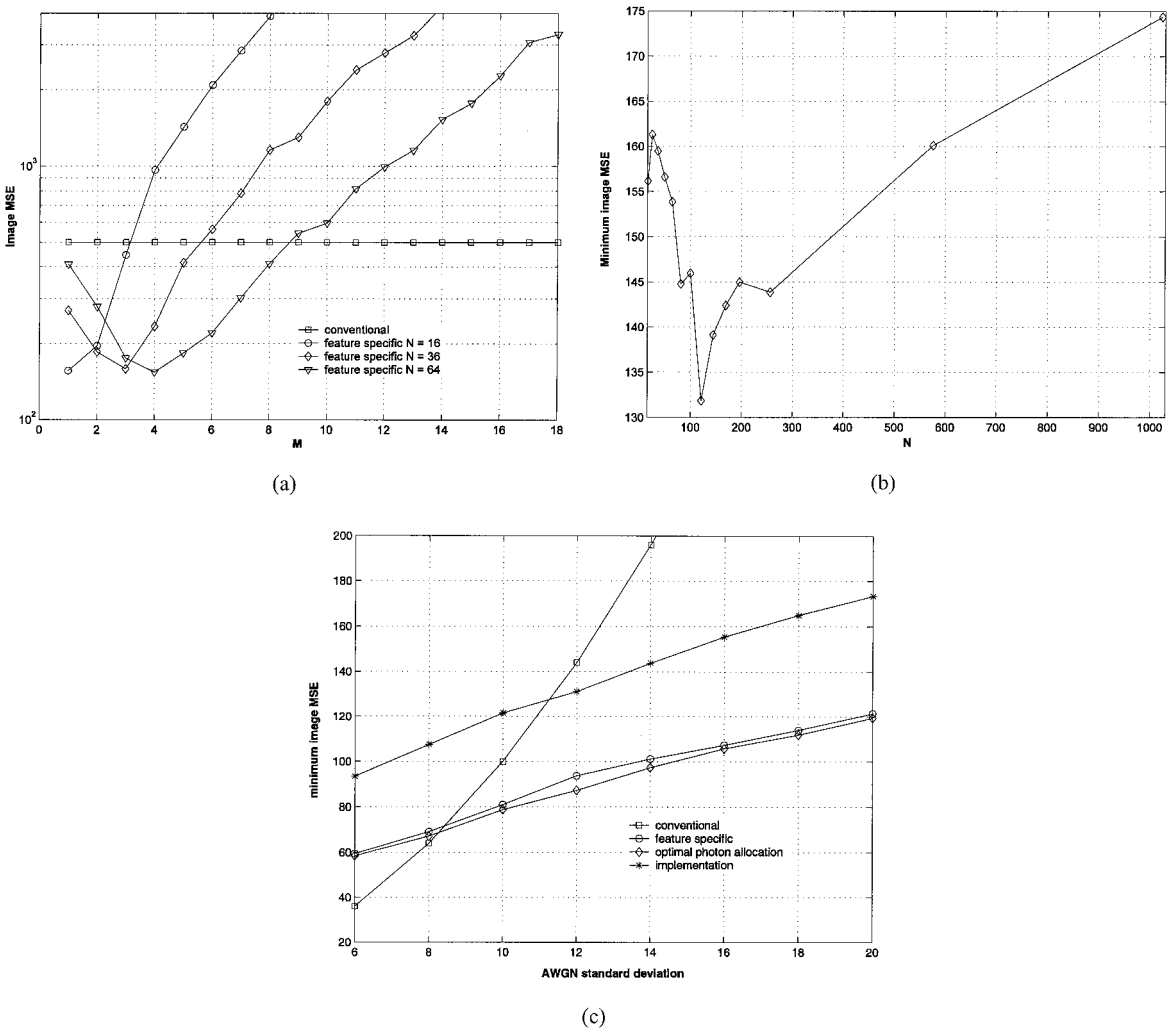


Fig. 7. Reconstruction MSE for KL feature-specific imaging corrupted by AWGN. (a) MSE versus number of features for $\sigma^2 = 500$ and $N = 16$, 36, and 64. (b) Minimum MSE versus blocksize for $\sigma^2 = 500$. (c) Optimum KL feature-specific imaging performance versus noise standard deviation.

the resulting MSE in the image estimate MSE_{kl}, can now be written as

$$\mathrm{MSE}_{kl} = \frac{1}{N}\sum_{i=M+1}^{N}\lambda_i + \frac{\sigma^2}{N}\sum_{i=1}^{M}\left[\left(\frac{C_+(\mathbf{a})}{a_i}\right)^2 + \left(\frac{C_-(\mathbf{b})}{b_i}\right)^2\right]. \quad (2)$$

From this equation we can see that it is possible to exploit the additional degrees of freedom in **a** and **b** to further minimize the MSE of the KL imaging system. The complex dependence on **a** and **b** that is represented by the maximum column sums $C_+(\mathbf{a})$ and $C_-(\mathbf{b})$, makes this a challenging optimization problem. We use a modified gradient search method for each value of M to find the optimal **a** and **b** vectors. From these optimized system designs we can select the best value of M for each value of noise. The result of this process is shown by the curve with open diamond symbols in Fig. 7(c). We notice that the optimal allocation of photons has resulted in a crossover point that is slightly lower ($\sigma_x = 8.25$), indicating a performance improvement of $10\log(8.5/8.25) = 0.13$ dB.

KL imaging uses projection matricies $\mathbf{P}^+$ and $\mathbf{P}^-$ that are based on the eigenvectors of the training-data covariance matrix. This guarantees optimal (noise-free) performance for test images that are selected from the same underlying image distribution as the training data. The previous results in this section are based on a test image that is similar to (but not contained in) the training data. In cases for which the testing data is dissimilar from the training data, we expect KL imaging performance to degrade. Using the same KL imaging system as we used to generate the (optimal) data in Fig. 7, we have quantified this degradation for the five test images shown in Fig. 8. In all five cases the crossover noise value is increased as compared with the optimal crossover value from Fig. 7(c) ($\sigma_x = 8.25$). Table 1 shows the crossover noise cost (in dB) for each of these test images.

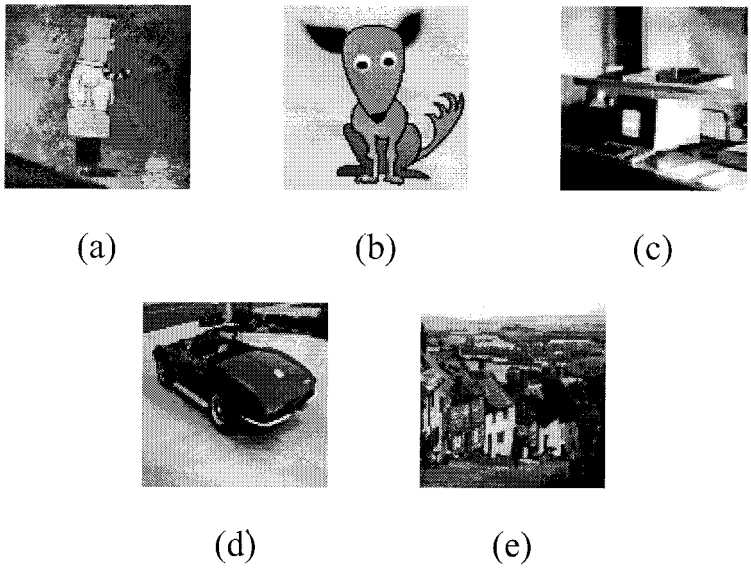


Fig. 8. Test images for evaluating training-set-sensitivity of KL feature-specific imaging: (a) Man, (b) Dog, (c) Desk, (d) Car, and (e) Goldhill.

Table 1. Crossover Noise Cost[a]

Image	Man	Dog	Desk	Car	Goldhill
Crossover noise cost	1.6	3.3	3.4	3.7	3.7

[a]For test images that differ from the training set.

5. Polarization-Based Optical System for Feature-Specific Imaging

Figure 9 shows a candidate optical architecture for feature-specific imaging. It is a polarization-based dual-rail optical pipeline processor. The system shown in this figure is suitable for the measurement of two features, however, additional processing stages can be cascaded as shown. Because the performance of feature-specific imaging (as compared with conventional imaging) so critically depends on the efficient use of all the collected photons, this system is designed to be photon efficient. Its operation will first be described for the case of a single feature. We assume that light entering the feature-specific

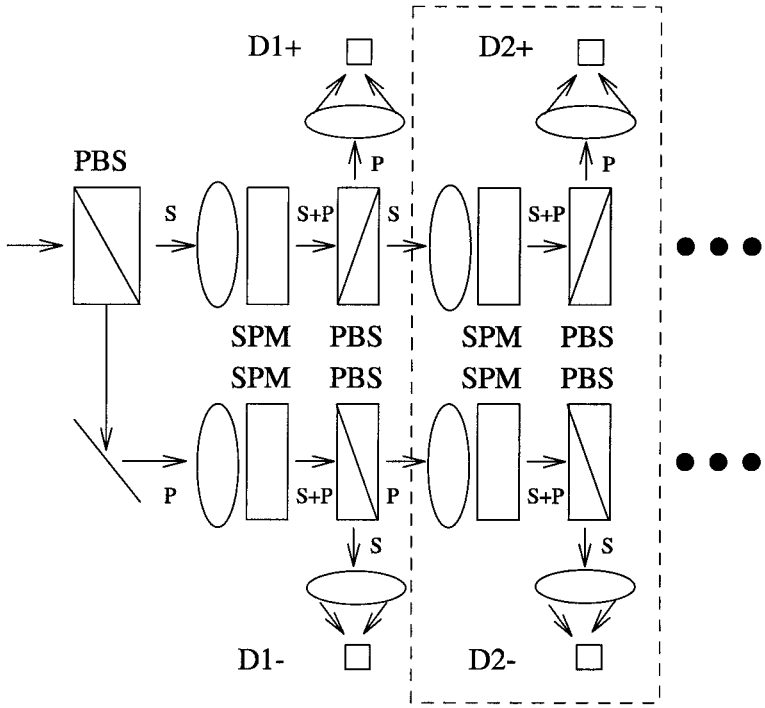


Fig. 9. Candidate feature-specific imaging architecture: PBS = polarizing beamsplitter, SPM = spatial polarization modulator, D1+ and D1− are detectors for the positive and negative components of the 1st feature measurement and D2+ and D2− are for the 2nd feature.

imager is unpolarized. This unpolarized incident light first passes through a polarizing beamsplitter (PBS) as shown. The PBS separates the incident light into *s*- and *p*-polarizations, directing each into a separate arm of the dual-rail system. The light in each arm is imaged onto a spatial polarization modulator (SPM) that rotates the polarization at each pixel of the optical irradiance pattern by an angle given by the desired projection $\mathbf{p}^1$. The rotations in the positive arm are given by $\theta 1_i^{+} = \sin^{-1}(\sqrt{p_i^1/C_+})$ if $p_i^1 > 0$ and $\theta 1_i^{+} = 0$ otherwise. For the negative arm we have the complementary scheme: $\theta 1_i^{-} = \sin^{-1}(\sqrt{p_i^1/C_-})$ if $p_i^1 < 0$ and $\theta 1_i^{-} = 0$ otherwise. Notice that normalization by the maximum column sums C_+ and C_- ensures that the maximum rotation imparted by any pixel is 90°. These rotations can be imparted for example by using an electrically addressable device such as a liquid-crystal (LC) modulator, or by using a fixed device such as a surface-relief profile etched into an optically active crystal. In either case the orthogonal component of polarization generated by the SPM in each arm is re-directed by use of polarizing beamsplitters and is integrated onto the photodetectors as shown in Fig. 9.

Now we describe the operation of this system in the case of two features. The prescription for the rotation angles $\theta 1_i^{+}$ and $\theta 1_i^{-}$ remains unchanged in this case, however, it is important to note two things. First, note that the un-rotated light propagates past the PBS elements and can be used for the computation of additional features. Second, note that the normalization of $\mathbf{P}^{+}$ and $\mathbf{P}^{-}$ by their respective maximum column sums ensures that there are sufficient photons for the computation of these successive features. The prescription for the rotation angles in stage two, however, must be modified to account for the reduced irradiance at each pixel of the stage 2 SPMs. Specifically we have $\theta 2_i^{+} = \sin^{-1}(\sqrt{p_i^2/(C_+ - p_i^1)})$ if $p_i^2 > 0$ and $\theta 2_i^{+} = 0$ otherwise. A complementary scheme is used for the negative arm. From Fig. 9 we see that the light exiting each arm of stage 1 is reimaged onto the SPMs in each arm of stage 2 so that feature 2 may be computed via the induced polarization rotations described above. Additional stages operate in the same way.

The optical system described above is photon-efficient by design; however, there are unavoidable implementational imperfections that will impact the performance of such a feature-specific imager. To quantify the impact of these imperfections we have formed a computer model of the system shown in Fig. 9. The model includes various sources of imperfection, such as (1) imperfect PBS devices with 98% transmission and 2% leakage, (2) imperfect coatings that give rise to surface reflections—all surfaces in our model are taken to be 99% transmissive, (3) LC SPM elements subject to a uniform noise term (± 0.05V) in their applied voltage, and of course AWGN at each detector. We have also included a logistic response function (shown in Fig. 10) to model each SPM; however, many other sources of experimental error have been overlooked in this preliminary analysis. We expect for example that SPM devices will not have a 100 areal fill-factor, and/or may not have a uniform response at all pixels or for all incident angles. These SPM characteristics have not yet been included in our model. The results of our preliminary implementational study are shown by the asterisk curve in Fig. 7(c). We see from this data that the presence of these device/system imperfections have degraded the performance of the KL imager, generating a crossover noise cost of 1.34 dB. We also notice that this degradation is most deliterious in low-noise environments and that significant MSE advantage is still provided by KL imaging when the detector noise level is high.

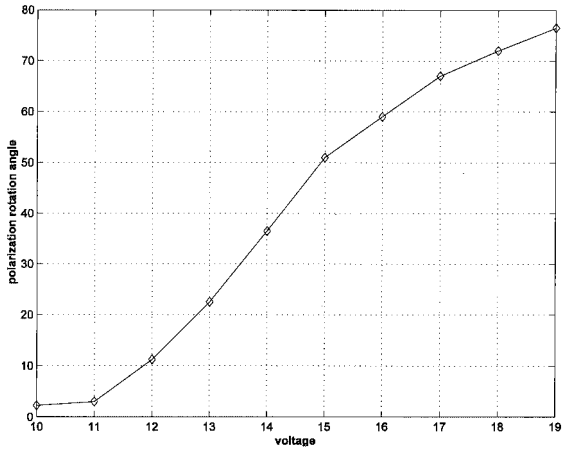


Fig. 10. SPM response function used in optical system tolerancing study.

6. Conclusions

This paper has presented an analysis of feature-specific imaging. An imager is often deployed as part of a larger system designed to achieve some diagnostic task. In cases for which the overall system goal can be described in terms of linear projections of the object irradiance distribution, we consider the possibility of direct optical feature measurement. We have quantified the fundamental limits of such an approach. We have compared the feature fidelity associated with direct feature measurements to the fidelity achieved by postprocessing a conventional image, and we find a tradeoff between the number of features and the corresponding feature fidelity. In both AWGN and quantization-noise–limited imaging environments this tradeoff results in higher fidelity for direct feature measurements (as compared with processing a conventional image) in cases for which relatively few features are necessary to adequately meet the goals of the overall system.

A common alternate approach to the measurement of linear object features is the optical correlator. It is instructive to consider the photon inefficiencies that result from the use of optical correlation within the feature-specific imaging environment described here. Two observations can be made in this regard.

First, a correlator is shift invariant so that the blockwise linear projections that we describe are obtained at the center of each block in the correlation plane. Photons arriving at other locations are unused resulting in poor feature fidelity. Second, a single correlator implements a Toeplitz projection matrix that depends upon only a single feature. Multiple correlators are therefore required (one for each row in the desired projection matrix) to realize the desired computation. Division of photons among these multiple correlators represents another source of photon inefficiency. In this paper we have shown that the use of a single projection matrix, augmented to optimize the efficient utilization of photons, can result in feature-specific imaging with high fidelity.

A specific example of feature-specific imaging based on KL projections was analyzed and yielded the important result that it is possible to improve image MSE via direct feature measurement in high AWGN environments. A candidate optical system was presented and a preliminary implementational study demonstrates that such a system can efficiently measure linear object features and can provide improved performance as compared with conventional imaging. This work is continuing along several important dimensions. In the course of this work we have noticed that KL features are suboptimal for minimum MSE image reconstructions when feature measurements are corrupted by noise. We are now seeking a new formulation for the optimum feature set in this case. Extension of the techniques described here to include multispectral imaging is also underway so that correlation among wavelength bands can be exploited to generate more efficient use of photons in these systems. We are also working to improve our implementational model with the eventual goal of constructing an optical system for feature-specific imaging.

References

1. M. P. Christensen, G. W. Euliss, M. J. McFadden, K. M. Coyle, P. Milojkovic, M. W. Haney, J. van der Gracht, and R. A. Athale, "Active-eyes: an adaptive pixel-by-pixel image-segmentation sensor architecture for high-dynamic-range hyperspectral imaging," Appl. Opt. **41,** 6093–6103 (2002).
2. E. Clarkson and H. H. Barrett, "Approximations to ideal-observer performance on signal-detection tasks," Appl. Opt. **39,** 1783–1793 (2000).
3. W. C. Chou, M. A. Neifeld, and R. Xuan, "Information-based optical design for binary-valued imagery," Appl. Opt. **39,** 1731–1742 (2000).
4. S. Prasad, "Information capacity of a seeing-limited imaging system," Opt. Commun. **177,** 119–134 (2000).
5. S. Tucker, W. T. Cathey, and E. Dowski, "Extended depth of field and aberration control for inexpensive digital microscope systems," Optics Express **4,** 467–474 (1999).
6. D. L. Marks, R. Stack, A. J. Johnson, D. J. Brady, and D. C. Munson, "Cone-beam tomography with a digital camera," Appl. Opt. **40,** 1795–1805 (2001).
7. Z. Liu, M. Centurion, G. Panotopoulos, J. Hong, and D. Psaltis, "Holographic recording of fast events on a CCD camera," Opt. Lett. **27,** 22–24 (2002).
8. P. Potuluri, M. Fetterman, and D. Brady, "High depth of field microscopic imaging using an interferometric camera," Opt. Express **8,** 624–630 (2001).
9. M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," IEEE Trans. Image Process. **1,** 205–220 (1992).
10. M. Harwit and N. J. A. Sloane, *Hadamard Transform Optics*, (Academic, New York, 1979), pp. 62–70.
11. K. I. Diamantaras and S. Y. Kung, *Principal Components Neural Networks: Theory and Applications*, (Wiley, New York, 1996), pp. 44–73.
12. R. Stephen and E. Richard, *Independent Component Analysis: Principles and Practice*, R. Stephen and E. Richard, eds. (Cambridge University, Cambridge, UK, 2001), Chap. 1, 1–70.
13. C. J. Oliver, "Optical image processing by multiplex coding," Appl. Opt. **15,** 93–106 (1976).
14. R. D. Swift, R. B. Wattson, J. A. Decker, R. Paganetti, and M. Harwit, "Hadamard transform imager and imaging spectrometer," Appl. Opt. **15,** 1595–1609 (1976).
15. P. B. Fellgett and E. H. Linfoot, "On the assessment of optical images," Philos. Trans. R. Soc. London **247,** 269–407 (1955).
16. T. J. Vickers, C. K. Mann, and J. Zhu, "Hadamard Multiplex Multichannel Spectroscopy to Achieve a Spectroscopic Power Distribution Advantage," Appl. Spectrosc. **45,** 42–49 (1991).
17. A. G. Marshall and M. B. Comisarow, "Fourier and Hadamard transform methods in spectroscopy," Anal. Chem. **47,** 491–504 (1975).
18. H. J. Landau and D. Slepian, "Some computer experiments in picture processing for bandwidth reduction," Bell Syst. Tech. J. **50,** 1525–1540 (1971).
19. I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," IEEE Trans. Inf. Theory **36,** 961–1005 (1990).
20. H. Feng and M. Effros, "On the rate-distortion performance and computational efficiency of the Karhunen-Loeve transform for lossy data compression," IEEE Trans. Image Process. **11,** 113–122 (2002).
21. S. J. Lee, S. B. Jung, J. W. Kwon, and S. H. Hong, "Face detection and recognition using PCA," TENCON 99 Proc. IEEE Region 10 Conference **1,** 84–87 (1999).
22. A. Aysin, L. F. Chaparro, I. Grave, and V. Shusterman, "Denoising of Non-Stationary Signals Using Optimized Karhunen-Loeve Expansion," Proc. IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis, 621–624 (1998).
23. A. Hyvarinen, "Emergence of phase- and shift-invariant features by decomposition of natural images into independent feature subspaces," Neural Comput. **12,** 1705–1720 (2000).
24. H. Saruwatari, T. Kawamura, and K. Shikano, "Fast-convergence algorithm for ICA-based blind source separation using array signal processing," Proc. IEEE 11th Signal Processing Workshop on Statistical Signal Processing, 464–467 (2001).
25. A. Hyvarinen, "Fast ICA by a fixed-point algorithm that maximizes non-Gaussianity," *Independent Component Analysis: Principles and Practice*, R. Stephen and E. Richard, eds. (Cambridge University, Cambridge, UK, 2001), Chap. 2, pp. 71–94.