Statistical and Information-Theoretic Analysis of Resolution in Imaging

Morteza Shahram, *Member, IEEE*, and Peyman Milanfar, *Senior Member, IEEE*

***Abstract*—In this paper, some detection-theoretic, estimation-theoretic, and information-theoretic methods are investigated to analyze the problem of determining resolution limits in imaging systems. The canonical problem of interest is formulated based on a model of the blurred image of two closely spaced point sources of unknown brightness. To quantify a measure of resolution in statistical terms, the following question is addressed: "What is the minimum detectable separation between two point sources at a given signal-to-noise ratio (SNR), and for prespecified probabilities of detection and false alarm (P_d and P_f)?". Furthermore, asymptotic performance analysis for the estimation of the unknown parameters is carried out using the Cramér–Rao bound. Although similar approaches to this problem (for one-dimensional (1-D) and oversampled signals) have been presented in the past, the analyzes presented in this paper are carried out for the general two-dimensional (2-D) model and general sampling scheme. In particular the case of under-Nyquist (aliased) images is studied. Furthermore, the Kullback–Liebler distance is derived to further confirm the earlier results and to establish a link between the detection-theoretic approach and Fisher information. To study the effects of variation in point spread function (PSF) and model mismatch, a perturbation analysis of the detection problem is presented as well.**

***Index Terms*—Aliasing, Cramér–Rao bound, detection, estimation, Fisher information, imaging, information-theoretic imaging, Kullback–Liebler distance, model mismatch, perturbation analysis, resolution, variational analysis.**

I. Introduction

THE problem of resolution historically has been of significant interest in different communities in science and engineering, for example in astronomy, optics, different applications in physics, array processing and imaging. In this paper, we focus on the problem of achievable resolution in imaging practice for the following reasons. First, the problem of resolving point sources can be considered as a canonical case study to investigate the performance of imaging systems and image restoration techniques. Second, developing different techniques to resolve point sources is indeed a concern in real-world application, e.g., astronomy. Our approach to this problem in this paper is based on statistical detection and estimation theory, since we assume the measured signal from which the sources are to be resolved, is noise-corrupted.

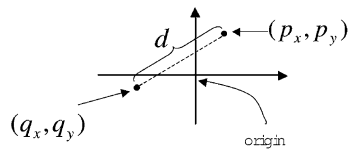

Fig. 1. Location of point sources.

To begin, let the original image of interest consist of two impulse functions positioned at points (p_x, p_y) and (q_x, q_y) with a small distance of $d = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2}$ apart (Fig. 1). That is, the signal model is

$$I(x, y) = \sqrt{\alpha}\delta(x - p_x, y - p_y) + \sqrt{\beta}\delta(x + q_x, y + q_y). \quad (1)$$

We consider the following two-dimensional (2-D) model for the measured (discrete) signal

$$\begin{aligned} g(x_k, y_l) &= s(x_k, y_l) + w(x_k, y_l) \\ &= \alpha h(x_k - p_x, y_l - p_y) \\ &\quad + \beta h(x_k + q_x, y_l + q_y) \\ &\quad + w(x_k, y_l), \quad k, l \in \{1, 2, \ldots, N\} \end{aligned} \quad (2)$$

where $h(x, y)$ is the blurring kernel[1] (representing the overall point spread function (PSF) of the imaging system) and $w(x_k, y_l)$ is assumed to be a zero-mean Gaussian white noise[2] process with variance σ^2.

According to the classical Rayleigh criterion, two point sources are "barely resolvable" when the central peak of the pattern generated by one point source falls exactly on the first zero of the pattern generated by the second one [6]. See Fig. 2 for a depiction. But in fact under certain conditions related to signal-to-noise ratio (SNR), resolution beyond the Rayleigh limit is indeed possible. This can be called the "super-resolution" limit.

The statistical analysis presented in this paper is formulated based on the ability to distinguish whether the measured image is generated by one point source or two point sources. Specifically, for the proposed model the equivalent question is whether the parameter d is equal to zero or not. If $d = 0$, then we only have one peak and if $d > 1$ then there are two well resolved

Manuscript received November 16, 2004; revised April 23, 2006. This work was supported in part by the National Science Foundation under CAREER Award CCR-9984246 and the Air Force Office of Scientific Research under Grant F49620-03-1-0387.

M. Shahram was with the Department of Electrical Engineering, University of California, Santa Cruz, CA 95064 USA. He is now with the Department of Statistics, Stanford University, Stanford, CA 94305 USA (e-mail: mshahram@stanford.edu).

P. Milanfar is with the Department of Electrical Engineering, University of California, Santa Cruz, CA 95064 USA (e-mail: milanfar@ee.ucsc.edu).

Communicated by A. Høst-Madsen, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2006.878180

[1]For convenience and ease of presentation only, we assume that the point spread function is a symmetric function throughout, i.e., $h(x, y) = h(-x, y) = h(x, -y) = h(-x, -y)$.

[2]We do not concern ourselves in this paper with the case of photon-limited systems.

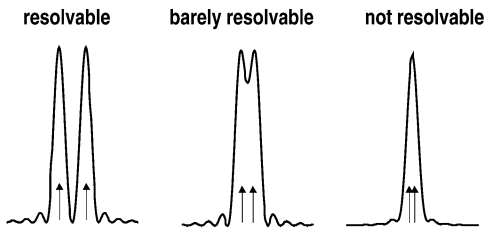

Fig. 2. Resolvability in the Rayleigh sense.

peaks according to the Rayleigh criterion.[3] So the problem of interest revolves around values of d in the range of $0 \leq d < 1$. This can be posed as a hypothesis testing problem, i.e.,

$$\begin{cases} \mathcal{H}_0 : d = 0 & \text{One point source} \\ \mathcal{H}_1 : d > 0 & \text{Two point sources} \end{cases} \tag{3}$$

or equivalently (see (4) at the bottom of the page). The problem of resolution from the statistical viewpoint has been well studied in the past [25], [2], [1], [7], [9]. The majority of researchers have used the Cramér–Rao bound to analyze the problem of resolution to study the mean-square error of unbiased estimators for the distance between the sources [8], [9], [1], [21], [22], [13], [16]. In addition, some papers [8], [2] have considered a hypothesis testing approach to determine the resolvability of sources in some limited cases for one-dimensional signals.

As for novel contributions in this paper, we formulate the problem of interest as a composite hypothesis test and derive an explicit relationship between the minimum detectable separation and the required SNR for any sampling strategy including aliased (sub-Nyquist) images. As a result, a quantitative measure of resolution is obtained which can be used to understand the effect of model parameters (e.g., PSF and sampling parameters). The following question is addressed: what is the minimum separation between two point sources (maximum attainable resolution) that is detectable with high probability at a given SNR. Besides, general analytical results are derived for arbitrary sampling schemes. Our previous work has been focused on the case of one-dimensional (1-D) signals and over-Nyquist sampling [14], [20], [19]. In this paper, we present extensions of these results to 2-D and to the under-sampled (aliased) case (for a general PSF).

We put forward a solution to (4) for its most general case where all the parameters involved in the signal model are unknown to the detector. Furthermore, we consider two cases, where the value of noise variance (σ^2) is known and where it is unknown. For both cases we develop corresponding detection strategies and obtain the explicit relationship between SNR and the parameters in the model. Also, to further analyze the problem of resolution. such frameworks will be explained in the following.

[3]Throughout this paper, we assume without loss of generality that $d = 1$ corresponds to the Rayleigh limit.

A significant question which has not been addressed in the past is to analyze the effect of a known or unknown perturbation of the PSF on the detection performance. Variation in PSF can be caused by other blurring elements in the system, for example an out-of-focus lens or atmospheric or underwater turbulence. We first address the problem of finding the change in the required SNR due to a variation of the PSF required for resolvability. This will help us to analyze sensitivity to model inaccuracies.

In the interest of completeness, we also compute the Fisher Information matrix (and Cramér–Rao (CR) bound) in closed form for two different cases. The Cramér-Rao lower bound formulation is used to study the limits to attainable precision of estimated distances of the two point sources. We carry out this analysis for the case of under-sampled images and for a general PSF.

Another appealing and informative analysis is to compute the symmetric Kullback–Liebler Distance (KLD) or Divergence [12, p. 26]. KLD is a measure of discriminating power between two hypotheses, and is directly related to the performance figure of a related optimal detector. To accurately compute the KLD for the underlying problem, we make some essential extensions to the conventional formula of approximating the KLD. We shall see an interesting and important connection between the KLD analysis and the detection-theoretic analysis.

As for the application of the presented analysis in this paper, we shall see that the framework of addressing the fundamental relationship in resolving point sources (i.e., the tradeoff between SNR and resolvability) will also suggest a corresponding detector which can be well used in practice. In other words, the established performance expression is simply the result of applying the proposed local detector.

The organization of the paper is as follows. In Section II, we will present and formulate our detection-theoretic approach. The asymptotic performance of maximum likelihood estimate of the unknown parameters in terms of the Fisher Information Matrix will be discussed in Section III. In Section IV, we will study an information-theoretic criterion for the problem of interest. We will explain our perturbation and variational analysis for the underlying detection problem and discuss the effect of model mismatches in Section V. We will also discuss some of the practical applications and possible extensions of the proposed framework in Section VI. Finally, some comments and conclusion will be presented in Section VII.

II. Detection-Theoretic Approach

In the test (4), when the model parameters are unknown, the probability density function (pdf) under both hypotheses is therefore not known exactly, resulting in a composite hypothesis test. It is therefore not possible to form the standard likelihood ratio test. A common approach in this case is the generalized likelihood ratio test (GLRT). GLRT first computes maximum likelihood (ML) estimates of the unknown parameters, and then will use the estimated values to form the standard Neyman–Pearson (NP) detector. While other useful (Bayesian)

$$\begin{cases} \mathcal{H}_0 : g(x_k, y_l) = & (\alpha + \beta) h(x_k, y_l) + w(x_k, y_l) \\ \mathcal{H}_1 : g(x_k, y_l) = & \alpha h(x_k - p_x, y_l - p_y) + \beta h(x_k + q_x, y_l + q_y) + w(x_k, y_l). \end{cases} \tag{4}$$

methods are also available for such composite problems, our focus will be on GLRT-type methods because of less restrictive assumptions and easier computation and implementation; but most importantly, because locally most powerful (LMP) tests can be developed for the range of small d (below the Rayleigh limit). Also, as will be shown later, the performance of such a detector is very close to that of an ideal detector, to which the values of all the parameters in the model are known. Hence, the performance of the suggested detector can be reasonably considered as an approximate performance bound in practice.

Applying the GLRT approach to the problem of interest directly will produce a highly nonlinear test statistic (see [20]). However, since the range of interest for the value of d is assumed to be small (below the Rayleigh limit $d < 1$), we can benefit from approximating the model of the signal for nearby point sources. The approximate model is obtained by expanding the signal in a Taylor series about the small parameter values around $(p_x, q_x, p_y, q_y) = (0, 0, 0, 0)$. By introducing the partial derivatives of $h(x, y)$ as

$$h_{ij}(x_k, y_l) = \left.\frac{\partial^{i+j} h(x,y)}{\partial x^i \partial y^j}\right|_{x=x_k, y=y_l} \tag{5}$$

we write the signal model in the following (lexicographically scanned) vector form (e.g., $[\mathbf{h}]_{lN+k} = h(x_k, y_l)$)

$$\tilde{\mathbf{s}} = \mathbf{H}\boldsymbol{\theta} \tag{6}$$

where

$$\mathbf{H} = [\mathbf{h}, \mathbf{h}_{10}, \mathbf{h}_{01}, \mathbf{h}_{20}, \mathbf{h}_{02}, \mathbf{h}_{11}] \tag{7}$$

$$\boldsymbol{\theta} = \begin{bmatrix} \alpha + \beta \\ -\alpha p_x + \beta q_x \\ -\alpha p_y + \beta q_y \\ \frac{1}{2}\left(\alpha p_x^2 + \beta q_x^2\right) \\ \frac{1}{2}\left(\alpha p_y^2 + \beta q_y^2\right) \\ \alpha p_x p_y + \beta q_x q_y \end{bmatrix}. \tag{8}$$

We elect to keep terms up to order 2 of the above Taylor expansion which gives a more accurate representation of the signal and avoids trivial approximations in cases where the first-order terms would simply vanish [19]. This approximation leads to a linear model detection problem and also is equivalent to the framework of locally most powerful tests [11, p. 218]. We first consider the case where the noise variance is known to the detector. By substituting the above approximated model into (4), the hypothesis test can be rewritten as

$$\begin{cases} \mathcal{H}_0: & \mathbf{A}\boldsymbol{\theta} = \mathbf{0} \\ \mathcal{H}_1: & \mathbf{A}\boldsymbol{\theta} \neq \mathbf{0} \end{cases} \tag{9}$$

where

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{10}$$

The test (9) is a problem of detecting a deterministic signal with unknown parameters. The GLRT for the approximated model yields [11, p. 274]:

$$T(\mathbf{g}) = \frac{1}{\sigma^2}\hat{\boldsymbol{\theta}}^T \mathbf{A}^T [\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1}\mathbf{A}\hat{\boldsymbol{\theta}} \tag{11}$$

where $\mathbf{g}$ is the (lexicographically scanned) vector form of the measured signal ($[\mathbf{g}]_{lN+k} = g(x_k, y_l)$) and

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{g} \tag{12}$$

is the unconstrained maximum likelihood estimate of $\boldsymbol{\theta}$. For any given data set $\mathbf{g}$, we decide $\mathcal{H}_1$ if the statistic exceeds a specified threshold

$$T(\mathbf{g}) > \gamma. \tag{13}$$

The choice of γ is motivated by the level of tolerable false alarm (or false-positive) in a given problem, but is typically kept very low. It is worth mentioning that since the hypothesis test in (3) is a one-sided test, the above formulations (the Taylor approximation and the generalized likelihood ratio setup for the problem in (9)) can be viewed as a locally most powerful detector [11, p. 218]. From (11), the performance of this detector is characterized by

$$P_f = Q_{\chi_5^2}(\gamma) \tag{14}$$

$$P_d = Q_{\chi_5'^2(\lambda)}(\gamma) \tag{15}$$

$$\lambda = \frac{1}{\sigma^2}\boldsymbol{\theta}^T\mathbf{A}^T[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1}\mathbf{A}\boldsymbol{\theta} \tag{16}$$

where $Q_{\chi_5^2}$ is the right tail probability for a Central Chi-Squared pdf with 5 degrees of freedom, and $Q_{\chi_5'^2(\lambda)}$ is the right tail probability for a noncentral Chi-Squared pdf with 5 degrees of freedom and noncentrality parameter λ. For a specific desired P_d and P_f, we can compute the implied value for the noncentrality parameter from (14) and (15). We call this value of the noncentrality parameter $\lambda(P_f, P_d)$. This notation is key in illuminating a very useful relationship between the SNR and the smallest separation which can be detected with high probability, and low false alarm rate. From (16) we can write

$$\sigma^2 = \frac{1}{\lambda(P_f, P_d)}\boldsymbol{\theta}^T\mathbf{A}^T[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1}\mathbf{A}\boldsymbol{\theta}. \tag{17}$$

Also, by defining the SNR as

$$\text{SNR} = \frac{1}{\sigma^2}\boldsymbol{\theta}^T\mathbf{H}^T\mathbf{H}\boldsymbol{\theta} \tag{18}$$

and replacing the value of σ^2 with the right-hand side of (17), the relation between the parameter set $\boldsymbol{\theta}$ and the required SNR can be made explicit:[4]

[4]To give an insight into the terms in the expression for required SNR, let us denote

$$\mathbf{H}^T\mathbf{H} = \begin{bmatrix} a & \mathbf{b}^T \\ \mathbf{b} & \mathbf{C} \end{bmatrix} \tag{19}$$

where

$$a = \mathbf{h}^T\mathbf{h} \tag{20}$$

$$\mathbf{b} = [\mathbf{h}^T\mathbf{h}_{10}, \mathbf{h}^T\mathbf{h}_{01}, \mathbf{h}^T\mathbf{h}_{20}, \mathbf{h}^T\mathbf{h}_{02}, \mathbf{h}^T\mathbf{h}_{11}]^T \tag{21}$$

$$\mathbf{C} = \mathbf{A}\mathbf{H}^T\mathbf{H}\mathbf{A}^T. \tag{22}$$

Then we can show

$$[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1} = \mathbf{C} - \frac{1}{a}\mathbf{b}\mathbf{b}^T. \tag{23}$$

This form gives a rather better intuition to the established relationship between the required SNR and the resolvability. For instance it can be readily proved that the appearance of the subtracted term in the denominator of (24) ($\frac{1}{a}\mathbf{b}\mathbf{b}^T$) is due to the energy of point sources ($\alpha + \beta$) being unknown to the detector. In other words, if only the value of $\alpha + \beta$ is known to the detector, this term would vanish and the related term in the denominator would be given by $\boldsymbol{\theta}^T\mathbf{A}^T\mathbf{C}\mathbf{A}\boldsymbol{\theta}$.

TABLE I

Case where ...	Effect of P_f and P_d is given by
all the parameters are known	(33) and (34)
all the parameters except the noise variance are unknown	(14)-(16) and (24)
all the parameters including the noise variance are unknown	(29)-(31) and (32)

$$\text{SNR} = \lambda(P_f, P_d)\frac{\boldsymbol{\theta}^T\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}}{\boldsymbol{\theta}^T\mathbf{A}^T[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1}\mathbf{A}\boldsymbol{\theta}}. \tag{24}$$

The above analysis can be well extended to the case where σ^2 is unknown *a priori.* The corresponding hypotheses for this case are given by

$$\begin{cases} \mathcal{H}_0: & \mathbf{A}\boldsymbol{\theta} = \mathbf{0}, \sigma^2 > 0 \\ \mathcal{H}_1: & \mathbf{A}\boldsymbol{\theta} \neq \mathbf{0}, \sigma^2 > 0 \end{cases} \tag{25}$$

The GLRT for (25) [11, p. 345] gives the following test statistic:

$$T_u(\mathbf{g}) = \frac{\hat{\boldsymbol{\theta}}^T\mathbf{A}^T[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1}\mathbf{A}\hat{\boldsymbol{\theta}}}{\mathbf{g}^T[\mathbf{I} - \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T]\mathbf{g}} > \gamma \tag{26}$$

where subscript "u" denotes the case of unknown noise variance, $\mathbf{I}$ is the identity matrix, and $\hat{\boldsymbol{\theta}}$ is the same unconstrained maximum likelihood estimation of $\boldsymbol{\theta}$ as in (12). For any given data set $\mathbf{g}$, we decide $\mathcal{H}_1$ if the statistic exceeds a specified threshold

$$T_u(\mathbf{g}) > \gamma'. \tag{27}$$

From (11), the performance of this detector is characterized by [11, p. 186]

$$P_f = Q_{F_{5,N-6}}(\gamma') \tag{28}$$

$$P_d = Q_{F'_{5,N-6}(\lambda_u)}(\gamma') \tag{29}$$

$$\lambda_u = \frac{1}{\sigma^2}\boldsymbol{\theta}^T\mathbf{A}^T[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1}\mathbf{A}\boldsymbol{\theta} \tag{30}$$

where $Q_{F_{5,N-6}}$ is the right tail probability for a Central F distribution with five numerator degrees of freedom and $N-6$ denominator degrees of freedom; and $Q_{F'_{5,N-6}(\lambda_u)}$ is the right tail probability for a noncentral F distribution with five numerator degrees of freedom and $N-6$ denominator degrees of freedom, and noncentrality parameter λ_u.

In this GLRT context, the following relation between the parameter set $\boldsymbol{\theta}$ and the required SNR (denoted by a subscript "u") can be obtained

$$\text{SNR}_u = \lambda_u(P_f, P_d)\frac{\boldsymbol{\theta}^T\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}}{\boldsymbol{\theta}^T\mathbf{A}^T[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1}\mathbf{A}\boldsymbol{\theta}} \tag{31}$$

which mirrors (24), with the only difference in performance being the change of coefficient from $\lambda(P_f, P_d)$ to $\lambda_u(P_f, P_d)$. It can be easily verified that $\lambda(P_f, P_d) < \lambda_u(P_f, P_d)$ for $P_d > P_f$.

In either case, an important question is to consider how different these obtained performance is from that of the "ideal" clairvoyant detector, to which all the parameters (i.e., $\boldsymbol{\theta}$ and σ^2) are known. We first note that in this case the hypothesis test in (9) will be a standard linear Gauss–Gauss detection problem. Also, we can further simplify the problem by seeing that the term $(\alpha + \beta)\mathbf{h}$ in the signal model in (6) is a common known term under both hypotheses and can be removed. As a result, the following relationship can be obtained for the completely known case:

$$\text{SNR}_{\text{id}} = \eta(P_f, P_d)\frac{\boldsymbol{\theta}^T\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}}{\boldsymbol{\theta}^T\mathbf{A}^T\mathbf{A}\mathbf{H}^T\mathbf{H}\mathbf{A}^T\mathbf{A}\boldsymbol{\theta}} \tag{32}$$

where the subscript "id" denotes the ideal case and $\eta(P_f, P_d)$ is the required deflection coefficient [11, p. 71] computed as

$$\eta = (Q^{-1}(P_f) - Q^{-1}(P_d))^2 \tag{33}$$

where $Q^{-1}(\cdot)$ is the inverse of the right-tail probability function for a standard Gaussian random variable (zero mean and unit variance). Comparing the expression in (24) to that of (33)

$$\frac{\text{SNR}}{\text{SNR}_{\text{id}}} = \frac{\lambda(P_f, P_d)}{\eta(P_f, P_d)}\frac{\boldsymbol{\theta}^T\mathbf{A}^T\mathbf{A}\mathbf{H}^T\mathbf{H}\mathbf{A}^T\mathbf{A}\boldsymbol{\theta}}{\boldsymbol{\theta}^T\mathbf{A}^T[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1}\mathbf{A}\boldsymbol{\theta}} \tag{34}$$

where we note that $\eta(P_f, P_d) < \lambda(P_f, P_d)$, provided $P_d > P_f$. Also, it can be proved that $\mathbf{A}\mathbf{H}^T\mathbf{H}\mathbf{A}^T - [\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1}$ is a positive definite matrix. As a result, as expected, $\text{SNR}_u > \text{SNR} > \text{SNR}_{\text{id}}$, always. However, as we will demonstrate in one of the following sections, the difference between SNR_u and SNR_{id} is quite small over most of the parameter range. Also, in [19] we have shown that in the limiting cases (where intensities are known and equal), the proposed detector indeed produces a uniformly optimal test (same as the ideal detector). Furthermore, for the general case as we will show later, the difference between performances of the proposed detector and ideal detector is very small. Therefore, we argue that the GLRT framework used to suggest a detector can be reasonably accounted to set a performance limit in practice.

For clarity's sake, we summarize the dependence of performance on P_f and P_d for the three cases discussed above in Table I.

A. Effects of Sampling Schemes on Performance

The expressions in (24) or (31) are in general applicable to any sampling schemes including nonuniformly sampled or undersampled (aliased) images. However, since the energy of a bandlimited signal in the continuous domain is that of its uniformly and supercritically discretized version divided by the sampling rate, the right-hand side of these expressions should be understood more generally as depending upon the sampling offsets (phases) of the discrete images. In particular, for undersampled images, the energy terms in $\mathbf{H}^T\mathbf{H}$ will vary significantly as sampling phase changes. We will study this effect in Section II-A2.

An interesting related question is how the availability of multiple observations of the image (i.e., several frames with different sampling phases) affects the performance. Let Ω_l denote

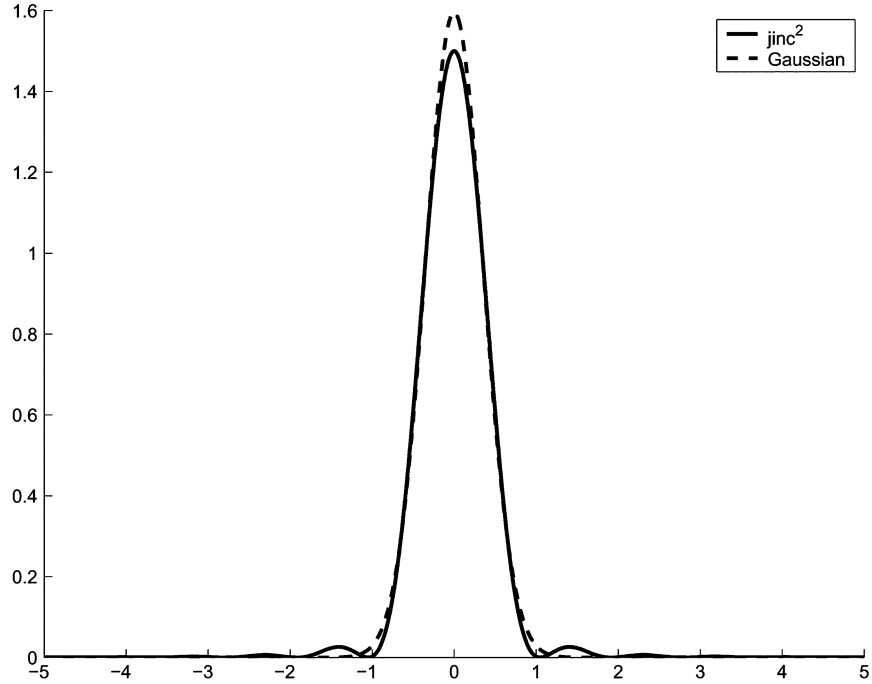

Fig. 3. Normalized 1-D cut of the point spread functions used to present the results in this paper.

the lth set of acquired samples (i.e., lth frame in a video sequence) out of a total of L frames and let $\mathbf{H}_l$ and SNR_l represent the corresponding $\mathbf{H}$ and the required SNR of the lth image. Then, the overall required SNR is given by

$$\text{SNR}_L = \lambda(P_f, P_d)\frac{\boldsymbol{\theta}^T\mathbf{H}_L^T\mathbf{H}_L\boldsymbol{\theta}}{\boldsymbol{\theta}^T\mathbf{A}^T[\mathbf{A}(\mathbf{H}_L^T\mathbf{H}_L)^{-1}\mathbf{A}^T]^{-1}\mathbf{A}\boldsymbol{\theta}} \tag{35}$$

where

$$\mathbf{H}_L = \sum_{l=1}^{L}\mathbf{H}_l. \tag{36}$$

Furthermore, it can be proved that

$$\text{SNR}_L \leq \frac{1}{L}\sum_{l=1}^{L}\text{SNR}_l \tag{37}$$

with equality sign for the case of oversampled frames. The reason behind this is that the energy of signal in over-Nyquist case is a constant value and does not depend on the phase of sampling. Hence, each (supercritically sampled) frame has the same effect on the detection performance. In other words, in this case $\mathbf{H}_m^T\mathbf{H}_m = \mathbf{H}_n^T\mathbf{H}_n = \frac{1}{L}\mathbf{H}_L^T\mathbf{H}_L$ for $1 < m, n < L$. However in under-Nyquist case, some frames (due to better placement of samples) provide more information for detectability than others.

In the following pages, we further analyze the performance results obtained earlier for oversampled images. Next we look into a case where the underlying image is undersampled to see the effect of aliasing on the performance. Having earlier derived the general expression for the performance, to facilitate the presentation, in what follows, we will often consider a particular case with the following set of conditions (we may call this the *symmetric case*) as follows:

- $p_x = p_y, q_x = q_y = 0$;
- $\alpha = \beta = 1$;
- $h(x, y) = h(\sqrt{x^2 + y^2}) = h(r)$ (angular symmetric kernel).

Some examples of angular symmetric kernels which will be used later in this paper are "jinc-squared" and Gaussian windows. The former $(\text{jinc}^2(r))$ is the PSF resulting from a single circular aperture and is characterized by

$$\begin{aligned}\text{jinc}(r) &= \frac{1}{2\pi r\sqrt{-1}}\int_0^{2\pi}\exp[\sqrt{-1}(\theta + r\cos\theta)]\mathrm{d}\theta \\ &= \frac{J_1(2\pi r)}{2\pi r}\end{aligned} \tag{38}$$

where $J_1(\cdot)$ is the first order Bessel function of the first kind [6]. The Gaussian kernel, on the other hand, can be considered as a typical approximation of the overall effect of various elements in the imaging systems (including aperture, charge-coupled device (CCD), out of focus lens, atmospheric or underwater turbulence, etc.); it is given by

$$h(r) = \frac{1}{\sqrt{2\pi\sigma_r^2}}\exp\left(-\frac{r^2}{2\sigma_r^2}\right). \tag{39}$$

A plot of the above kernels is shown in Fig. 3, where we note the corresponding Rayleigh spacings for both functions are 1 (this corresponds to $\sigma_r = 0.35$ in (39).

1) Over-Nyquist Sampling: In this section, we further simplify the earlier results for the over-Nyquist case. To begin, we note that the energy terms can be written as

$$\begin{aligned}\mathbf{P} &= \mathbf{H}^T\mathbf{H} \\ &= \begin{bmatrix} E_0 & 0 & 0 & -E_{10} & -E_{01} & 0 \\ 0 & E_{10} & 0 & 0 & 0 & 0 \\ 0 & 0 & E_{01} & 0 & 0 & 0 \\ -E_{10} & 0 & 0 & E_{20} & E_{11} & 0 \\ -E_{01} & 0 & 0 & E_{11} & E_{02} & 0 \\ 0 & 0 & 0 & 0 & 0 & E_{11} \end{bmatrix}\end{aligned} \tag{40}$$

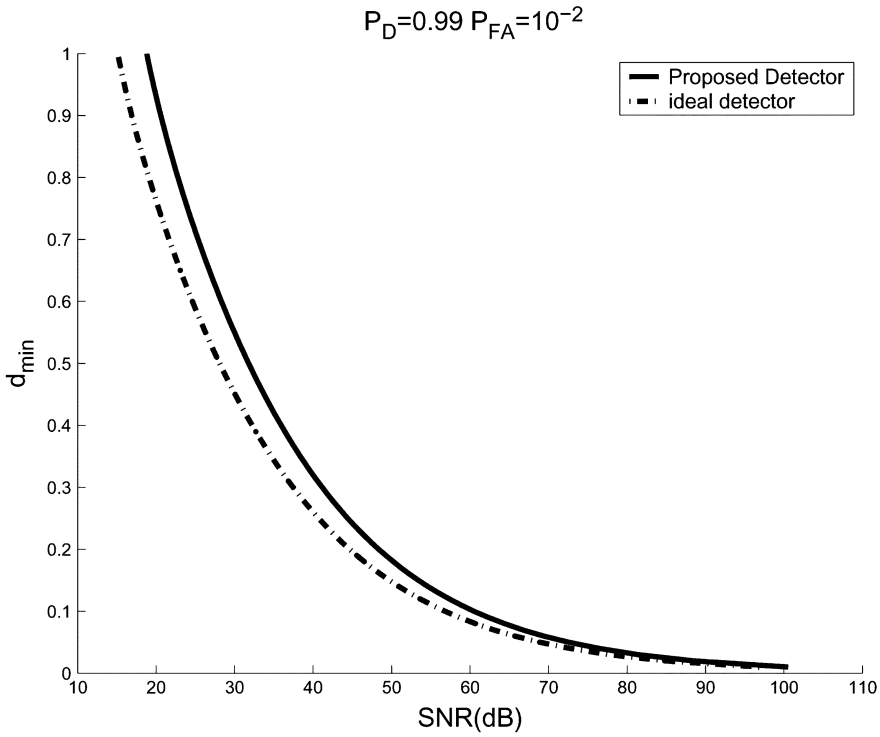

Fig. 4. Minimum detectable d as a function of SNR (in dB) (just above the Nyquist rate); Gaussian PSF; (31) and (32).

where

$$\begin{aligned} E_{ij} &= \mathbf{h}_{ij}^T \mathbf{h}_{ij} \\ &= \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} u^i v^j |H(u,v)|^2 du dv \\ &= f_s^2 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left[\frac{\partial^{i+j} h(x,y)}{\partial x^i \partial y^j} \right]^2 \mathrm{d}x \mathrm{d}y. \end{aligned} \tag{41}$$

The zero elements in (40) are due to the orthogonality of some of the partial derivatives with each other.[5] With the above notation, we will also have

$$\mathbf{Q} = \mathbf{A}^T [\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1}\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & E_{10} & 0 & 0 & 0 & 0 \\ 0 & 0 & E_{01} & 0 & 0 & 0 \\ 0 & 0 & 0 & E_{20} - \frac{E_{10}^2}{E_0} & E_{11} - \frac{E_{10}E_{01}}{E_0} & 0 \\ 0 & 0 & 0 & E_{11} - \frac{E_{10}E_{01}}{E_0} & E_{02} - \frac{E_{01}^2}{E_0} & 0 \\ 0 & 0 & 0 & 0 & 0 & E_{11} \end{bmatrix}. \tag{43}$$

[5]For instance

$$\begin{aligned} &\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x,y) \frac{\partial h(x,y)}{\partial x} \mathrm{d}x \mathrm{d}y \\ &\quad = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{2} \\ &\qquad \times \frac{\partial h^2(x,y)}{\partial x} \mathrm{d}x \mathrm{d}y \\ &\quad = \int_{-\infty}^{+\infty} \frac{1}{2} h^2(x,y) \Big|_{-\infty}^{+\infty} \mathrm{d}y = 0 \end{aligned} \tag{42}$$

since the PSF has finite energy and $\lim_{x \to \pm\infty} h(x,y) = 0$.

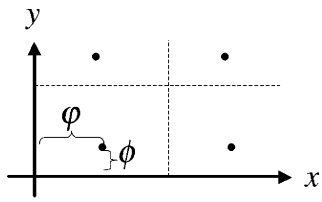

Fig. 5. Sampling phases in x and y directions (ϕ and ψ; dots indicate the locations at which signal is being sampled).

Let us now consider the symmetric case. For the oversampled case, as $d \to 0$, we will have

$$\text{SNR} \approx \frac{\lambda(P_f, P_d)}{N^2} \frac{64}{d^4} \frac{E_{10}^2}{E_0 E_{20} - E_{10}^2} \tag{44}$$

where we can show that

$$E_0 = \int_0^{+\infty} r h^2(r) \mathrm{d}r \tag{45}$$

$$E_{10} = \pi \int_0^{+\infty} r \left[\frac{\partial h(r)}{\partial r} \right]^2 \mathrm{d}r \tag{46}$$

$$E_{20} = \frac{\pi}{4} \int_0^{+\infty} \frac{3}{r} \left[\frac{\partial h(r)}{\partial r} \right]^2 + 3r \left[\frac{\partial^2 h(r)}{\partial r^2} \right]^2 \mathrm{d}r \tag{47}$$

as energy terms. See Appendix A for detailed calculation of these energy terms.[6] Fig. 4 shows the minimum detectable d versus SNR for the proposed local detector in (31) and the ideal

[6]We note that due to the Cauchy–Schwartz inequality $E_{10}^2 < E_0 E_{20}$, the right-hand side of (44) is always positive. Also, as discussed in [20], this term will vanish if the amplitude of the original scene $(\alpha + \beta)$ is known to the detector. This is to say

$$\text{SNR} \approx \frac{\lambda(P_f, P_d)}{N^2} \frac{64}{d^4} \frac{E_0}{E_2}. \tag{48}$$

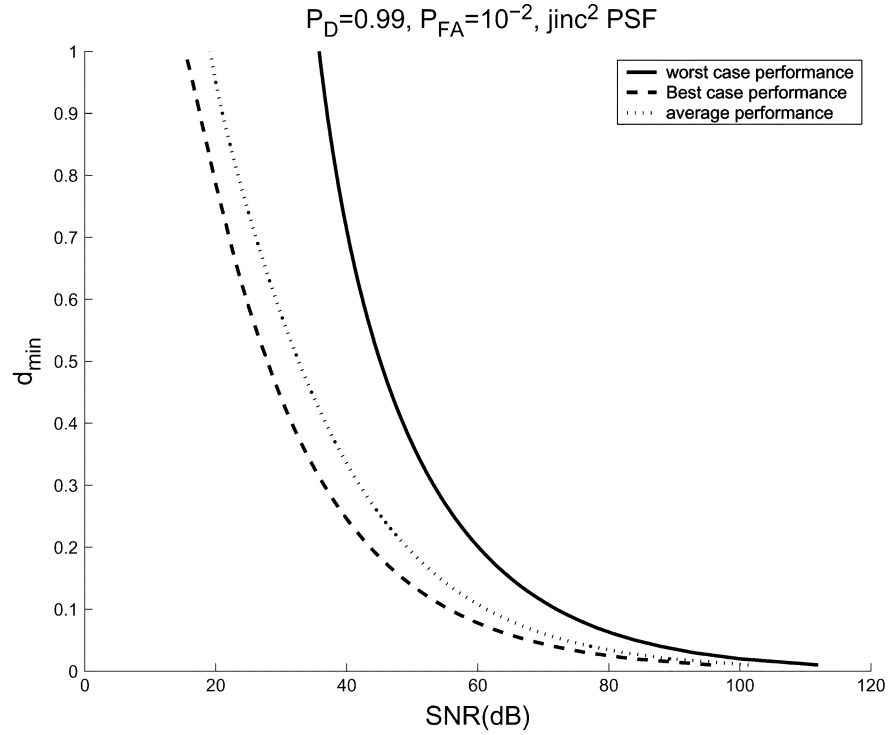

Fig. 6. Minimum detectable d as a function of SNR (in dB); best, worst, and average performance over the possible range of sampling phases (one set of uniform samples 50% below the Nyquist rate); GLRT detector; $h(r) = \mathrm{jinc}^2(r)$; known σ^2; (24).

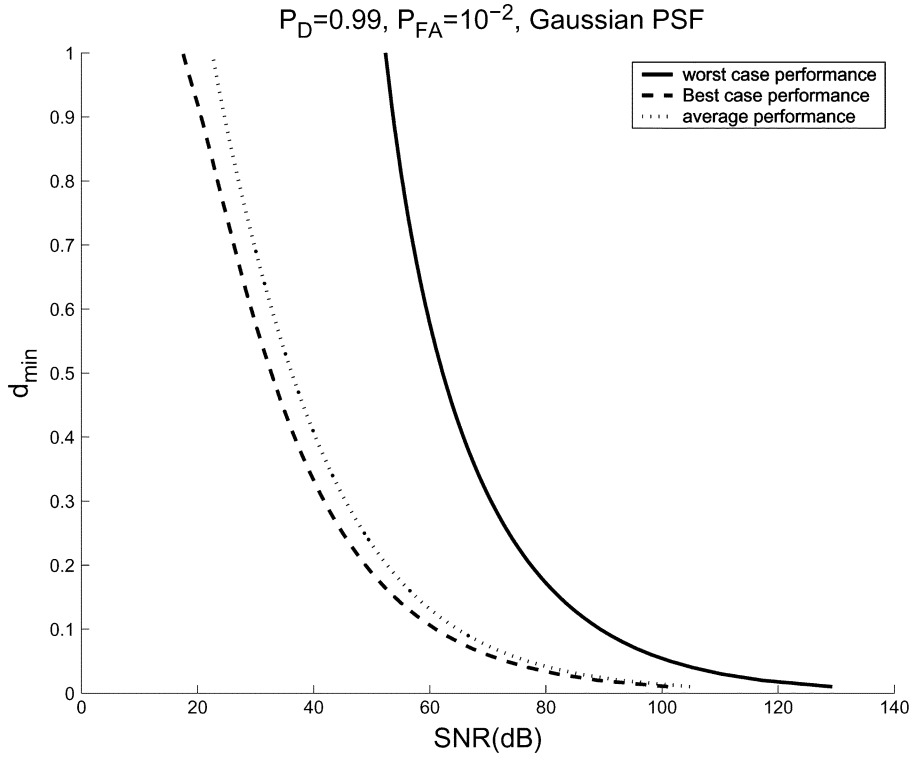

Fig. 7. Minimum detectable d as a function of SNR (in dB); best, worst, and average performance over the possible range of sampling phases (one set of uniform samples 50% below the Nyquist rate); GLRT detector; Gaussian kernel; known σ^2; (24).

detector for the case of Gaussian PSF (symmetric) and over-Nyquist sampling. The former detector has been suggested for the case where all the model parameters including the noise variance is unknown to the detector. However as can be seen, for any given d, the difference between the required SNR for these detectors is at most 3–4 dB.

Interestingly, the inverse proportionality of the fourth root of SNR and resolution (separation between point sources) was been reported earlier by our paper [19] and other researchers in different frameworks in array processing [21], [22], [13], [16] and also in optical imaging applications [8], [9], [1], [19] and references therein.

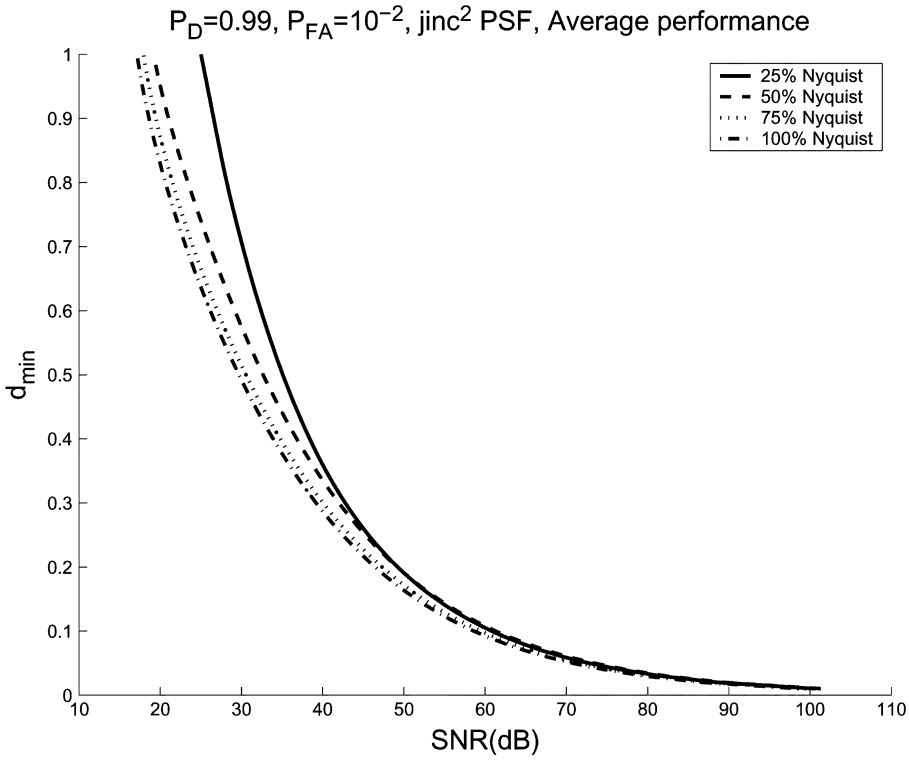

Fig. 8. Minimum detectable d as a function of SNR (in dB); average performance over the possible range of sampling phases (one set of uniform samples at different sampling rates); GLRT detector; jinc-squared kernel; known σ^2; (24).

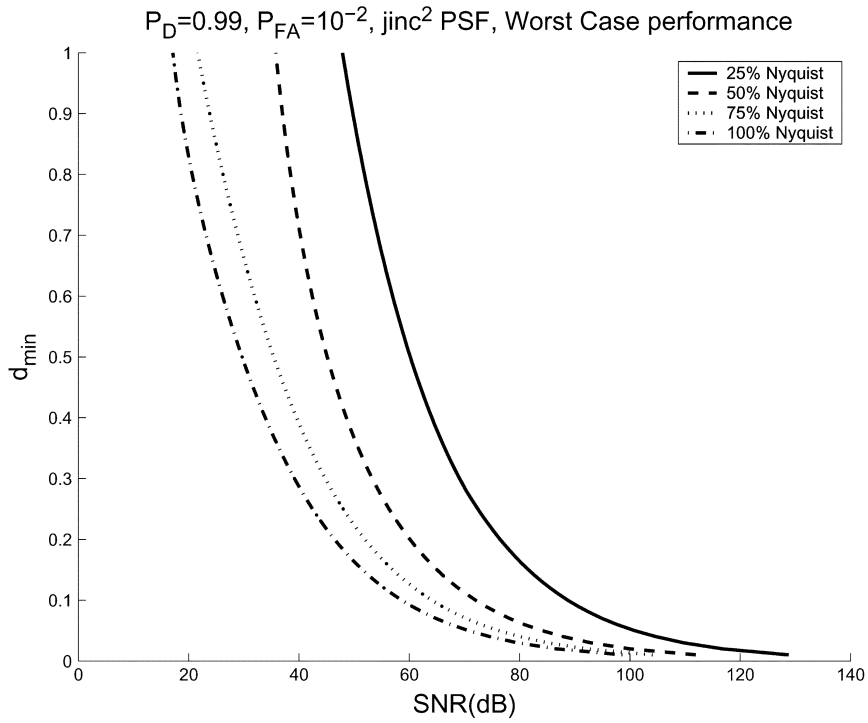

Fig. 9. Minimum detectable d as a function of SNR (in dB); worst case performance over the possible range of sampling phases (one set of uniform samples at different sampling rates); GLRT detector; jinc-squared kernel; known σ^2; (24).

2) Under-Nyquist Sampling: Supercritical sampling of a bandlimited function (e.g., $h(x,y)$ and its derivatives) preserves its energy by a factor related to the sampling rate, regardless of sampling offset. However, for the under-Nyquist (aliased) case every element of $\mathbf{H}^T\mathbf{H}$ will be a function of sampling phases in x and y directions (we call these sampling phases ϕ and ψ, see Fig. 5). This leads to a rather complicated expression for the required SNR versus d. Let us define a number of terms which will be used to describe the dependence of sampling parameters on different variables and matrices. Suppose $H(u,v)$ is band limited to $-B_u < u < B_u$ and $-B_v < v < B_v$ in the frequency domain and define $L_u = \lceil \frac{2B_u}{f_s} \rceil$ and $L_v = \lceil \frac{2B_v}{f_s} \rceil$ where $\lceil(\cdot)\rceil$ denotes the smallest integer greater than or equal to its argument. To clarify further, let $\mathbf{P}_{\text{sub}}$ be the product $\mathbf{H}^T\mathbf{H}$ for the under-Nyquist case. In general, this matrix will have the following form[7]:

$$\mathbf{P}_{\text{sub}} = \mathbf{P} + \sum_{m=0}^{L_u}\sum_{n=0}^{L_v}\left[\mathbf{P}_c^{m,n}\sum_{l=0}^{L}\cos(2\pi m\phi_l + 2\pi n\psi_l) + \mathbf{P}_s^{m,n}\sum_{l=0}^{L}\sin(2\pi m\phi_l + 2\pi n\psi_l)\right] \quad (49)$$

where $\mathbf{P}$ is defined in (40) (for the over-sampled case) and $\mathbf{P}_c^{m,n}$'s and $\mathbf{P}_s^{m,n}$'s are the matrices resulting from the aliased components.

Figs. 6 and 7 show the minimum, maximum and average values of the required SNR over the possible range of sampling phases for jinc-squared and Gaussian kernels, respectively. These curves are generated by using (24), which is the expression for the required SNR corresponding to the proposed detector in (26). For a given value of d, the required SNR is computed for the values of ϕ and ψ in the range of $[0, 2\pi]$. Then the maximum, minimum, and average of these values are computed.

The curves are shown for the case where only one set of uniformly sampled data is available. Nevertheless, it can be proved that the maximum, minimum, and average values of the resolvability (or required SNR) remain the same for arbitrary number of frames. This follows by noting that the number of frames or the sampling rate are embedded inside the "SNR" on the left-hand side of (24) and (31). For instance, for resolving a particular separation, doubling the sampling rate does not change the required SNR, but rather implies that the same detection performance can be achieved with twice the noise variance as compared to the original sampling rate.

It is seen that the required SNR for the case where the PSF is a Gaussian is on average 3 dB (and 16 dB at "worst case") more than for that for the case of simple circular aperture. This can be explained by noting that the jinc-squared kernel contains more energy in its second derivative. The more energy in the second partial derivative means bigger difference between the PDF's under the two hypotheses and therefore better detectability. Also, we note that the jinc-squared window has a larger effective bandwidth which lets more high frequency information through. This phenomenon will also be observed in the following sections.

Finally, Figs. 8 and 9 show the average and maximum resolvability at different sampling rates below Nyquist. As observed, change of sampling rate has much less effect on the average performance (i.e., the required *total* SNR) compared to that of the performance at worst case (Fig. 9). These worst cases occur when at some sampling scenarios (e.g., at 50% Nyquist), the discrete measured signal includes two strong peaks located far from each other (or roughly speaking, where the acquired samples are far from the point sources). This phenomenon clearly degrades the performance of the detector. On the other hand, if there exist some samples which are positioned closely to the point sources, the detector collects more information about the underlying signal.

[7]See Appendix B.

III. Estimation-Theoretic Approach, Fisher Information

To complement the earlier results, in this section, we carry out the Fisher Information derivations for the general signal model. Many papers have computed the Cramér–Rao bound to study the mean-square error of unbiased estimators for the distance between the point sources [8], [9], [1], [21], [22]. The result of the CRLB analysis assists us to first confirm the earlier results obtained by the detection-theoretic approach, to better understand the effect of estimation accuracy on the performance of the local detector developed in Section II, and finally to derive the KLD in Section IV, when discussing the information-theoretic framework. We present the analysis for the case where $B_u < f_s < 2B_u$ and $B_v < f_s < 2B_v$ and when two frames are measured, with sampling phases of (ϕ_1, ψ_1) and (ϕ_2, ψ_2), respectively. As such, the vector of unknown parameters of the signal model in (2) is

$$\mathbf{t} = \left[\underbrace{p_x, q_x, p_y, q_y}_{\mathbf{t}_r}, \underbrace{\alpha, \beta, \phi_2, \psi_2}_{\mathbf{t}_s}\right]^T \quad (50)$$

in which we identify two sets of parameters: parameters of interest ($\mathbf{t}_r$) and nuisance parameters ($\mathbf{t}_s$). We have assumed ϕ_1 and ψ_1 to be known, since the first frame can be considered as the reference frame.

The CRLB for the separation d (i.e., a nonlinear function of $\mathbf{t}$) is given by (51)–(52) shown at the bottom of the next page [10] where $\mathbf{\Lambda}$ is the Fisher information matrix defined by

$$\begin{aligned}[\mathbf{\Lambda}]_{ij} &= -E\left[\frac{\partial^2 \ln p(\mathbf{g},\mathbf{t})}{\partial \mathbf{t}_i \partial \mathbf{t}_j}\right] \\ &= \frac{1}{\sigma^2}\sum_k\sum_l\left[\frac{\partial}{\partial \mathbf{t}_i}\alpha h(x_k - p_x, y_l - p_y)\right. \\ &\quad +\beta h(x_k + q_x, y_l + q_y)\frac{\partial}{\partial \mathbf{t}_j}\alpha h(x_k - p_x, y_l - p_y) \\ &\quad \left.+\beta h(x_k + q_x, y_l + q_y)\right] \end{aligned} \quad (53)$$

where $p(\cdot)$ denotes the probability density function of the measured signal defined by (2). The matrix $\mathbf{\Lambda}$ can be partitioned with respect to the parameter sets $\mathbf{t}_r$ and $\mathbf{t}_s$ as

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_{rr} & \mathbf{\Lambda}_{rs} \\ \mathbf{\Lambda}_{rs} & \mathbf{\Lambda}_{ss} \end{bmatrix}. \quad (54)$$

The derivation of the Fisher information matrix for the general sampling scheme is presented in Appendix C. For the over-

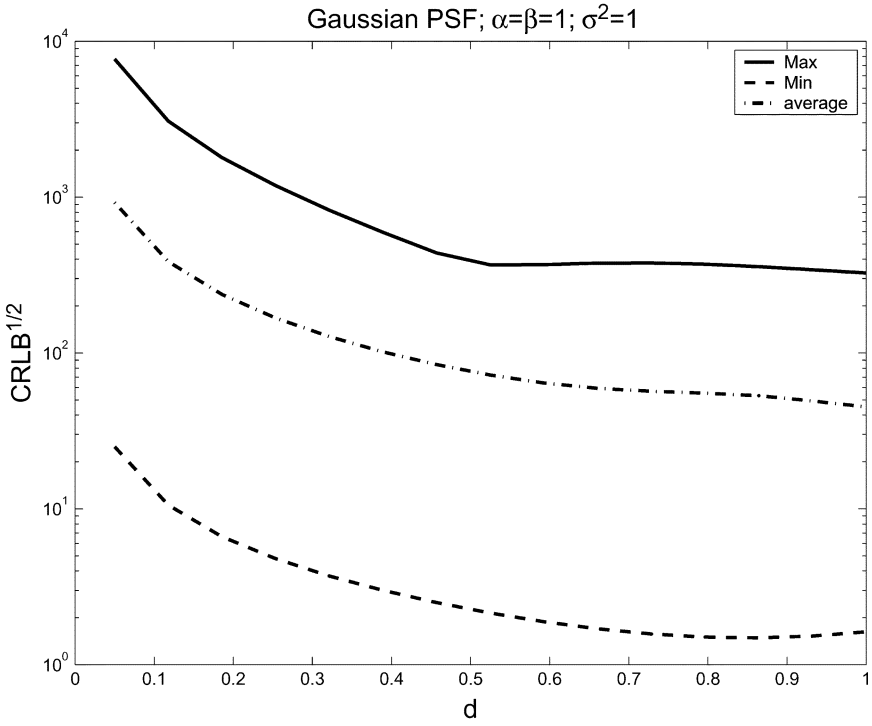

Fig. 10. $\sqrt{\text{CRLB}(d)}$ versus d; Maximum, minimum and average values over the possible range of sampling phases resulted from two sets of uniform samples 50% below the Nyquist rate (Periodically nonuniform sampling); $\alpha = \beta = 1$; Gaussian PSF; known σ^2.

Nyquist case, we note that the summations in Fisher information matrix can be simply substituted by continuous integrations. Furthermore these integrations can be rather easily computed in the frequency domain for a given point spread function (see [20] for some examples).

As for the under-Nyquist case, similar to the earlier calculation of energy terms in Section II, we can see that the Fisher Information Matrix has the following components

$$\mathbf{\Lambda}_{\text{sub}} = \mathbf{\Lambda} + \sum_{n=0}^{L_u}\sum_{m=0}^{L_v}\left[\mathbf{\Lambda}_c^{n,m}\sum_{l=0}^{L}\cos(2\pi n\phi_l + 2\pi m\psi_l) + \mathbf{\Lambda}_s^{n,m}\sum_{l=0}^{L}\sin(2\pi n\phi_l + 2\pi m\psi_l)\right] \quad (55)$$

where $\mathbf{\Lambda}$ is the Fisher information matrix provided there is no aliasing (i.e., same as what was computed for the over-Nyquist case) and $\mathbf{\Lambda}_c^{n,m}$'s and $\mathbf{\Lambda}_s^{n,m}$'s are the related matrices due to aliasing. Similar to Section II, for a given value of d, the square root of the CRLB is computed for the values ϕ and ψ in the range $[0, 2\pi]$. Then the maximum, minimum and average values are calculated to be shown in the following figures. For example, these values are displayed in Fig. 10 for a Gaussian kernel when $\alpha = \beta = 1$. As seen, the estimation task becomes much harder as d decreases.

Figs. 11 and 12 show the average of the square-root of CRLB versus d for two cases, where $\alpha = \beta$ and where $\alpha \neq \beta$. For the case where the PSF is jinc-squared, the estimation accuracy is better, due to the larger energy in the higher order derivatives of a jinc-squared function. The effect of the difference between intensities $(\alpha - \beta)$ on the estimation variance is shown in Figs. 13 and 14. These curves indicate that the estimation task is harder for the case of unequal intensities, as expected.

So far we have explored some detection-theoretic and estimation-theoretic approaches to the problem of achievable resolution. We have investigated the corresponding performance figures (required SNR for a specific resolvability and the lower

$$\text{var}(\hat{d}) \geq \begin{bmatrix} \frac{\partial d}{\partial p_x} \\ \frac{\partial d}{\partial q_x} \\ \frac{\partial d}{\partial p_y} \\ \frac{\partial d}{\partial q_y} \end{bmatrix}^T \begin{bmatrix} [\mathbf{\Lambda}^{-1}]_{11} & [\mathbf{\Lambda}^{-1}]_{12} & [\mathbf{\Lambda}^{-1}]_{13} & [\mathbf{\Lambda}^{-1}]_{14} \\ [\mathbf{\Lambda}^{-1}]_{21} & [\mathbf{\Lambda}^{-1}]_{22} & [\mathbf{\Lambda}^{-1}]_{23} & [\mathbf{\Lambda}^{-1}]_{24} \\ [\mathbf{\Lambda}^{-1}]_{31} & [\mathbf{\Lambda}^{-1}]_{32} & [\mathbf{\Lambda}^{-1}]_{33} & [\mathbf{\Lambda}^{-1}]_{34} \\ [\mathbf{\Lambda}^{-1}]_{41} & [\mathbf{\Lambda}^{-1}]_{42} & [\mathbf{\Lambda}^{-1}]_{43} & [\mathbf{\Lambda}^{-1}]_{44} \end{bmatrix} \begin{bmatrix} \frac{\partial d}{\partial p_x} \\ \frac{\partial d}{\partial q_x} \\ \frac{\partial d}{\partial p_y} \\ \frac{\partial d}{\partial q_y} \end{bmatrix} \quad (51)$$

$$= \frac{1}{d^2}\begin{bmatrix} p_x(p_x+q_x) \\ q_x(p_x+q_x) \\ p_y(p_y+q_y) \\ q_y(p_y+q_y) \end{bmatrix}^T \begin{bmatrix} [\mathbf{\Lambda}^{-1}]_{11} & [\mathbf{\Lambda}^{-1}]_{12} & [\mathbf{\Lambda}^{-1}]_{13} & [\mathbf{\Lambda}^{-1}]_{14} \\ [\mathbf{\Lambda}^{-1}]_{21} & [\mathbf{\Lambda}^{-1}]_{22} & [\mathbf{\Lambda}^{-1}]_{23} & [\mathbf{\Lambda}^{-1}]_{24} \\ [\mathbf{\Lambda}^{-1}]_{31} & [\mathbf{\Lambda}^{-1}]_{32} & [\mathbf{\Lambda}^{-1}]_{33} & [\mathbf{\Lambda}^{-1}]_{34} \\ [\mathbf{\Lambda}^{-1}]_{41} & [\mathbf{\Lambda}^{-1}]_{42} & [\mathbf{\Lambda}^{-1}]_{43} & [\mathbf{\Lambda}^{-1}]_{44} \end{bmatrix} \begin{bmatrix} p_x(p_x+q_x) \\ q_x(p_x+q_x) \\ p_y(p_y+q_y) \\ q_y(p_y+q_y) \end{bmatrix} \quad (52)$$

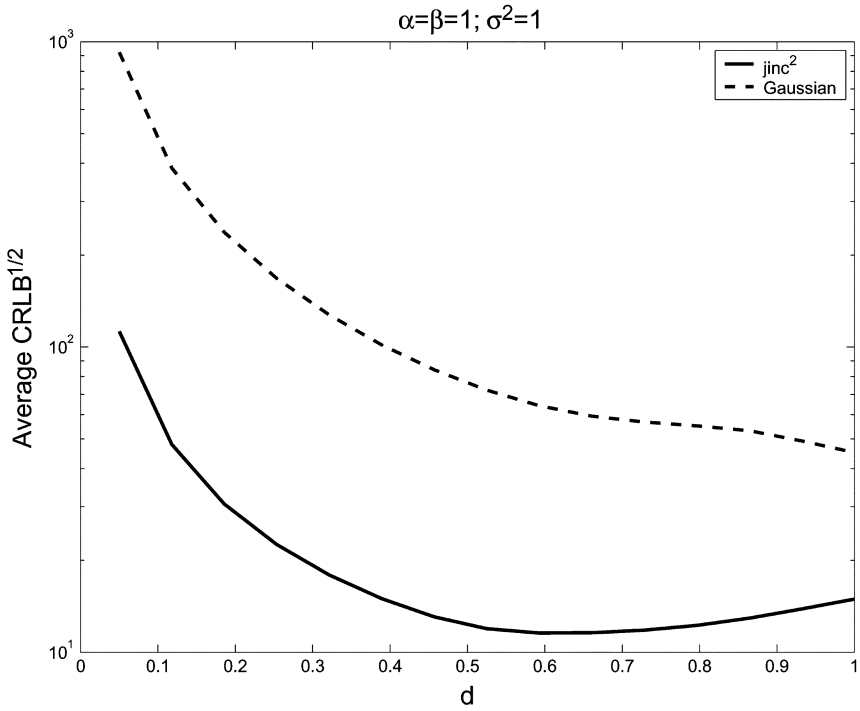

Fig. 11. $\sqrt{\mathrm{CRLB}(d)}$ versus d; average values over the possible range of sampling phases resulted from two sets of uniform samples 50% below the Nyquist rate (Periodically nonuniform sampling) for Gaussian and jinc-squared PSFs; $\alpha = \beta = 1$; known σ^2.

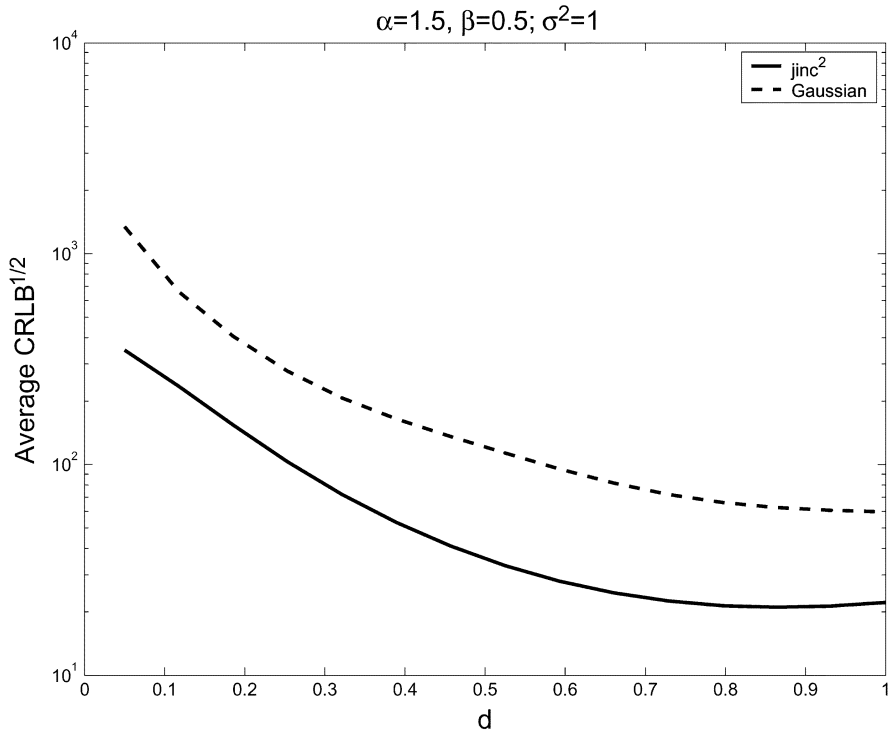

Fig. 12. $\sqrt{\mathrm{CRLB}(d)}$ versus d; average values over the possible range of sampling phases resulted from two sets of uniform samples 50% below the Nyquist rate (Periodically nonuniform sampling) for Gaussian and jinc-squared PSFs; $\alpha = 1.5, \beta = 0.5$; known σ^2.

bound on error of estimating the separation). In the following section we use (and extend) a well-known information-theoretic measure in distinguishing two hypotheses. This measure nicely links the asymptotic detection performance to the Fisher information derived in this section.

IV. Information-Theoretic Analysis, Kullback–Liebler Distance

In the interest of completeness and also verifying the earlier result from yet another perspective, we investigate the problem of the achievable resolution by an information-theoretic

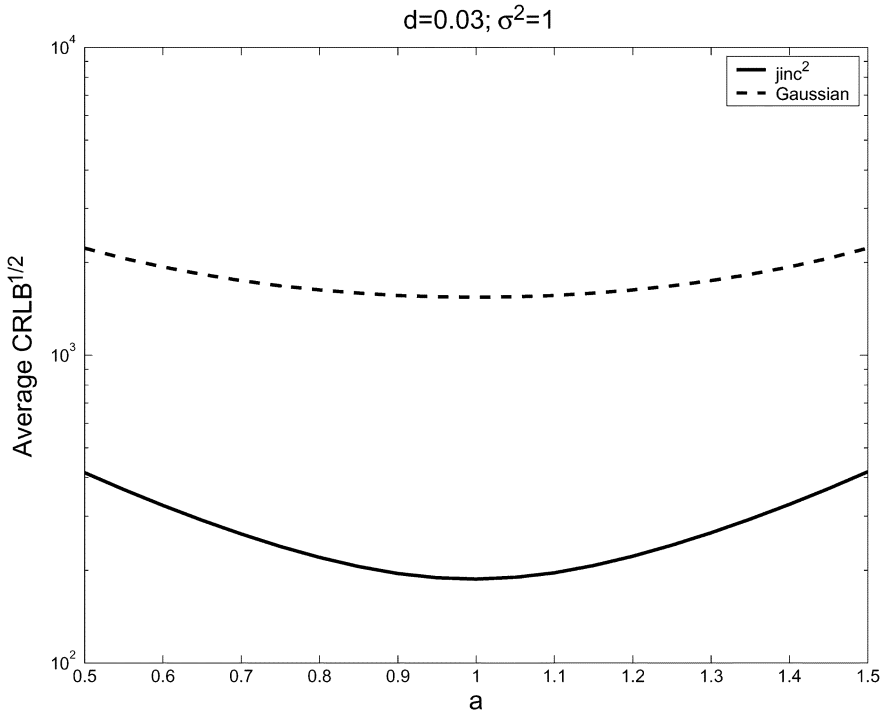

Fig. 13. $\sqrt{\text{CRLB}(d)}$ versus $\alpha(=2-\beta)$; average values over the possible range of sampling phases resulted from two sets of uniform samples 50% below the Nyquist rate (Periodically nonuniform sampling) for Gaussian and jinc-squared PSFs; $d=0.03$; known σ^2.

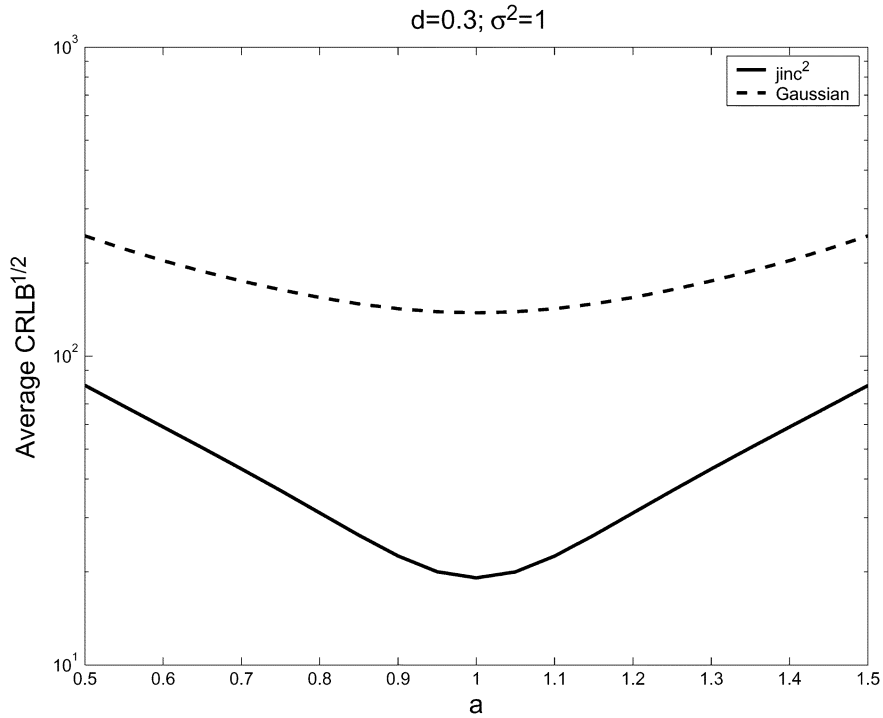

Fig. 14. $\sqrt{\text{CRLB}(d)}$ versus $\alpha(=2-\beta)$; average values over the possible range of sampling phases resulted from two sets of uniform samples 50% below the Nyquist rate (Periodically nonuniform sampling) for Gaussian and jinc-squared PSFs; $d=0.03$; known σ^2.

approach. Namely, we compute the symmetric Kullback–Liebler Distance (KLD) or Divergence $(\mathcal{J})$ [12, p. 26] for the underlying hypothesis testing problem. KLD is a measure of discrimination between two hypotheses, and can be directly related to the performance of the optimal detector. However, since KLD analysis does not take the effect of nuisance parameters[8] into account, it will indicate a somewhat loose bound on the detection performance for our problem.

Quantitatively, KLD is related to detection performance in the following way. Asymptotically, as the number of independent samples (N) grows to infinity [12]

[8]Parameters which are unknown to the detector but are common under both hypotheses.

$$P_f(1-P_d) \to \exp(-N\mathcal{J}). \tag{56}$$

To gain better insight, we first carry out an analysis for the case where the point sources are symmetric (that is to say $p_x = q_x, p_y = q_y$ and $\alpha = \beta$). Having computed this simpler case, we will extend the result to the general case. To begin, let $p(\mathbf{g}, d)$ and $p(\mathbf{g}, 0)$ be the PDF's of the measured signal under hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ in (4). Then, we will have (see Appendix D)

$$\begin{aligned}\mathcal{J}(d) &= \int_{\mathcal{D}} [p(\mathbf{g}, d) - p(\mathbf{g}, 0)] \log\left(\frac{p(\mathbf{g}, d)}{p(\mathbf{g}, 0)}\right) \mathrm{d}\mathbf{g} \qquad (57)\\ &\approx \frac{d^4}{4\sigma^2} \sum_k \sum_l \left(\frac{\partial^2 h_{02}(x_k, y_l)}{\partial d^2}\right)^2\\ &= \frac{d^4}{4\sigma^2} \mathbf{h}_{02}^T \mathbf{h}_{02} \qquad (58)\end{aligned}$$

as $d \to 0$, where $\mathcal{D}$ is the observation (signal) space. We note that the KLD measure behaves as the minimum detectable d raised to the power of 4 (confirming the power law we have derived for the inverse of the required SNR in earlier sections.).

Now, let us consider the more general model of unequal and asymmetric point sources. First, we observe from the above analysis that for the underlying problem, KLD computation requires an extension to higher order terms (see Appendix D again). To this end, we extend the (low order) formula typically used, as in for example [12, p. 26]

$$\mathcal{J}(\mathbf{t}_r) \approx \mathbf{t}_r^T \, \mathbf{\Lambda}_{rr}|_{\mathbf{t}_r = \mathbf{0}} \, \mathbf{t}_r \tag{59}$$

by a second-order approximation

$$\begin{aligned}\mathcal{J}(\mathbf{t}_r) &\approx \mathbf{t}_r^T \, \mathbf{\Lambda}_{rr}|_{\mathbf{t}_r = \mathbf{0}} \, \mathbf{t}_r\\ &\quad + \left[\mathrm{col}\left(\mathbf{t}_r \mathbf{t}_r^T\right)\right]^T \frac{\partial^2 \mathbf{\Lambda}_{rr}}{\partial \mathbf{t}_r^2}\bigg|_{\mathbf{t}_r = \mathbf{0}} \left[\mathrm{col}\left(\mathbf{t}_r \mathbf{t}_r^T\right)\right] \qquad (60)\end{aligned}$$

where $\mathrm{col}(\,\cdot\,)$ denotes the lexicographical (columnwise) scanning operator. After some algebra (60) will lead to[9]

$$\mathcal{J}(\boldsymbol{\theta}) \approx \frac{1}{\sigma^2} \boldsymbol{\theta}^T \mathbf{A}^T \mathbf{A} \mathbf{H}^T \mathbf{H} \mathbf{A}^T \mathbf{A} \boldsymbol{\theta} \tag{61}$$

This is again in general applicable to any arbitrary sampling scenario and point spread function. It is worth mentioning that the matrix $\mathbf{A}^T \mathbf{A} \mathbf{H}^T \mathbf{H} \mathbf{A}^T \mathbf{A}$ is in fact the Fisher information for the parameter set $\boldsymbol{\theta}$ in the quadratic approximated model in (6). Interestingly, the above framework shows that computing KLD in the context of [12] (that is, an approximation based on small variations of parameter(s) of interest) is in spirit similar to computing the original KLD for the approximated model. The latter approach, of course, does not require any concern about higher order terms.

For an alias-free signal (61) can be further simplified to (62) shown at the bottom of the page whereas for the aliased case, the right-hand side of (61) will depend on the sampling phases.

For undersampled image, similar to earlier analyzes, KLD varies with sampling phases

$$\begin{aligned}\mathcal{J}_{\mathrm{sub}}(\boldsymbol{\theta}) &= \mathcal{J}(\boldsymbol{\theta})\\ &\quad + \sum_{m=0}^{L_u} \sum_{n=0}^{L_v} \Bigg[\mathcal{J}_c^{m,n}(\boldsymbol{\theta}) \sum_{l=0}^{L} \cos(2\pi m \phi_l + 2\pi n \psi_l)\\ &\quad + \mathcal{J}_s^{m,n}(\boldsymbol{\theta}) \sum_{l=0}^{L} \sin(2\pi m \phi_l + 2\pi n \psi_l)\Bigg] \qquad (63)\end{aligned}$$

where $\mathcal{J}(\boldsymbol{\theta})$ is the KLD for actual over-sampled image and $\mathcal{J}_c^{m,n}(\boldsymbol{\theta})$'s and $\mathcal{J}_s^{m,n}(\boldsymbol{\theta})$'s are the related terms caused by aliasing. As sampling rate increases, the terms resulting from aliasing $(\mathcal{J}^{m,n}(\boldsymbol{\theta}) m + n > 0)$ will vanish and the expression in (62) is obtained. Fig. 15 shows variation of KLD (63) over the range of sampling phases for the symmetric case. We have used the expression in (57) to find the maximum, minimum and average values of KLD for any given separation d below the Rayleigh limit.

V. Variations in Point-Spread Function

The purpose of this section is to analyze how other parameters in real-world imaging can affect the achievable resolution. The analysis here helps us to compute the effect of small unmodeled variations in PSF or changes of PSF by other blurring functions (e.g., the effect of lens or charge coupled devices (CCD)). In this section we are interested in studying three cases. In the first case we present an analysis for an imaging system in which samples are acquired through a CCD. We study how the detection performance is affected by using such a system as compared to the idealized point sampling. For the second case, we study the effect of variations in PSF on the resolvability in a general framework. We assume that the variation in PSF is known to the detector and we derive the sensitivity of the required SNR versus the small change in PSF. The third, and perhaps most important case, is the scenario where the model (PSF) based on which we design our detector is slightly different from the actual PSF. The result of these analyzes will (for example) help quantify the importance of precisely knowing the blurring kernel on the resolving

[9]Note the difference between $\boldsymbol{\theta}$ and $\mathbf{t}_r$ which are defined in (8) and (50), respectively.

$$\mathcal{J}(\boldsymbol{\theta}) \approx \frac{1}{\sigma^2} \begin{bmatrix} -\alpha p_x + \beta q_x \\ -\alpha p_y + \beta q_y \\ \frac{1}{2}\left(\alpha p_x^2 + \beta q_x^2\right) \\ \frac{1}{2}\left(\alpha p_y^2 + \beta q_y^2\right) \\ \alpha p_x p_y + \beta q_x q_y \end{bmatrix}^T \begin{bmatrix} E_{10} & 0 & 0 & 0 & 0 \\ 0 & E_{01} & 0 & 0 & 0 \\ 0 & 0 & E_{20} & E_{11} & 0 \\ 0 & 0 & E_{11} & E_{02} & 0 \\ 0 & 0 & 0 & 0 & E_{11} \end{bmatrix} \begin{bmatrix} -\alpha p_x + \beta q_x \\ -\alpha p_y + \beta q_y \\ \frac{1}{2}\left(\alpha p_x^2 + \beta q_x^2\right) \\ \frac{1}{2}\left(\alpha p_y^2 + \beta q_y^2\right) \\ \alpha p_x p_y + \beta q_x q_y \end{bmatrix} \tag{62}$$

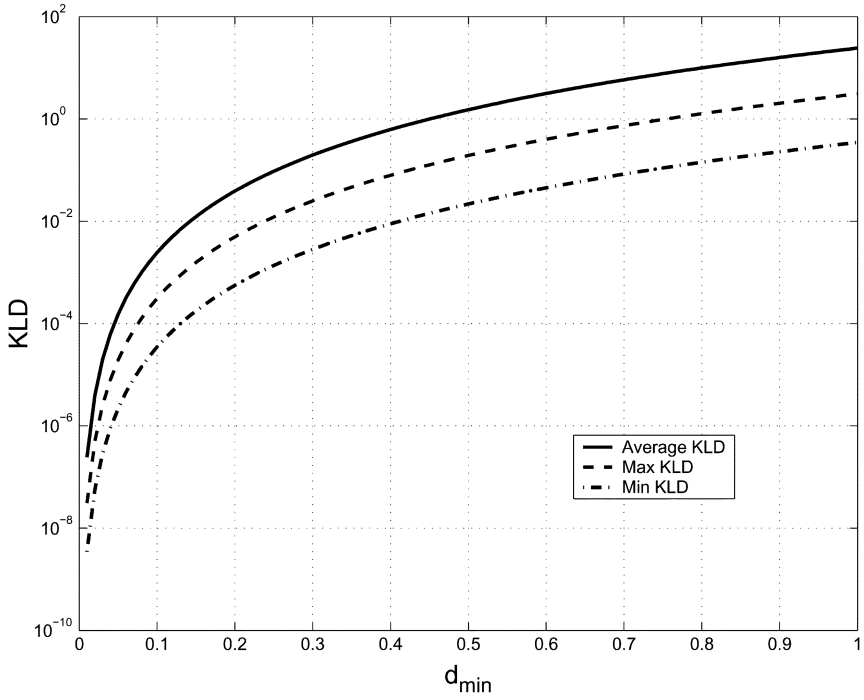

Fig. 15. Maximum, minimum and average values of KLD versus d for the possible range of sampling phases (50% below the Nyquist rate); $h(r) = \mathrm{jinc}^2(r)$; symmetric case.

power. To gain more intuition, we concentrate on alias-free images throughout this section.

A. Imaging With Spatial Integration: CCD Sampling

In real-world imaging there are other possible blurring sources which change the total PSF of the imaging system. These blurring effects are usually modeled as a space-invariant functions and can be therefore represented by a linear convolution with the assumed PSF of theimaging system. For instance, imaging with a CCD can be properly modeled in such a way. To see this effect, let us first recall that in uniform standard (point) sampling scheme, we have

$$s(x_k, y_l) = s(x,y)|_{x=k/f_s, y=l/f_s} \tag{64}$$

where f_s is the sampling frequency. On the other hand, using a CCD in image gathering will result in spatial integration of the light-field coming from a (continuous) physical scene. As an example, CCDs with square area which are uniformly sensitive to light will generate the following discretized output:

$$s_{\mathrm{ccd}}(x_k, y_l) = \frac{1}{W^2}\int_{(l-1/2)/f_s - W/2}^{(l-1/2)/f_s + W/2}\int_{(k-1/2)/f_s - W/2}^{(k-1/2)/f_s + W/2} s(x,y)\mathrm{d}x\mathrm{d}y \tag{65}$$

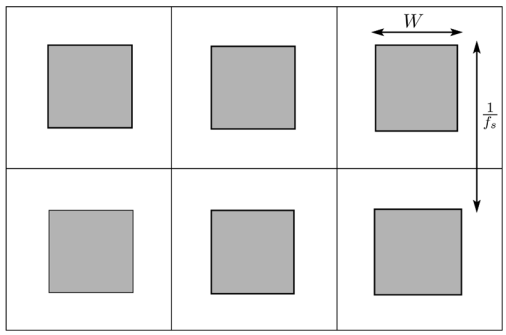

Fig. 16. Simple structure illustrating the spatial integration caused by CCD.

where W is the dimension of each CCD cell,[10] as depicted in Fig. 16. By defining

$$\mathrm{rect}(x,y) = \begin{cases} \frac{1}{W^2} & |x|, |y| \le W/2 \\ 0 & \text{otherwise} \end{cases} \tag{66}$$

as the blurring kernel of the above CCD, it can be seen that the expression in (65) directly leads to the following:

$$\begin{aligned} &s_{\mathrm{ccd}}(x_k, y_l) \\ &\quad = s(x,y) * *\mathrm{rect}(x,y)|_{x=k/f_s, y=l/f_s} &(67)\\ &\quad = I(x,y) * * \underbrace{h(x,y) * *\mathrm{rect}(x,y)}_{h_{\mathrm{ccd}}(x,y)}|_{x=k/f_s, y=l/f_s} &(68)\end{aligned}$$

where $I(x,y)$ is the original scene to be imaged, $**$ denotes the two-dimensional convolution operator and $h_{\mathrm{ccd}}(\cdot,\cdot)$ is the

[10]We ignore the effect of fill factor not being 100%.

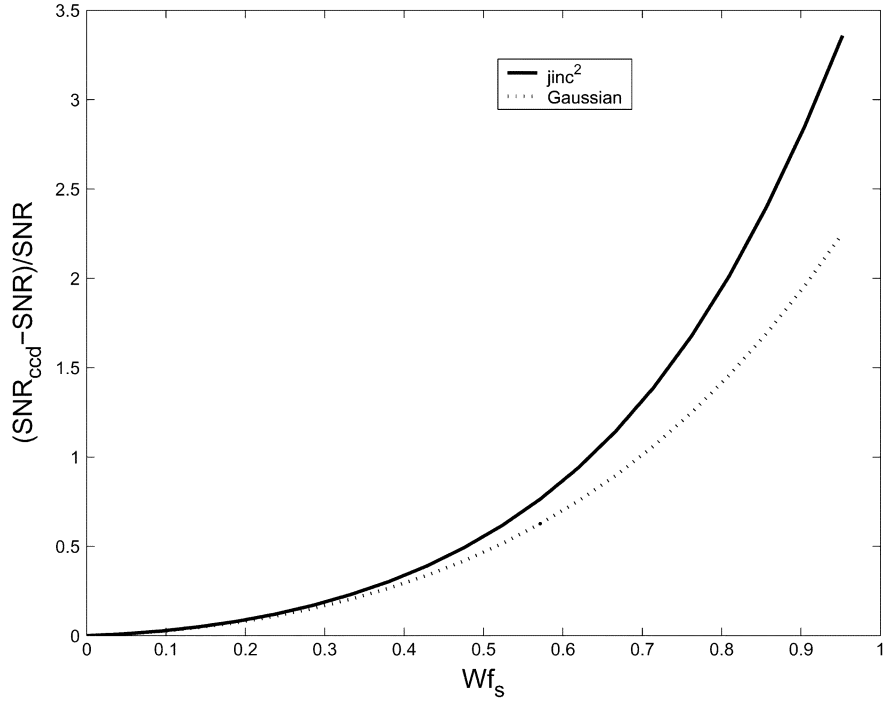

Fig. 17. Relative difference between the required SNR for CCD sampling and that of point sampling.

overall PSF. Now let $\mathrm{SNR}_{\mathrm{ccd}}$ denote the required SNR for the image collected by the above CCD sampling scheme. Fig. 17 shows the relative difference between this quantity and the required SNR for the point sampling as a function of W. In this example we have considered the symmetric and over-Nyquist case and have presented the results for both jinc-squared and Gaussian PSFs.

From a system design point of view, since typically a CCD with larger effective area has better noise characteristics (i.e., smaller σ^2), one can optimize the size of the CCD by using such a curve in effecting a trade-off with other parameters (like noise variance versus sampling rate).

In what follows, we investigate the effect of any (small) variations in the PSF on the required SNR in a general framework using perturbation analysis.

B. Variational Derivative Approach

In this section we concentrate on calculating the sensitivity of the required SNR to (known) variations in PSF. We recall from Section II that the expression for required SNR is a function of the PSF and in particular its partial derivatives (up to the second order). We use well-known techniques in calculus of variations to compute the overall variation in the required SNR.

To begin, consider the expression for SNR in (24) and suppose that the point spread function is changed by $h(x,y) \rightarrow h(x,y) + \Delta h(x,y)$. By using the concept of variational derivative [5], [24], [26], we can compute the variation in SNR (denoted by $\Delta\mathrm{SNR}(h)$) due to the variation $\Delta h(x,y)$. Namely

$$\Delta\mathrm{SNR}(h) = \frac{\lambda(P_f,P_d)}{N^2}\frac{\boldsymbol{\theta}^T\Delta\mathbf{P}\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{Q}\boldsymbol{\theta} - \boldsymbol{\theta}^T\Delta\mathbf{Q}\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{P}\boldsymbol{\theta}}{\boldsymbol{\theta}^T\mathbf{Q}\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{Q}\boldsymbol{\theta}} \tag{69}$$

$$= \frac{\lambda(P_f,P_d)}{N^2}\frac{\boldsymbol{\theta}^T(\Delta\mathbf{P}\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{Q} - \Delta\mathbf{Q}\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{P})\boldsymbol{\theta}}{\boldsymbol{\theta}^T\mathbf{Q}\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{Q}\boldsymbol{\theta}} \tag{70}$$

where $\Delta\mathrm{SNR}(h) = \mathrm{SNR}(h+\Delta h) - \mathrm{SNR}(h)$ and $\Delta\mathbf{P}$ and $\Delta\mathbf{Q}$ are perturbations of matrices $\mathbf{P}$ and $\mathbf{Q}$ defined in (40) and (43). We have detailed the derivation of $\Delta\mathbf{Q}$ and $\Delta\mathbf{P}$ in Appendix E. Following these derivations, we present the result for the case where point sources are located symmetrically , $p_x = p_y, q_x = q_y = 0, \alpha = \beta = 1$, and $h(x,y) = h(\sqrt{x^2+y^2}) = h(r)$ (angular symmetric kernel). In this case the required SNR is given by (44). Then the variation is computed as the equation shown at the bottom of the page. As a result, the relative change in SNR is given by[11]

$$\frac{\Delta\mathrm{SNR}(h)}{\mathrm{SNR}} \approx \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} A_h(x,y)\Delta h(x,y)\mathrm{d}x\mathrm{d}y \tag{71}$$

[11]See Appendix E for details.

$$\Delta\mathrm{SNR}(h) \approx \frac{\lambda(P_f,P_d)}{N^2}\times\frac{64}{d^4}\frac{2E_0\Delta E_0\left(E_0E_{20}-E_{10}^2\right) - E_0^2(E_0\Delta E_{20}+\Delta E_0E_{20}-2E_{10}\Delta E_{10})}{\left(E_0E_{20}-E_{10}^2\right)^2}.$$

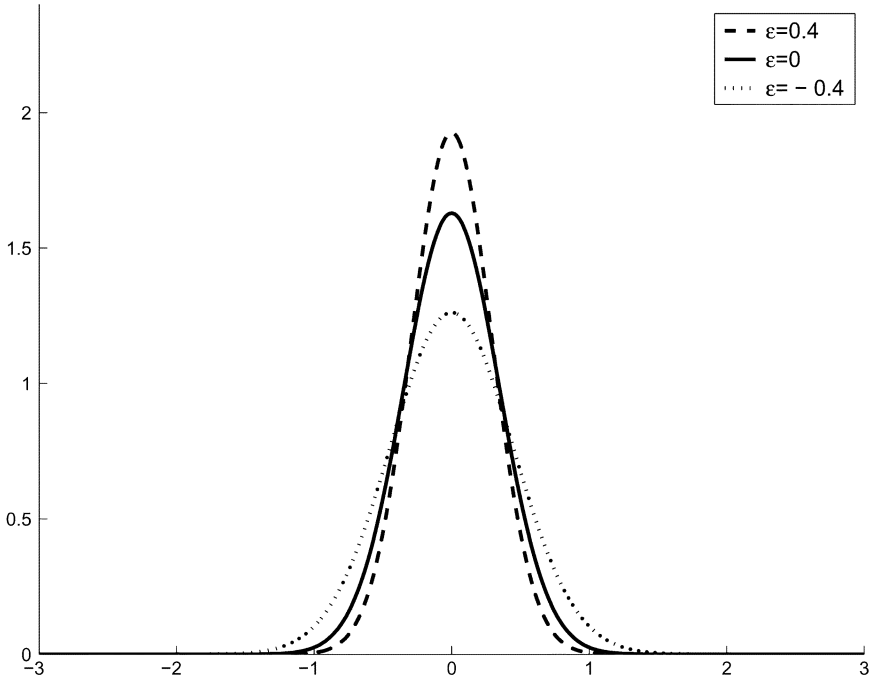

Fig. 18. 1-D cut of stretched ($\epsilon = 0.4$) and compressed ($\epsilon = -0.4$) versions of the Gaussian PSF.

where

$$A_h(x,y) = \frac{\left(E_0E_{20} - 2E_{10}^2\right)h(x,y) + 2E_0E_{10}\frac{\partial^2 h(x,y)}{\partial x^2} - E_0^2\frac{\partial^4 h(x,y)}{\partial x^4}}{E_0E_{20} - E_{10}^2}. \tag{72}$$

Invoking the Cauchy–Schwartz inequality, we have (73) shown at the bottom of the page which indicates that the relative change in SNR is bounded by the energy in the variation times a term related to energies of the PSF and its derivatives. As an example, consider the following variation which corresponds to a "stretching" ($\epsilon < 0$) or "compressing" ($\epsilon > 0$) of the PSF

$$\Delta h(x,y) = \rho(\epsilon)h([1+\epsilon]x,[1+\epsilon]y) - h(x,y) \tag{74}$$

where

$$\rho(\epsilon) = \frac{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h^2(x,y)\mathrm{d}x\mathrm{d}y}{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h^2([1+\epsilon]x,[1+\epsilon]y)\mathrm{d}x\mathrm{d}y} = \frac{E_0}{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h^2([1+\epsilon]x,[1+\epsilon]y)\mathrm{d}x\mathrm{d}y} \tag{75, 76}$$

is merely an energy normalization factor.[12] A plot of a 1-D cut of stretched and compressed versions of the Gaussian PSF is depicted in Fig. 18).

Fig. 19 shows the normalized variation in SNR versus ϵ for the jinc-squared kernel. As expected, with a narrower point spread function ($\epsilon > 0$) less SNR is required to resolve the point sources and vice versa. In fact, we can obtain a closed-form relationship for ΔSNR in this specific case assuming that the image is sampled supercritically. Let us consider the expression in (44) which includes the energy of the PSF and its partial derivatives. Now let $E_{ij}(\epsilon)$ denote the energy terms for the new kernel $\rho(\epsilon)h([1+\epsilon]x,[1+\epsilon]y)$. It can be shown that

$$E_{ij}(\epsilon) = (1+\epsilon)^{i+j}E_{ij}(0) \tag{77}$$

[12]The expression in (74) can be approximated by (see the equation at the bottom of the page.)

$$\frac{\Delta\text{SNR}(h)}{\text{SNR}} \le \sqrt{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[\Delta h(x,y)]^2\mathrm{d}x\mathrm{d}y}\sqrt{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[A_h(x,y)]^2\mathrm{d}x\mathrm{d}y}$$

or

$$\frac{\Delta\text{SNR}(h)}{\text{SNR}} \le \frac{\sqrt{E_0}\sqrt{E_0^3E_{40} - E_0^2E_{20}^2 - 4E_0^2E_{10}E_{30} + 8E_0E_{10}^2E_{20} - 4E_{10}^4}}{E_0E_{20} - E_{10}^2}\sqrt{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[\Delta h(x,y)]^2\mathrm{d}x\mathrm{d}y} \tag{73}$$

$$\Delta h(x,y) \approx \epsilon\left(-\frac{2\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}h(x,y)\left[x\frac{\partial h(x,y)}{\partial x} + y\frac{\partial h(x,y)}{\partial y}\right]\mathrm{d}x\mathrm{d}y}{E_0^2}h(x,y) + \left[x\frac{\partial h(x,y)}{\partial x} + y\frac{\partial h(x,y)}{\partial y}\right]\right).$$

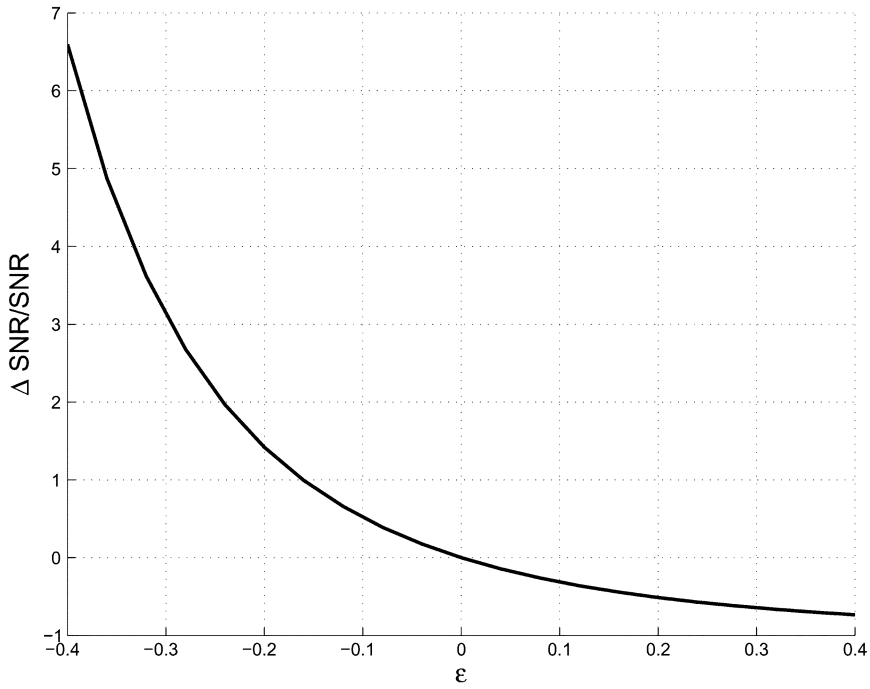

Fig. 19. Variation in the required SNR versus parameter variation in PSF.

After some algebra we have

$$\frac{\Delta\text{SNR}}{\text{SNR}} = \left(\frac{1}{(1+\epsilon)^4} - 1\right) \tag{78}$$

which holds true for any $h(x,y)$. The curve in Fig. 19 simply shows this relationship.

C. Effects of Model Mismatch on Performance

Another type of analysis is to study the case when the assumed model of the measured signal does not match the true model. This is an interesting case study which provides answers to the question of how much performance degrades due to modeling inaccuracies or mismatch.

Let us assume that the actual point spread function is $h(x,y) + \Delta h(x,y)$, whereas the optimal detector is designed for the point spread function $h(x,y)$; so that $\Delta h(x,y)$ is the mismatch (unknown to the detector) between the actual PSF and the assumed PSF. We observe that the (approximated) signal model in this case is now given by

$$\tilde{\mathbf{s}} + \Delta\tilde{\mathbf{s}} = (\mathbf{H} + \Delta\mathbf{H})\boldsymbol{\theta} = \tilde{\mathbf{s}} + \Delta\mathbf{H}\boldsymbol{\theta} \tag{79}$$

where

$$\Delta\mathbf{H} = [\Delta\mathbf{h}, \Delta\mathbf{h}_{10}, \Delta\mathbf{h}_{01}, \Delta\mathbf{h}_{20}, \Delta\mathbf{h}_{02}, \Delta\mathbf{h}_{11}].$$

The measured signal will then be $\mathbf{g} + \Delta\mathbf{H}\boldsymbol{\theta}$. Consequently, the estimate of the parameter vector will be changed to

$$\hat{\boldsymbol{\theta}} + \Delta\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{g} + \underbrace{(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\Delta\mathbf{H}\boldsymbol{\theta}}_{\mathbf{b}} \tag{80}$$

in which the second term on the right-hand side is identified as the "estimator bias." We are now able to show that $\hat{\boldsymbol{\theta}} + \Delta\hat{\boldsymbol{\theta}}$ follows a Gaussian PDF with mean $\boldsymbol{\theta} + \mathbf{b}$ and variance $\sigma^2(\mathbf{H}^T\mathbf{H})^{-1}$. Using this, we can compute the PDF of $\mathbf{A}\hat{\boldsymbol{\theta}} + \mathbf{A}\Delta\hat{\boldsymbol{\theta}}$ which is again a Gaussian pdf with mean $\mathbf{A}\boldsymbol{\theta} + \mathbf{b}$ and variance $\sigma^2\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T$. The pdf of the suggested test statistics which is compared to a prespecified threshold similar to the expression in (13)

$$\frac{1}{\sigma^2}(\hat{\boldsymbol{\theta}} + \Delta\hat{\boldsymbol{\theta}})^T\mathbf{A}^T[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1}\mathbf{A}(\hat{\boldsymbol{\theta}} + \Delta\hat{\boldsymbol{\theta}}) > \gamma \tag{81}$$

is characterized by a noncentral Chi-squared pdf under both hypotheses [11, p. 32]. To this end, we conclude that the detection performance in the presence of mismatch is characterized by

$$P_f(h + \Delta h) = Q_{\chi'^2_5(\lambda_1)}(\gamma) \tag{82}$$

$$P_d(h + \Delta h) = Q_{\chi'^2_5(\lambda_2)}(\gamma) \tag{83}$$

where

$$\lambda_1 = \frac{1}{\sigma^2}\mathbf{b}^T\mathbf{A}^T[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1}\mathbf{A}\mathbf{b} \tag{84}$$

$$\lambda_2 = \frac{1}{\sigma^2}(\boldsymbol{\theta} + \mathbf{b})^T\mathbf{A}^T[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1}\mathbf{A}(\boldsymbol{\theta} + \mathbf{b}) \tag{85}$$

are the resulting noncentrality parameters. It is also worth emphasizing here that in order to obtain λ_1 and λ_2, the value of σ^2 in the expression above needs to be computed according to the desired P_d and P_f by using the formula in (16).

Now as an example we again use (74) to compute the variation in probability of detection and false alarm rate. Hereafter, we present the results for the case where the desired detection and false alarm rates are 0.99 and 0.01, respectively. Figs. 20 and 21 show the variation in probability of error versus ϵ (which controls the stretching or compressing the PSF as in (74) for two

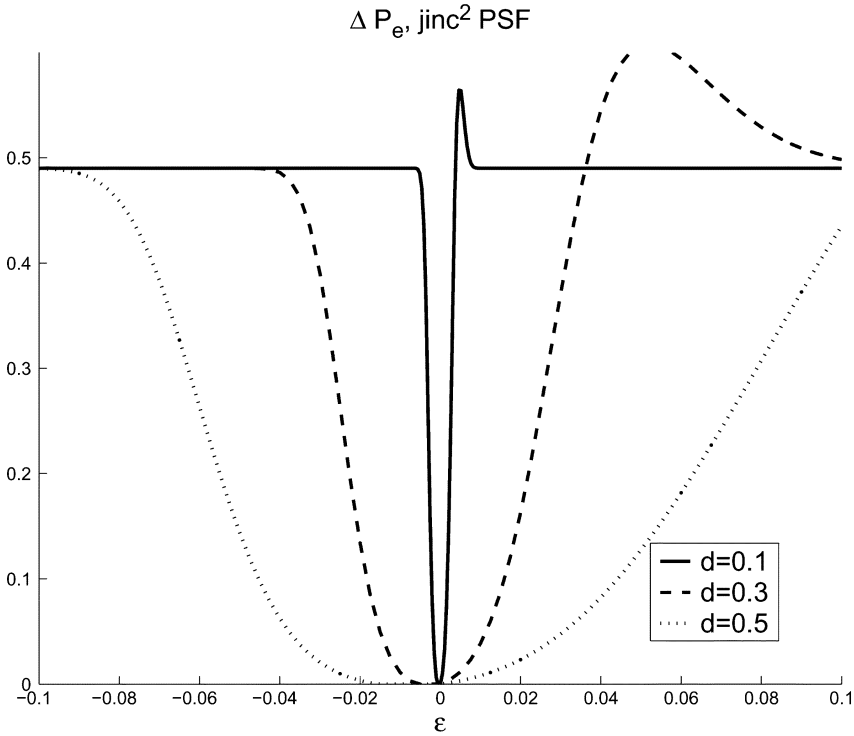

Fig. 20. Variation in the error rate versus parameter variation in PSF; $h(r) = \mathrm{jinc}^2(r)$; (88).

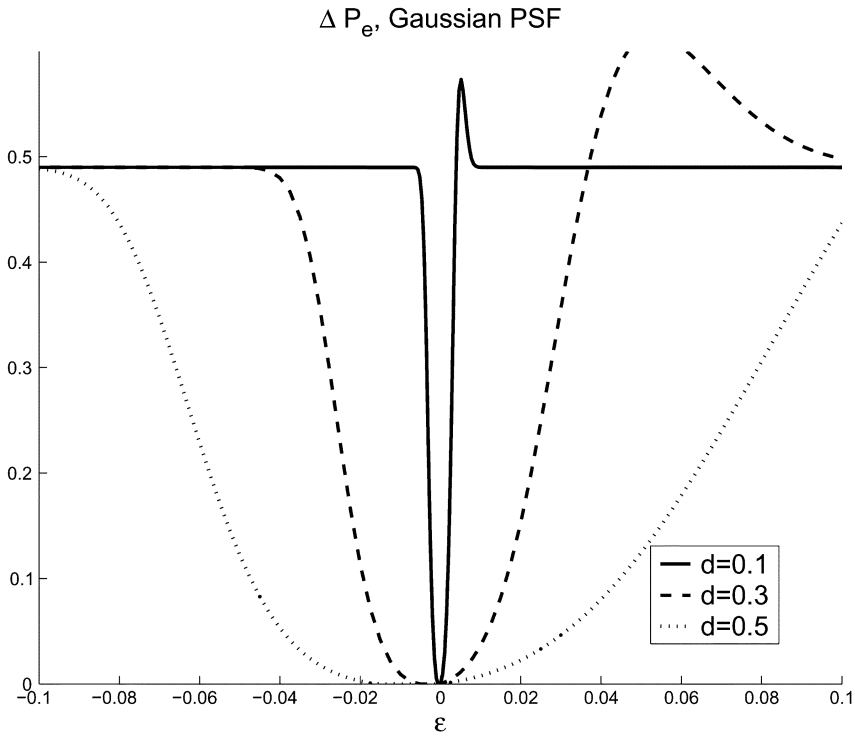

Fig. 21. Variation in the error rate versus parameter variation in PSF; Gaussian window; (88).

different kernels, each of which at different values of d. The change in the total probability of error is computed by

$$P_e(h) = \frac{1}{2}[1 - P_d(h) + P_f(h)] \tag{86}$$

$$P_e(h+\Delta h) = \frac{1}{2}[1 - P_d(h+\Delta h) + P_f(h+\Delta h)] \tag{87}$$

$$\begin{aligned}\Delta P_e &= P_e(h+\Delta h) - P_e(h) \\ &= \frac{1}{2}[\underbrace{P_f(h+\Delta h) - P_f(h)}_{\Delta P_f} \\ &\quad + \underbrace{P_d(h) - P_d(h+\Delta h)}_{-\Delta P_d}]\end{aligned} \tag{88}$$

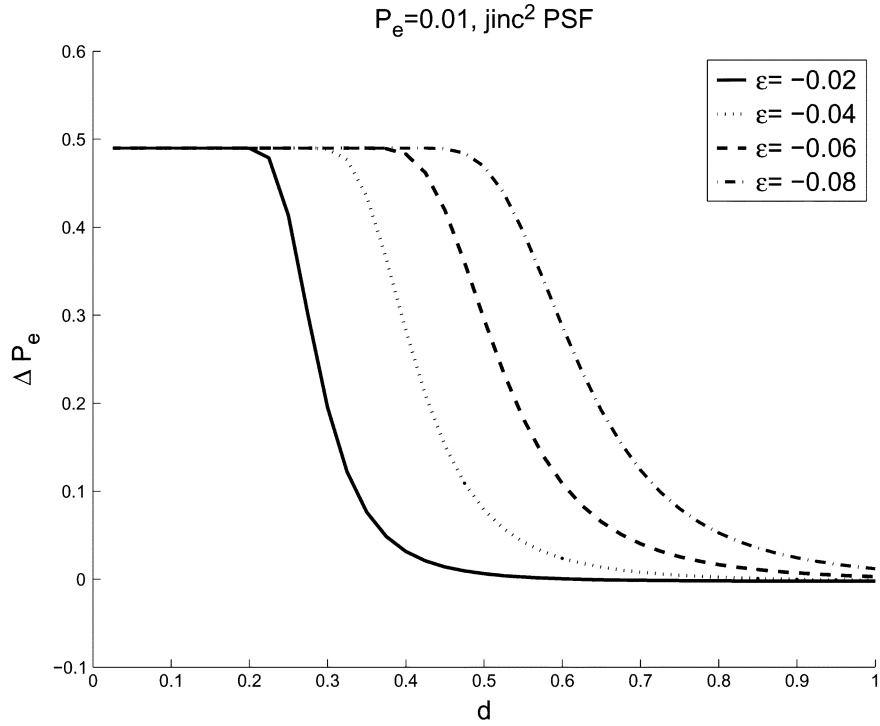

Fig. 22. Variation in the error rate versus d for fixed ϵ; jinc-squared PSF.

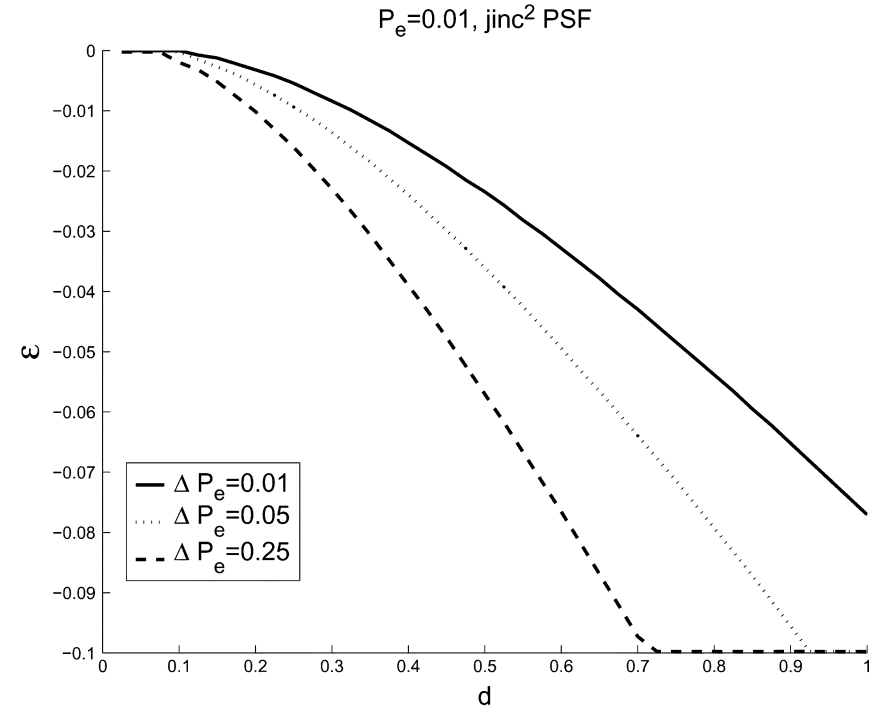

Fig. 23. The maximum tolerable ϵ versus d for the fixed probability of error; jinc-squared PSF.

Firstly we observe that the detector performance is severely affected for a smaller d (e.g., $d = 0.1$). Roughly speaking, for the range of $d < 0.3$, the proposed detector completely fails if $|\epsilon|$ exceeds $d^2/2$.

In Fig. 22 we observe how the probability of error changes as a function of d for a given ϵ (i.e., variation in PSF).[13] Also Fig. 23 depicts the variations in PSF which can be tolerated such that the probability of error is lower than a certain level. Clearly, this amount highly depends on the separation between point sources. For a small separation, even a minimal variation in PSF can cause dramatically unpleasant results. It is worth noting that the SNR used to generate these figures is the same as that required for $P_e = 0.01$ under no model mismatch.

[13]Hereafter, we only consider negative ϵ.

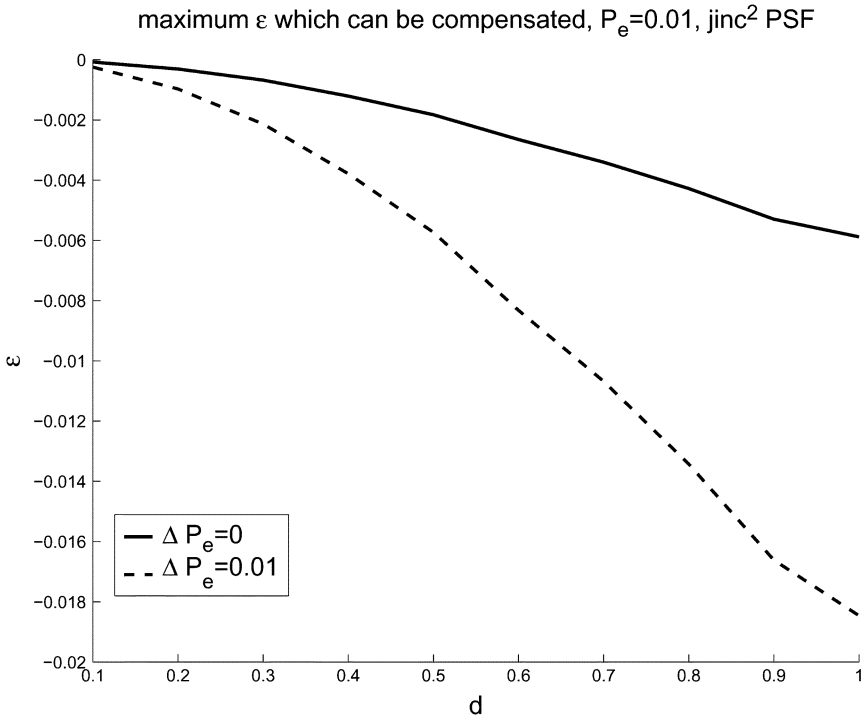

Fig. 24. Maximum tolerable ϵ which can be compensated by sufficiently increasing SNR versus given d; jinc-squared PSF.

Another interesting question in this regard would be how much extra SNR is required to compensate the error caused by a model mismatch. To answer this question let us consider the case where $P_d = 0.99$ and $P_f = 0.01$ are the desired detection and false alarm rate, respectively. To satisfy these conditions we can calculate the threshold γ which must be equal to 15.1 in (13). However, if there exists any mismatch caused by the variation in PSF, according to (84) and (85), by some calculations, we have $\lambda_1 > 33.5$ and $\lambda_2 < 0.008$ in order to achieve the error rates above. In other words, for a given $\Delta h(x, y)$, increasing the SNR can provide the desired detection accuracy only if the following inequalities hold simultaneously:

$$\frac{1}{\sigma^2}\mathbf{b}^T\mathbf{A}^T[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1}\mathbf{A}\mathbf{b} < 0.008 \tag{89}$$

$$\frac{1}{\sigma^2}(\boldsymbol{\theta}+\mathbf{b})^T\mathbf{A}^T[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1}\mathbf{A}(\boldsymbol{\theta}+\mathbf{b}) > 33.5 \tag{90}$$

We note that the first inequality enforces the pdf of the test statistic under $\mathcal{H}_0$ to have smaller noncentrality parameter so that the required P_d is accessible, whereas the second inequality plays the reverse role for the pdf under $\mathcal{H}_1$. We can unify the inequalities in (89) and (90) as

$$\frac{(\boldsymbol{\theta}+\mathbf{b})^T\mathbf{A}^T[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1}\mathbf{A}(\boldsymbol{\theta}+\mathbf{b})}{\mathbf{b}^T\mathbf{A}^T[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1}\mathbf{A}\mathbf{b}} > 4187.5. \tag{91}$$

Under the above condition a sufficiently high value of SNR can compensate the effect of model mismatch. To be more realistic let us also carry out the analysis for the case where we allow the probability of error to be equal to 0.02 (i.e., $P_d = 0.98$ and $P_f = 0.02$). The threshold γ remains the same. However, to satisfy the new conditions we only need $\lambda > 30.5$ and $\lambda < 0.66$. In other words it is possible to have an error rate less than 0.02 if

$$\frac{(\boldsymbol{\theta}+\mathbf{b})^T\mathbf{A}^T[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1}\mathbf{A}(\boldsymbol{\theta}+\mathbf{b})}{\mathbf{b}^T\mathbf{A}^T[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1}\mathbf{A}\mathbf{b}} > 452.23. \tag{92}$$

As an illustration of the above analysis, let us consider the case where PSF undergoes the same effect as in (74). Fig. 24 shows the amount of mismatch which can be compensated by presumably high SNR (theoretically as $\mathrm{SNR} \to \infty$) for a given distance d. We demonstrate two cases: the case where no extra error can be tolerated ($\Delta P_e = 0$) and the case where we allow $\Delta P_e = 0.01$. The results clearly indicate that specially for small distance between point sources the detector is extremely sensitive to (unknown) variation of PSF.

VI. Practical Aspects and Extensions

In this section, we briefly discuss some future extensions and practical aspects of the analyzes developed in the paper.

A. Application in Astronomical Imaging and Image Processing: A Primary Case Study

A complete extension of the current work to a real problem in astronomy requires a careful study and is beyond the scope of this paper. Typically, the question of resolvability in astronomy introduces new aspects to the problem such as presence of other interfering phenomena, objects, clutter, non-Gaussian noise and also time-varying PSF. Each of these issues should be carefully treated and incorporated in the analysis. As an example, we have addressed the issue of variation and mismatches in PSF in Section V.

Fig. 25. Original image.

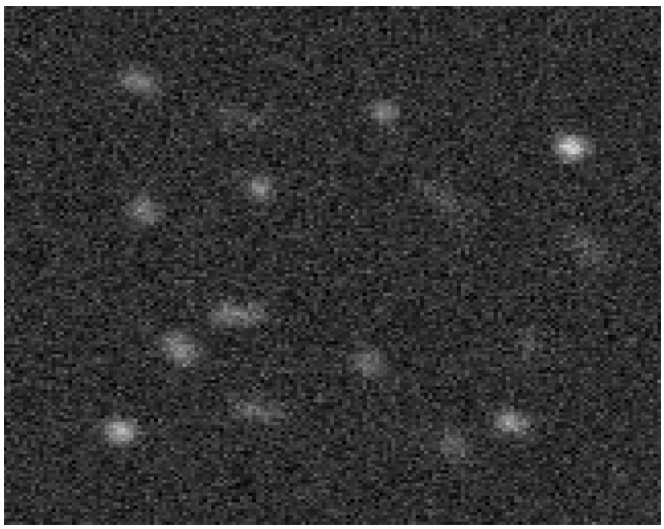

Fig. 26. Measured image.

Fig. 27. Result of applying CLEAN method to the image in Fig. 26.

Fig. 28. The result of modified CLEAN (CLEAN associated with the local detectors).

In this section we present a basic application of the proposed detector to a related but slightly simplified problem. The underlying image to be restored is a collection of point sources (or more precisely pairs of closely spaced point sources) and the collected data is a discrete noisy and blurred version of it as shown respectively in Figs. 25 and 26.

In the astronomical data analysis community, a well-known deconvolution method named CLEAN is widely used [23]. CLEAN strictly assumes that the underlying image consists of point sources and estimates their locations and intensities. Furthermore, the PSF is assumed to be known. It first finds the point which has the maximum brightness (x_0, y_0) and subtracts the effect of the (estimated) point sources (i.e., the convolution of $\delta(x - x_0, y - y_0)$ and the PSF). Then the resulted signal is iteratively used to find and remove other point sources until the energy of the residual signal reaches a prespecified noise level. The result of applying this method to the image in Fig. 25 is demonstrated in Fig. 27. Obviously, the use of CLEAN is limited by the width of PSF, since any pair of point sources which are spaced within a distance related to the width of the PSF is identified as a single point source.

We argue that the resolvability of CLEAN can be significantly improved by employing the local detector we have proposed in this article. We include our local detection machinery in this approach in the following way. First instead of searching for the point with the maximum brightness, we find the point with the largest intensity after convolving the data with PSF.[14] This approach will help us to locate the midpoint of the pairs of point sources within a properly chosen window. Once this point is identified, one can conveniently set up the local detector to determine the number of point sources in the neighborhood of the point and also estimate the distance between point sources and their intensities.

A basic implementation of this approach was carried out and the result of applying it to Fig. 25 is shown in Fig. 28, which closely follows the structure of the original data. To have a better

[14]A quantitative proof of the optimality of this approach is given in [20], [18].

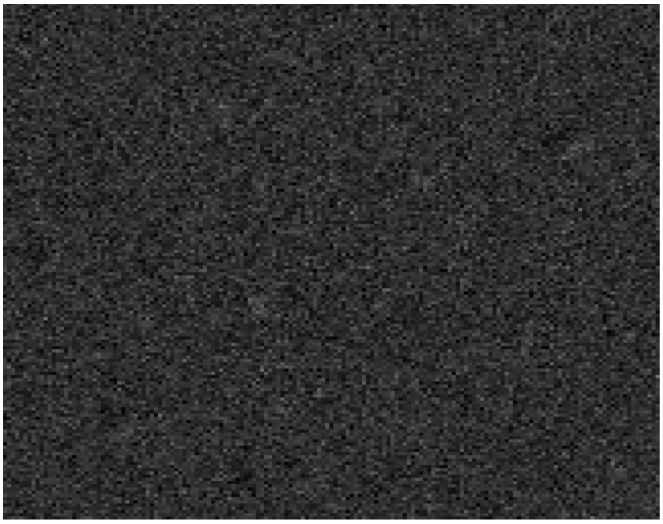

Fig. 29. Residual: difference between the measured image and convolution of Fig. 28 and PSF.

indication of the claimed improvement, see the image in Fig. 29 which is the difference between the collected image (Fig. 26) and the convolution of the output image (Fig. 28) with the PSF.

B. An Application to Physical Fault Detection in Integrate Circuit Manufacturing

Integrated circuit manufacturing processes require transferring a circuit pattern (including logic gates, memory cells, etc.) onto the silicon wafers. This process is referred to as lithography. However, because of several factors (like irregularity of surface or mask imperfection) the above process may introduce some errors in replicating the pattern. These errors translate into short or open circuits which make the manufactured IC useless. Hence, there is a need for metrology and inspection to detect these physical defects or faults. The commonly employed method is to image the printed wafer using an SEM (Scanning Electron Microscope) and to inspect this image by matching with the desired, expected pattern. Over time, the dimensions of the printed circuits are becoming increasingly small (currently 65 nm), and this, coupled with the resolution limits of the SEM, lead to a heavily blurred and noisy version of the desired image of the wafer. We have briefly explained how to employ and extend the proposed detection frameworks to such inspection tasks in [17]. However the developed detector evaluates the correctness of a circuit or mask in a simple way and for comprehensive and practical purposes the method we propose needs to be generalized. Such generalization requires better understanding of the physical models of the underlying systems.

VII. Conclusion

In this paper we investigated the problem of resolving two closely spaced point sources (beyond the Rayleigh limit) from noise-corrupted, blurred, and possibly undersampled images. We have studied three different frameworks to derive performance limits on minimum resolvability, namely, a detection-theoretic approach, an estimation-theoretic analysis, and an information-theoretic approach. We have discussed the case of under-Nyquist sampling in several parts of this paper because of its significance in, for example, image super-resolution reconstruction [3], [15], [4]. To our knowledge, this paper presents the first direct analysis of resolution limits in imaging for undersampled images.

For the detection-theoretic approach, we have put forward a hypothesis testing framework and derived the relationship between the required SNR and the detectable distance between point sources. We explicitly found a general, fundamental, and informative relationship to quantify a (statistical) measure of resolution, and also to reveal the effect of point spread function and sampling parameters on resolvability. The established expressions are in general applicable to any point spread function and any arbitrary sampling scheme. The analysis for the undersampled images demonstrates explicitly how the performance is affected by the sampling phases (sampling offset), while in the over-Nyquist case the performance is independent from the sampling offset.

In the process of addressing the fundamental relationship between SNR and resolvability, we also proposed a corresponding detector which can be implemented in practice to estimate the locations and intensities of point sources from a degraded (noisy and blurred) image. As an example of implementing such a technique, one can first apply a running windowed energy detector to determine whether a pattern (generated by a point source) within the window is present or not. Then as a refinement step, the proposed detector will be applied to resolve the overlapping patterns originated by two closely-spaced point sources.

We studied the asymptotic performance of the ML estimate of the unknown parameters, using the Fisher information matrix. Deriving a lower bound on the variance of the parameter d, in particular, was rather helpful in confirming the results of detection-theoretic analysis and also justifying the effect of estimation accuracy on the performance of the proposed detector. We also derived the symmetric Kullback-Liebler distance for the underlying problem by extending its standard form to higher order terms. This analysis provides an upper bound on the detection performance we have derived in Section II and also connects the Fisher information matrix with this performance bound.

As a practical matter, we presented a sensitivity/variational analysis to find the effect of known or unknown variation of PSF on the performance of the detector discussed in Section II. The results of this part are useful to compute the effect of a known variation of PSF on performance of an imaging system and to quantify the degradation of such a system due to imperfect knowledge of the model of the imaging system.

Furthermore, we have briefly discussed some of the practical uses of the proposed methodology to other fields such as astronomical data analysis tasks and circuit inspection applications.

As for future directions, one open area is to extend this framework to a more complicated scenario where the PSF is partially known or is associated with some stochastic parameters. Another direction is to study the case of unknown motion between frames. Also, as mentioned before, the problem of resolving point sources is a canonical case study in imaging and image processing. We believe more general imaging problems can be treated with the same techniques employed in this paper, with proper extension (and complication) of the imaging models.

Resolution limit in photon-limited imaging systems is another important case study that needs to be investigated. The

noise in these systems is no longer additive readout noise and follows the Poisson process. Also, in many cases the underlying image itself may contain different forms of noise and in particular clutter or unwanted interference which are also described by a probabilistic function.

Another challenging direction is to investigate the performance of the imaging systems with time-varying characteristics (for example, time-varying PSF particularly in astronomical applications). Such behavior may be modeled by some probabilistic descriptions as well. The analysis here requires a somewhat more sophisticated machinery.

We have briefly studied the resolution problem for the case where the measured signal is contaminated by Poisson noise (for example in photon-limited imaging) in [17]. This is an important case study in astronomical applications and SEM imaging.

Appendix A
Angular Symmetric PSF

Let us consider the case of angular symmetric PSF ($h(x,y) = q(\sqrt{x^2+y^2}) = q(r)$) such as that corresponding to circular aperture. Following relationships for the partial derivatives can be written in the polar form (r,θ):

$$\frac{\partial h(x,y)}{\partial x} = \frac{\partial q(r)}{\partial r}\cos(\theta)$$
$$\frac{\partial h(x,y)}{\partial y} = \frac{\partial q(r)}{\partial r}\sin(\theta)$$
$$\frac{\partial^2 h(x,y)}{\partial x^2} = \frac{\partial^2 q(r)}{\partial r^2}\cos^2(\theta) + \frac{\partial q(r)}{\partial r}\frac{\sin^2(\theta)}{r}$$
$$\frac{\partial^2 h(x,y)}{\partial y^2} = \frac{\partial^2 q(r)}{\partial r^2}\sin^2(\theta) + \frac{\partial q(r)}{\partial r}\frac{\cos^2(\theta)}{r}$$
$$\frac{\partial^2 h(x,y)}{\partial y \partial x} = \frac{\partial^2 q(r)}{\partial r^2}\sin(\theta)\cos(\theta) - \frac{\partial q(r)}{\partial r}\frac{\sin(\theta)\cos(\theta)}{r}.$$

After doing some math and by applying

$$\int_0^{2\pi}\sin(\theta)\cos(\theta)\mathrm{d}\theta = \int_0^{2\pi}\sin(\theta)\cos^3(\theta)\mathrm{d}\theta = 0$$
$$\int_0^{2\pi}\sin^2(\theta)\mathrm{d}\theta = \int_0^{2\pi}\cos^2(\theta)\mathrm{d}\theta = \pi$$
$$\int_0^{2\pi}\sin^4(\theta)\mathrm{d}\theta = \int_0^{2\pi}\cos^4(\theta)\mathrm{d}\theta = \frac{3\pi}{4}$$
$$\int_0^{2\pi}\sin^2(\theta)\cos^2(\theta)\mathrm{d}\theta = \frac{\pi}{4}.$$

We can show that

$$E_0 = \int_0^{+\infty} r q^2(r)\mathrm{d}r \tag{93}$$

$$E_{10} = E_{01} = \pi\int_0^{+\infty} r\left[\frac{\partial q(r)}{\partial r}\right]^2\mathrm{d}r \tag{94}$$

$$E_{20} = E_{02} = \frac{\pi}{4}\int_0^{+\infty}\frac{3}{r}\left[\frac{\partial q(r)}{\partial r}\right]^2 + 3r\left[\frac{\partial^2 q(r)}{\partial r^2}\right]^2\mathrm{d}r \tag{95}$$

$$E_{11} = \frac{\pi}{4}\int_0^{+\infty}\frac{1}{r}\left[\frac{\partial q(r)}{\partial r}\right]^2 + r\left[\frac{\partial^2 q(r)}{\partial r^2}\right]^2\mathrm{d}r \tag{96}$$

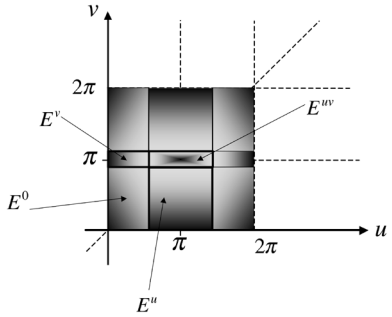

Fig. 30. E^0, E^u, E^v, and E^{uv} are the energy of the discrete signals within depicted areas.

by using

$$\int_0^{+\infty}\frac{\partial q(r)}{\partial r}\frac{\partial^2 q(r)}{\partial r^2}\mathrm{d}r = 0.$$

Therefore

$$\mathrm{SNR} = \frac{\lambda(P_f,P_d)}{N^2}\frac{64E_0 - 16d^2E_{10} + d^4E_{20}}{d^4\left(E_{20} - \frac{E_{10}^2}{E_0}\right)} \tag{97}$$

Appendix B
Sampling Theory for Under-Nyquist Measurements

To begin, let us first mention that for the case of over-Nyquist sampling, following relationships are hold:

$$E_{ij} = \mathbf{h}_{ij}^T\mathbf{h}_{ij} = \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}u^i v^j|H(u,v)|^2\mathrm{d}u\mathrm{d}v \tag{98}$$

$$= f_s^2\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left[\frac{\partial^{i+j}p(x,y)}{\partial x^i\partial y^j}\right]^2\mathrm{d}x\mathrm{d}y. \tag{99}$$

Whereas in the under-Nyquist case, assuming that $H(u,v)$ is symmetric (real $h(x,y)$) and band limited to $-B_u < u < B_u$ and $-B_v < v < B_v$, we will have (100)–(103) shown at the bottom of the next page where $\Re$ denote the real part.[15] For instance, if $B_u < f_s < 2B_u$ and $B_v < f_s < 2B_v$ we will similarly have

$$E_i = E_i^0 + E_i^v\cos(2\pi\phi) + E_i^u\cos(2\pi\psi) + E_i^{uv}\cos(2\pi\phi + 2\pi\psi) \tag{105}$$

where E_i^0 denote the energy of signal considering there is no aliasing effect and E_i^u, E_i^v and E_i^{uv} are the terms related to aliasing in horizontal, vertical and diagonal directions as depicted in Fig. 30.

It is worth mentioning that for the general energy terms $(E_{ij} i \neq j)$ we will have $\sin(\cdot)$ terms as well

$$E_{ij} = E_{ij}^0 + \Re\left[E_{ij}^v e^{\sqrt{-1}2\pi\phi} + E_{ij}^u e^{\sqrt{-1}2\pi\psi} + E_{ij}^{uv}e^{\sqrt{-1}2\pi\phi + \sqrt{-1}2\pi\psi}\right]. \tag{106}$$

[15]We have used the following identity to derive the above results

$$\left|\sum_i a_i\right|^2 = \sum_i |a_i|^2 + 2\sum_i\sum_{i'>i}\Re\{a_i a_{i'}\} \tag{104}$$

APPENDIX C
DERIVATION OF THE FISHER INFORMATION MATRIX

Let Ω_1 and Ω_2 denote the sets of samples of the first and the second frames, respectively. From (53), the Fisher Information matrix can be written as

$$\mathbf{\Lambda} = \mathbf{Z}^T\mathbf{Z} \tag{107}$$

where $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \ldots, \mathbf{z}_8]$ is defined in (108)–(115) shown at the bottom of the next page.

APPENDIX D
COMPUTING THE KULLBACK–LIEBLER DISTANCE IN (57)

Directly using the results in [12, p. 26], we can obtain the following expression for KLD:

$$\mathcal{J}(d) \approx d^2 I(0) \tag{116}$$

where $I(d)$ is the Fisher Information measure [10, p 40].

$$I(d) = \frac{1}{\sigma^2}\sum_k\sum_l \left(\frac{\partial s(x_k, y_l)}{\partial d}\right)^2. \tag{117}$$

However, for the hypothesis test of interest in (4), $I(0)$ is zero and (116) is not directly applicable. Here we extend the approach in [12, p. 26] by considering higher order terms. Consider the following Taylor expansion:

$$\mathcal{J}(d) = \int_{\mathcal{D}} [p(\mathbf{g}, d) - p(\mathbf{g}, 0)] \log\left(\frac{p(\mathbf{g}, d)}{p(\mathbf{g}, 0)}\right) \mathrm{d}\mathbf{g} \tag{118}$$

$$= \mathcal{J}(0) + d\left.\frac{\partial \mathcal{J}}{\partial d}\right|_{d=0} + \frac{d^2}{2}\left.\frac{\partial^2 \mathcal{J}}{\partial d^2}\right|_{d=0} + \frac{d^3}{6}\left.\frac{\partial^3 \mathcal{J}}{\partial d^3}\right|_{d=0} + \frac{d^4}{24}\left.\frac{\partial^4 \mathcal{J}}{\partial d^4}\right|_{d=0} + O(d^6). \tag{119}$$

Noting that[16]

$$\left.\frac{\partial^i p(\mathbf{g}, d)}{\partial d^i}\right|_{d=0} = 0 \quad i = 1, 3. \tag{120}$$

We will have

$$\mathcal{J}(0) = 0$$

$$\left.\frac{\partial \mathcal{J}}{\partial d}\right|_{d=0} = \int_{\mathcal{D}} \frac{\partial p(\mathbf{g}, d)}{\partial d} \log\left(\frac{p(\mathbf{g}, d)}{p(\mathbf{g}, 0)}\right) + [p(\mathbf{g}, d) - p(\mathbf{g}, 0)] \left.\frac{\frac{\partial p(\mathbf{g}, d)}{\partial d}}{p(\mathbf{g}, d)}\right|_{d=0} \mathrm{d}\mathbf{g} = 0$$

$$\left.\frac{\partial^2 \mathcal{J}}{\partial d^2}\right|_{d=0} = \int_{\mathcal{D}} \left.\frac{2\left[\frac{\partial p(\mathbf{g}, d)}{\partial d}\right]^2}{p(\mathbf{g}, d)}\right|_{d=0} \mathrm{d}\mathbf{g} = 0$$

$$\left.\frac{\partial^3 \mathcal{J}}{\partial d^3}\right|_{d=0} = \int_{\mathcal{D}} \left.\frac{6\frac{\partial p(\mathbf{g}, d)}{\partial d}\frac{\partial^2 p(\mathbf{g}, d)}{\partial d^2}}{p(\mathbf{g}, d)} - \frac{3\left[\frac{\partial p(\mathbf{g}, d)}{\partial d}\right]^3}{[p(\mathbf{g}, d)]^2}\right|_{d=0} \mathrm{d}\mathbf{g} = 0,$$

$$\left.\frac{\partial^4 \mathcal{J}}{\partial d^4}\right|_{d=0} = \int_{\mathcal{D}} \left.\frac{8\frac{\partial p(\mathbf{g}, d)}{\partial d}\frac{\partial^3 p(\mathbf{g}, d)}{\partial d^3} + 6\left[\frac{\partial^2 p(\mathbf{g}, d)}{\partial d^2}\right]^3}{p(\mathbf{g}, d)} - \frac{18\left[\frac{\partial p(\mathbf{g}, d)}{\partial d}\right]^2 \frac{\partial^2 p(\mathbf{g}, d)}{\partial d^2}}{[p(\mathbf{g}, d)]^2} + \frac{8\left[\frac{\partial p(\mathbf{g}, d)}{\partial d}\right]^4}{[p(\mathbf{g}, d)]^3}\right|_{d=0} \mathrm{d}\mathbf{g}$$

[16]Since $p(\mathbf{g}, d)$ is an even (and differentiable) function around $d = 0$.

$$E_0 = \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{1}{L_u L_v}\left|\sum_{i=0}^{L_u-1}\sum_{j=0}^{L_v-1} H(u - 2\pi i, v - 2\pi j) e^{\sqrt{-1}[(u-2\pi i)\phi + (v-2\pi j)\psi]}\right|^2 \mathrm{d}u\mathrm{d}v \tag{100}$$

$$= \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{1}{L_u L_v}\sum_{i=0}^{L_u-1}\sum_{j=0}^{L_v-1} |H(u - 2\pi j, v - 2\pi j)|^2 \mathrm{d}u\mathrm{d}v + \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{2}{L_u L_v}\sum_{i=0}^{L_u-1}\sum_{i=0}^{L_v-1}\sum_{i'=i+1}^{L_u-1}\sum_{j'=j+1}^{L_v-1} \mathrm{d}u\mathrm{d}v \tag{101}$$

$$\times \Re[H(u - 2\pi i, v - 2\pi j) H^*(u - 2\pi i', v - 2\pi j') \times e^{\sqrt{-1}[(u-2\pi i)\phi + (v-2\pi j)\psi - (u-2\pi i')\phi - (v-2\pi j')\psi]}]$$

$$= \frac{1}{4\pi^2}\int_{-\pi L_v}^{\pi L_v}\int_{-\pi L_u}^{\pi L_u} |H(u, v)|^2 \mathrm{d}u\mathrm{d}v + \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{2}{L_u L_v}\sum_{i=0}^{L_u-1}\sum_{i=0}^{L_v-1}\sum_{i'=i+1}^{L_u-1}\sum_{j'=j+1}^{L_v-1} \mathrm{d}u\mathrm{d}v \tag{102}$$

$$\times \Re[H(u - 2\pi i, v - 2\pi j) H^*(u - 2\pi i', v - 2\pi j') e^{\sqrt{-1}[2\pi(i-i')\phi + 2\pi(j-j')\psi]}]$$

$$= f_s^2 \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h^2(x, y) \mathrm{d}x\mathrm{d}y + \left\{\begin{array}{c} \frac{2}{L_u L_v}\sum_{i=0}^{L_u-1}\sum_{i=0}^{L_v-1}\sum_{i'=i+1}^{L_u-1}\sum_{j'=j+1}^{L_v-1} \cos(2\pi(i-i')\phi + 2\pi(j-j')\psi) \\ \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} H(u - 2\pi i, v - 2\pi j) H(u - 2\pi i', v - 2\pi j') \mathrm{d}u\mathrm{d}v \end{array}\right\} \tag{103}$$

$$= \int_{\mathcal{D}} \frac{6\left[\frac{\partial^2 p(\mathbf{g},d)}{\partial d^2}\right]^2}{p(\mathbf{g},d)}\Bigg|_{d=0} \mathrm{d}\mathbf{g}.$$

As a result, we can write (118) as

$$\mathcal{J}(d) \approx \frac{d^4}{4} \int_{\mathcal{D}} \frac{\left[\frac{\partial^2 p(\mathbf{g},d)}{\partial d^2}\right]^2}{p(\mathbf{g},d)}\Bigg|_{d=0} \mathrm{d}\mathbf{g}. \tag{121}$$

On the other hand, we observe that

$$\begin{aligned}
\frac{\partial^2 I(d)}{\partial d^2}\Big|_{d=0} &= -\frac{\partial^2}{\partial d^2}\int_{\mathcal{D}} \frac{\partial^2 \ln p(\mathbf{g},d)}{\partial d^2} p(\mathbf{g},d)\Big|_{d=0} \mathrm{d}\mathbf{g} \\
&= \frac{\partial^2}{\partial d^2}\int_{\mathcal{D}} \frac{\left[\frac{\partial \ln p(\mathbf{g},d)}{\partial d}\right]^2}{p(\mathbf{g},d)}\Bigg|_{d=0} \mathrm{d}\mathbf{g} \\
&= \int_{\mathcal{D}} \frac{2\frac{\partial p(\mathbf{g},d)}{\partial d}\frac{\partial^3 p(\mathbf{g},d)}{\partial d^3} + 2\left[\frac{\partial^2 p(\mathbf{g},d)}{\partial d^2}\right]^3}{p(\mathbf{g},d)} \\
&\quad - \frac{5\left[\frac{\partial p(\mathbf{g},d)}{\partial d}\right]^2 \frac{\partial^2 p(\mathbf{g},d)}{\partial d^2}}{[p(\mathbf{g},d)]^2} + \frac{2\left[\frac{\partial p(\mathbf{g},d)}{\partial d}\right]^4}{[p(\mathbf{g},d)]^3}\Bigg|_{d=0} \mathrm{d}\mathbf{g} \\
&= \int_{\mathcal{D}} \frac{2\left[\frac{\partial^2 p(\mathbf{g},d)}{\partial d^2}\right]^2}{p(\mathbf{g},d)}\Bigg|_{d=0} \mathrm{d}\mathbf{g}.
\end{aligned} \tag{122}$$

Therefore, we show (123) and (124) at the bottom of the next page. As we see from (123), the divergence for the underlying hypothesis testing problem is directly related to the second derivative of the Fisher information matrix evaluated at $d = 0$.

Appendix E
Derivation of $\Delta\mathbf{Q}$ and $\Delta\mathbf{P}$ in (70)

Computing $\Delta\mathbf{Q}$ and $\Delta\mathbf{P}$ requires the following:

$$\begin{aligned}
\Delta E_0(h) &= f_s \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(x,y)\Delta h(x,y)\mathrm{d}x\mathrm{d}y \\
\Delta E_{10}(h) &= f_s \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{\partial h(x,y)}{\partial x}\frac{\partial \Delta h(x,y)}{\partial x}\mathrm{d}x\mathrm{d}y \\
&= f_s \int_{-\infty}^{+\infty}\left[\lim \frac{\partial h(x,y)}{\partial x}\Delta h(x,y)\Big|_{x\to-\infty}^{x\to+\infty} \right. \\
&\quad \left. - \int_{-\infty}^{+\infty} \frac{\partial^2 h(x,y)}{\partial x^2}\Delta h(x,y)\mathrm{d}x\right]\mathrm{d}y \\
&= -f_s \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{\partial^2 h(x,y)}{\partial x^2}\Delta h(x,y)\mathrm{d}x\mathrm{d}y \\
\Delta E_{01}(h) &= f_s \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{\partial h(x,y)}{\partial y}\frac{\partial \Delta h(x,y)}{\partial y}\mathrm{d}x\mathrm{d}y \\
&= -f_s \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{\partial^2 h(x,y)}{\partial y^2}\Delta h(x,y)\mathrm{d}x\mathrm{d}y \\
\Delta E_{20}(h) &= f_s \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{\partial^2 h(x,y)}{\partial x^2}\frac{\partial^2 \Delta h(x,y)}{\partial x^2}\mathrm{d}x\mathrm{d}y \\
&= f_s \int_{-\infty}^{+\infty}\left[\lim \frac{\partial^2 h(x,y)}{\partial x^2}\frac{\partial \Delta h(x,y)}{\partial x}\Big|_{x\to-\infty}^{x\to+\infty}\right.
\end{aligned}$$

$$[\mathbf{z}_1]_{lN+k} = \frac{\partial s(x,y)}{\partial p_x}\Big|_{x=x_k,y=y_l} = -\alpha h_{10}(x_k - p_x, y_l - p_y) \tag{108}$$

$$[\mathbf{z}_2]_{lN+k} = \frac{\partial s(x,y)}{\partial q_x}\Big|_{x=x_k,y=y_l} = \beta h_{10}(x_k + q_x, y_l + q_y) \tag{109}$$

$$[\mathbf{z}_3]_{lN+k} = \frac{\partial s(x,y)}{\partial p_y}\Big|_{x=x_k,y=y_l} = -\alpha h_{01}(x_k - p_x, y_l - p_y) \tag{110}$$

$$[\mathbf{z}_4]_{lN+k} = \frac{\partial s(x,y)}{\partial q_y}\Big|_{x=x_k,y=y_l} = \beta h_{01}(x_k + q_x, y_l + q_y) \tag{111}$$

$$[\mathbf{z}_5]_{lN+k} = \frac{\partial s(x,y)}{\partial \alpha}\Big|_{x=x_k,y=y_l} = h(x_k - p_x, y_l - p_y) \tag{112}$$

$$[\mathbf{z}_6]_{lN+k} = \frac{\partial s(x,y)}{\partial \beta}\Big|_{x=x_k,y=y_l} = h(x_k + q_x, y_l + q_y) \tag{113}$$

$$[\mathbf{z}_7]_{lN+k} = \frac{\partial s(x,y)}{\partial \phi_2}\Big|_{x=x_k,y=y_l} \tag{114}$$

$$= \begin{cases} 0 & (x_k,y_l) \in \Omega_1 \\ \alpha h_{10}(x_k - p_x, y_l - p_y) + \beta h_{10}(x_k + q_x, y_l + q_y) & (x_k,y_l) \in \Omega_2 \end{cases}$$

$$[\mathbf{z}_8]_{lN+k} = \frac{\partial s(x,y)}{\partial \psi_2}\Big|_{x=x_k,y=y_l} \tag{115}$$

$$= \begin{cases} 0 & (x_k,y_l) \in \Omega_1 \\ \alpha h_{01}(x_k - p_x, y_l - p_y) + \beta h_{01}(x_k + q_x, y_l + q_y) & (x_k,y_l) \in \Omega_2. \end{cases}$$

$$
\begin{aligned}
&- \int_{-\infty}^{+\infty} \frac{\partial^3 h(x,y)}{\partial x^3} \frac{\partial \Delta h(x,y)}{\partial x} \mathrm{d}x \Bigg] \mathrm{d}y \\
&= f_s \int_{-\infty}^{+\infty} \Bigg[- \lim \frac{\partial^3 h(x,y)}{\partial x^3} \Delta h(x,y) \Big|_{x \to -\infty}^{x \to +\infty} \\
&\quad + \int_{-\infty}^{+\infty} \frac{\partial^4 h(x,y)}{\partial x^4} \Delta h(x,y) \mathrm{d}x \Bigg] \mathrm{d}y \\
&= f_s \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\partial^4 h(x,y)}{\partial x^4} \Delta h(x,y) \mathrm{d}x \mathrm{d}y \\
\Delta E_{02}(h) &= f_s \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\partial^2 h(x,y)}{\partial y^2} \frac{\partial^2 \Delta h(x,y)}{\partial y^2} \mathrm{d}x \mathrm{d}y \\
&= f_s \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\partial^4 h(x,y)}{\partial y^4} \Delta h(x,y) \mathrm{d}x \mathrm{d}y \\
\Delta E_{11}(h) &= f_s \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\partial^2 h(x,y)}{\partial x \partial y} \frac{\partial^2 \Delta h(x,y)}{\partial x \partial y} \mathrm{d}x \mathrm{d}y \\
&= f_s \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\partial^4 h(x,y)}{\partial x^2 \partial y^2} \Delta h(x,y) \mathrm{d}x \mathrm{d}y
\end{aligned}
$$

where we assume that

$$
\lim h(x,y) = \lim \frac{\partial^{i+j} h(x,y)}{\partial x^i \partial y^j} = 0
$$

as x or $y \to \pm\infty$. Therefore, we show (125)–(126) at the bottom of the page, where

$$
\begin{aligned}
&\Delta \left(E_{20} - \frac{E_{10}^2}{E_0} \right) \\
&\quad = \Delta E_{20} - 2 \frac{E_{10}}{E_0} \Delta E_{10} + \frac{E_{10}^2}{E_0^2} \Delta E_0 \\
&\quad = f_s \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Bigg[h_{40}(x,y) - 2 \frac{E_{10}}{E_0} h_{20}(x,y) \\
&\qquad + \frac{E_{10}^2}{E_0^2} h(x,y) \Bigg] \Delta h(x,y) \mathrm{d}x \mathrm{d}y \\
&\Delta \left(E_{11} - \frac{E_{10} E_{01}}{E_0} \right) \\
&\quad = \Delta E_{11} - \frac{E_{01}}{E_0} \Delta E_{10} \\
&\qquad - \frac{E_{10}}{E_0} \Delta E_{01} + \frac{E_{10} E_{01}}{E_0^2} \Delta E_0 \\
&\quad = f_s \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Bigg[h_{22}(x,y) - \frac{E_{01}}{E_0} h_{20}(x,y) \\
&\qquad - \frac{E_{10}}{E_0} h_{02}(x,y) + \frac{E_{10} E_{01}}{E_0^2} h(x,y) \Bigg] \\
&\qquad \times \Delta h(x,y) \mathrm{d}x \mathrm{d}y
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{J}(d) &\approx \frac{d^4}{8} \left. \frac{\partial^2 I(d)}{\partial d^2} \right|_{d=0} \qquad (123) \\
&= \frac{d^4}{4\sigma^2} \sum_k \sum_l \left(\frac{\partial^2 s(x_k, y_l)}{\partial d^2} \right)^2 \Bigg|_{d=0} \\
&\quad + \left. \frac{\partial s(x_k, y_l)}{\partial d} \frac{\partial^3 s(x_k, y_l)}{\partial d^3} \right|_{d=0} \\
&= \frac{d^4}{4\sigma^2} \sum_k \sum_l \left(\frac{\partial^2 s(x_k, y_l)}{\partial d^2} \right)^2 \Bigg|_{d=0} + 0 \\
&= \frac{d^4}{4\sigma^2} \sum_k \sum_l \left(\frac{\partial^2 [h(x_k + d/2, y_l) + h(x_k - d/2, y_l)]}{\partial d^2} \right)^2 \Bigg|_{d=0} \\
&= \frac{d^4}{4\sigma^2} \sum_k \sum_l h_{20}^2(x_k, y_l) \Bigg|_{d=0}. \qquad (124)
\end{aligned}
$$

$$
\Delta \mathbf{P} = \begin{bmatrix}
\Delta E_0 & 0 & 0 & -\Delta E_{10} & -\Delta E_{01} & 0 \\
0 & \Delta E_{10} & 0 & 0 & 0 & 0 \\
0 & 0 & \Delta E_{01} & 0 & 0 & 0 \\
-\Delta E_{10} & 0 & 0 & \Delta E_{20} & \Delta E_{11} & 0 \\
-\Delta E_{01} & 0 & 0 & \Delta E_{11} & \Delta E_{02} & 0 \\
0 & 0 & 0 & 0 & 0 & \Delta E_{11}
\end{bmatrix} \qquad (125)
$$

and

$$
\Delta \mathbf{Q} = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 \\
0 & \Delta E_{10} & 0 & 0 & 0 & 0 \\
0 & 0 & \Delta E_{01} & 0 & 0 & 0 \\
0 & 0 & 0 & \Delta \left(E_{20} - \frac{E_{10}^2}{E_0} \right) & \Delta \left(E_{11} - \frac{E_{10} E_{01}}{E_0} \right) & 0 \\
0 & 0 & 0 & \Delta \left(E_{11} - \frac{E_{10} E_{01}}{E_0} \right) & \Delta \left(E_{02} - \frac{E_{01}^2}{E_0} \right) & 0 \\
0 & 0 & 0 & 0 & 0 & \Delta E_{11}
\end{bmatrix} \qquad (126)
$$

$$\Delta\left(E_{02}-\frac{E_{01}^2}{E_0}\right) = \Delta E_{02} - 2\tfrac{E_{01}}{E_0}\Delta E_{01} + \tfrac{E_{01}^2}{E_0^2}\Delta E_0$$
$$= f_s \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left[h_{04}(x,y) - 2\frac{E_{01}}{E_0}h_{02}(x,y) + \frac{E_{01}^2}{E_0^2}h(x,y)\right]\Delta h(x,y)\mathrm{d}x\mathrm{d}y$$

ACKNOWLEDGMENT

The authors wish to thank the reviewers for their comments and suggestions, which have helped improve the content of the paper.

REFERENCES

[1] E. Bettens, D. V. Dyck, A. J. den Dekker, J. Sijbers, and A. van den bos, "Model-based two-object resolution from observations having counting statistics," *Ultramicroscopy*, vol. 77, pp. 37–48, 1999.

[2] A. J. den Dekker, "Model-based optical resolution," *IEEE Trans. Instrum. Meas.*, vol. 46, pp. 798–802, 1997.

[3] M. Elad and A. Feuer, "Restoration of single super-resolution image from several blurred, noisy and down-sampled measured images," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1646–1658, Dec. 1997.

[4] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multi-frame super-resolution," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1327–1344, Oct. 2004.

[5] I. M. Gelfand and S. V. Fomin, *Calculus of Variations*. New York: Dover, 2000.

[6] J. W. Goodman, *Introduction to Fourier Optics.*. New York: McGraw-Hill, 1996.

[7] C. W. Helstrom, "The detection and resolution of optical signals," *IEEE Trans. Inf. Theory*, vol. IT-10, pp. 275–287, 1964.

[8] ——, "Detection and resolution of incoherent objects by a background-limited optical system," *J. Opt. Soc. Am.*, vol. 59, pp. 164–175, 1969.

[9] ——, "Resolvability of objects from the standpoint of statistical parameter estimation," *J. Opt. Soc. Am.*, vol. 60, pp. 659–666, 1970.

[10] S. M. Kay, *Fundamentals of Statistical Signal Processing, Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[11] ——, *Fundamentals of Statistical Signal Processing, Detection Theory*. Upper Saddle River, NJ: Prentice-Hall, 1998.

[12] S. Kullback, *Information Theory and Statistics*. New York: Dover, 1968.

[13] L. McWhorter and L. Scharf, "Cramer-Rao bounds for deterministic modal analysis," *IEEE Trans. Signal Process.*, vol. 41, pp. 1847–1862, May 1993.

[14] P. Milanfar and A. Shakouri, "A statistical analysis of diffraction-limited imaging," in *Proc. Int. Conf. Image Process.*, Rochester, NY, Sep. 2002, pp. 864–867.

[15] N. Nguyen, P. Milanfar, and G. Golub, "A computationally efficient image superresolution algorithm," *IEEE Trans. Image Process.*, vol. 10, no. 4, pp. 573–583, Apr. 2001.

[16] L. L. Scharf and P. H. Moose, "Information measures and performance bounds for array processors," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 11–21, Jan. 1976.

[17] M. Shahram, "Statistical and information-theoretic analysis of resolution in imaging and array processing," Ph.D. dissertation, University of California, Santa Cruz, CA, 2005.

[18] M. Shahram and P. Milanfar, "On the resolvability of sinusoids with nearby frequencies in the presence of noise," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2579–2588, Jul. 2005.

[19] ——, "A statistical analysis of achievable resolution in incoherent imaging," in *Proc. SPIE Annu. Meet., Signal Data Process. Small Tergets 2003*, San Diego, CA, Aug. 1–9, 2003, vol. 5204.

[20] ——, "Imaging below the diffraction limit, a statistical analysis," *IEEE Trans. Image Process.*, vol. 13, no. 5, pp. 677–689, May 2004.

[21] S. Smith, "Accuracy and resolution bounds for adaptive sensor array processing," in *Proc. 9th IEEE Signal Proc. Workshop Statist. Signal Array Proc.*, Portland, OR, 1998, pp. 37–40.

[22] ——, "Statistical resolution limits and the complexified Cramer-Rao bound," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1597–1609, 2005, to appear in.

[23] J. L. Starck and F. Murtagh, *Astronomical Image and Data Analysis*. New York: Springer-Verlag, 2002.

[24] G. Strang, *Introduction to Applied Mathematics*. Cambridge, MA: Wellesley-Cambridge, 1986.

[25] A. van den Bos, "Ultimate resolution: A mathematical framework," *Ultramicroscopy*, vol. 47, pp. 298–306, 1992.

[26] C. R. Wylie, *Advanced Engineering Mathematics*. New York: McGraw-Hill, 1975.