Theory and Design of Local Interpolators

A. SCHAUM

Naval Research Laboratory, Code 5620, Washington, DC 20375-5000

This paper shows that an error spectrum can be used to describe the performance of any convolutional interpolator used to shift an oversampled image. This spectrum is linear in the image power spectrum and in an error factor that depends only on the interpolator and the shift. The same form is shown to describe the interpolation of undersampled data, in an average sense. Simple formulas are derived for the error factor in either Fourier or real space, and standard interpolators are evaluated with them. Optimal interpolators are derived for various theoretical spectra: constant inband, Lorentzian, power law, and Gaussian. Practical methods of interpolator design are devised for use with image spectra that are known only partially or are not easily characterized analytically. © 1993 Academic Press, Inc.

1. INTRODUCTION AND BACKGROUND

This paper develops a general mathematical method for evaluating and optimizing any local interpolator that operates on uniformly spaced sample points. The method is used to examine standard interpolators and to design optimal ones in a variety of contexts, such as for low-frequency dominant spectra and for mean squared error minimization.

An important application of interpolation is in the registration of image pairs, a process that is usually implemented in two steps. The first is the estimation of the relative shift or local distortion, for which scene-based algorithms have been developed [1] that can be accurate to a level of approximately 1/100 of a sample, depending on image statistics. Residual registration error after use of displacement estimates is often dominated by inaccuracy in the second component of registration, resampling, which is interpolation at a discrete set of points.

Registration "plays a crucial role in the correction of raw satellite image data" [2] in remote sensing applications. The importance of algorithmic methods of image alignment then lies in the relaxation of optical/mechanical alignment tolerances [3], which is a vital cost/risk factor for satellite-based systems. Applications involving image registration are expanding rapidly in medical imaging [4].

In digital subtraction angiography, images are compared before and after injection of X-ray absorbing elements, and patient or organ motion causes relative translation and rotation between successive images, which then require registration and resampling for comparison [5]. Improvements in the quality of such images can allow "reduction of injected contrast agents or X-ray doses" [6]. Interpolation also has many traditional signal processing applications: speech processing, frequency multiplexing of single sideband systems, digital beamforming [7], time delay estimation, and data compression.

The detection of changes in a pair of images is also the basis for monitoring geologic, agricultural, and oceanic evolution in ecological and planetary studies. In military applications, autonomous surveillance and search and track systems often base the detection of moving targets on digital background subtraction followed by thresholding. The accuracy of all these applications ultimately can hinge on the fidelity of the interpolator.

Progress in interpolator design can be traced to the former use of a suboptimal polynomial interpolator called "Cubic Convolution" [8] to reconstruct Landsat digital imagery [9]. This method was shown [10] to be a special instance of a class of 4-point interpolators that were named "Parametric Cubic Convolution" (PCC), and the optimal value of the relevant parameter α was found to differ from that in popular use. The meaning of optimality depends on context, and in [10] values of α were found that minimized a particular error measure: (a) in an image-independent sense, and (b) absolutely, for arbitrary but known power spectrum of the signal to be resampled.

For (a), image energy was assumed to be concentrated at low spatial frequencies. All members of PCC have perfect dc ($\nu = 0$) responses, and α was selected to enhance this low-frequency behavior by requiring the error measure to be a maximally flat function of frequency near dc. This low-frequency optimal choice of parameter value ($\alpha = -\frac{1}{2}$) supplanted the original choice ($\alpha = -1$), and eventually the name Cubic Convolution (CC) came to refer to PCC with the new choice of α. Here we also use CC to mean PCC with $\alpha = -\frac{1}{2}$.

The error measure used in [10] is shown in Section 4 to be an average resampling error over all possible locations of the resampling point, which is a reasonable measure for the *interpolation* problem, i.e., the construction of a dense set of resampled points. However, here we show how to find and minimize the error for any given value of the shift s, which is the distance of the resampled points from the nearest samples. It is shown further that the minimum-error 4-point resampler in the sense of (a) above is indeed a cubic convolution, but is is not a member of the class PCC.

The function defined on the continuum by resampling an arbitrary signal at all shifts is called the interpolated function f_i. The class of PCC interpolators was defined in part by constraining $f_i(x)$ to be smooth, that is, to have continuous values and first derivatives. The imagery that results from reconstructing a discrete signal on the continuum with PCC has no steps or kinks. Therefore, PCC is actually only a subclass of interpolators: Smooth Cubic Convolutions.

The optimal solution in the image-independent sense (a) for any fixed number N of sampling points is here called LF-N, for Low-Frequency optimal. It is a local version of the Lagrange interpolator. The LF-N solutions generally can produce discontinuous first derivatives in interpolated imagery, and when N is odd the functional values can be discontinuous as well. LF-N is smooth only in the limit $N \to \infty$.

When optimized in the absolute sense of (b), PCC results in spectrum-dependent values of α, but the result is, of course, always a cubic interpolator. It is shown here that usually the absolutely optimal, i.e., minimum error, solution is not a polynomial type.

2. STANDARD INTERPOLATORS

Here the philosophy of Parker *et al.* [5] and others is adopted, in which resampling is treated as a two-step process (Fig. 1):

- interpolation to form a function defined on the continuum, followed by
- sampling the new function at whatever shift s is desired.

Letting $f(x)$ represent the underlying image intensity at position x, the sampled values of f, $f(n)$ (n an integer), are used to produce an interpolation estimate f_i of f:

$$f_i(x) = \sum_n r(x - n)f(n). \qquad (1)$$

The kernel r defines the method of interpolation. The treatment here is in one dimension. Two-dimensional interpolation can be accomplished by a cascade of 2 one-dimensional operations.
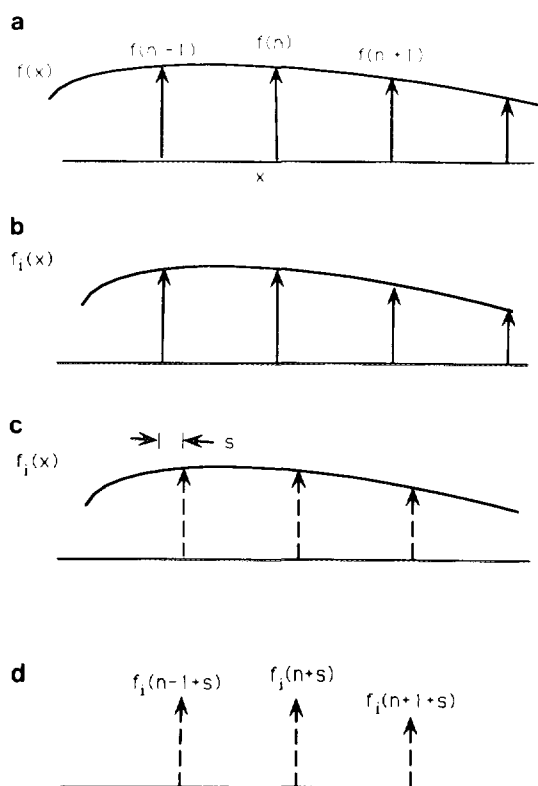

FIG. 1. (a) Discrete set $f(n)$ of sampled values from a continuous image $f(x)$. (b) Interpolation, based on the sampled values, forming a continuous image $f_i(x)$. (c) Sampling the interpolated image at a shift s to form (d) a new set of resampled values.

Because the sampling grid is defined here to have unit spacing, the Nyquist frequency is $\nu_{Nyq} = \frac{1}{2}$ cycles/sample. If the Fourier transform of f has no components with frequency $\nu > \frac{1}{2}$ we call f "oversampled." Such a function is bandlimited and is sampled at a rate greater than twice its highest frequency component. According to the Nyquist Reconstruction Theorem, the kernel

$$r(x) = \text{SINC}(x) \equiv \frac{\sin(\pi x)}{\pi x} \qquad (2)$$

can be used in (1) to reproduce f exactly, that is, $f_i = f$, whenever f is oversampled.

In practice, r must be of finite support (the range of x over which r is nonzero) to make (1) contain only a finite number of terms. For example, the SINC function can be truncated at $\pm N/2$. We call such an interpolator SINC-N. Two other common interpolators are Nearest Neighbor (NN) and Linear (LIN). Figure 2 illustrates $r(x)$ for NN, LIN, and SINC. The supports for NN and LIN are 1 and 2 respectively. The SINC has infinite support, but when truncated at $\pm N/2$ to SINC-N, its support becomes N. Table 1 lists the corresponding analytic forms for r.

TABLE 1
Common Interpolators (r, $\hat{r}$ are 0 where not specified)

Interpolator	$r(x)$	$\hat{r}(\nu)$
SINC ($|x| < \infty$)	$\dfrac{\sin(\pi s)}{\pi x}$	1 for $|\nu| < \frac{1}{2}$
NN ($|x| < \frac{1}{2}$)	1	$\text{sinc}(\nu)$
LIN ($|x| < 1$)	$1 - |x|$	$\text{sinc}^2(\nu)$
DFT-N ($|x| < N/2$)		
(N odd)	$\dfrac{\sin(\pi x)}{N \sin(\pi x/N)}$ [a]	$\displaystyle\sum_{k=-(N-1)/2}^{(N-1)/2} \text{sinc}(N\nu - k)$
(N even)	$\dfrac{\sin(\pi x)}{N \tan(\pi x/N)}$	$\displaystyle\sum_{k=-((N/2)-1)}^{(N/2)-1} \text{sinc}(N\nu - k)$
		$+\dfrac{1}{2}\left\{\text{sinc}\left[N\left(\nu - \dfrac{1}{2}\right)\right] + \text{sinc}\left[N\left(\nu + \dfrac{1}{2}\right)\right]\right\}$
PCC		
$|x| < 1$	$(\alpha + 2)|x|^3 - (\alpha + 3)|x|^2 + 1$	$\dfrac{3}{(\pi\nu)^2}[\text{sinc}^2(\nu) - \text{sinc}(2\nu)]$
$1 < |x| < 2$	$\alpha(|x|^3 - 5|x|^2 + 8|x| - 4)$	$+\dfrac{2\alpha}{(\pi\nu)^2}[3\,\text{sinc}^2(2\nu) - 2\,\text{sinc}(2\nu) - \text{sinc}(4\nu)]$

[a] This form of $r(x)$ for DFT-N was first derived by R. Lucke, Naval Research Laboratory.

Another interpolator whose performance has been studied with methods of experimental mathematics [11] is based on the Discrete Fourier Transform (DFT). The method is inspired by the Fourier-shift theorem. The idea is to compute the DFT, perhaps using modern Fast Fourier Transform algorithms, and then shift the phases of this finite, discrete transform as if it were the continuous Fourier transform. The result is then Fourier-inverted to produce a grid of resampled values. DFT-4, which interpolates with four points, also is illustrated in Fig. 2. The DFT kernels have simple analytic forms: r and $\hat{r}$ are listed in Table 1.
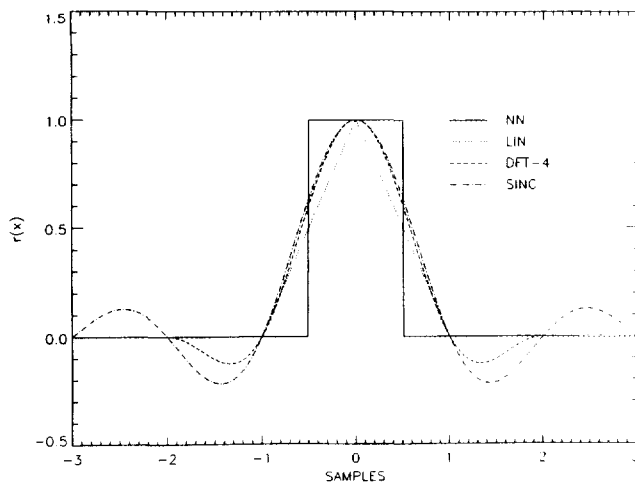

FIG. 2. Interpolation kernels $r(x)$ for Nearest Neighbor, Linear, DFT-4, and SINC, with supports 1, 2, 4, and ∞ samples, respectively.

3. DERIVATION OF THE ERROR FORMULA

The two most intuitive ways of designing interpolators are to require similarity between:

- the kernel r and the SINC function, which is the perfect interpolator (of oversampled functions), or
- the Fourier transforms of r and SINC.

As an example of the first design method, the class PCC was defined not only by constraining r to be smooth like the SINC, but also by requiring $r(x)$ to agree with SINC(x) at all integral x. Also, the original cubic convolution interpolator corresponds to the parametric value $\alpha = -1$ in PCC; this choice equates the derivative of the kernel $r(x)$ to that of SINC at $x = \pm 1$. Other parametric choices can also be understood through similar interpretations in real space. Figure 3 compares PCC for several values of the parameter α to SINC, the ideal interpolator of oversampled functions. The closeness of the $\alpha = -1$ curve to the SINC in the range $|x| \leq 1$ might also explain the early popularity of this parametric choice. Table 1 includes the analytic forms for PCC [10].

Analogous comparisons in Fourier space are often made between practical interpolators and the theoretical ideal. We use " ^ " to denote a Fourier transform, for example,

$$\hat{r}(\nu) = \int e^{-i\omega x} r(x)\, dx \quad (\omega \equiv 2\pi\nu). \tag{3}$$

(All integration ranges are taken to be $-\infty$ to $+\infty$.)
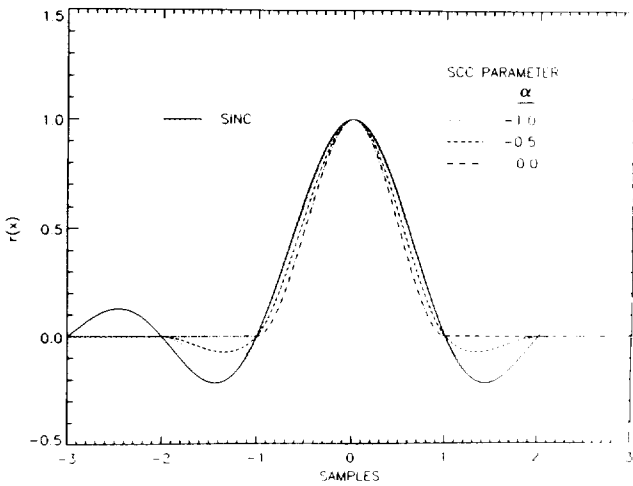
FIG. 3. Kernels (smooth) for Parametric Cubic Convolution (PCC) compared to SINC.

The transforms of the Fig. 2 interpolators are listed in Table 1 and plotted in Figure 4 for $|\nu| \leq 3$. Among the examples shown, the transform of the SINC is the only one of truly finite frequency support, with unit value in-band ($|\nu| < \frac{1}{2}$) and zero in the sidebands ($|\nu| > \frac{1}{2}$). Interpolator design is often based on minimizing differences between $\hat{r}$ and the SINC transform over some frequency range of interest. Alternatively, the transform of the interpolated function,

$$\hat{f}_i(\nu) = \sum_n f(n)e^{-i\omega n}\hat{r}(\nu),$$ (4)

can be compared to the true value $\hat{f}(\nu)$.

Here we use the squared error at shifts s

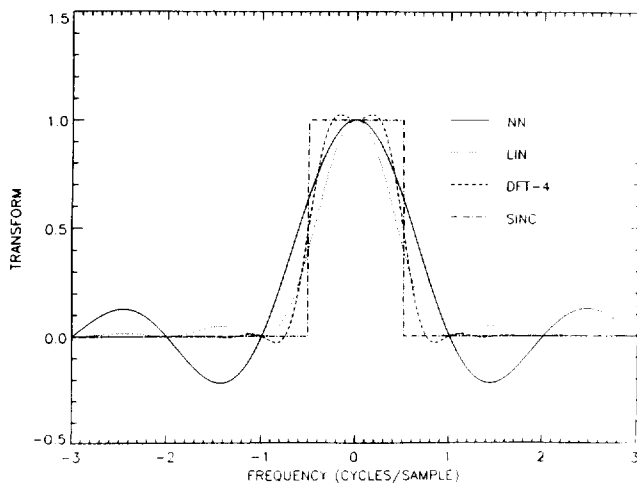$$d_s^2 \equiv \sum_n [f_i(n + s) - f(n + s)]^2$$ (5)

FIG. 4. Transforms $\hat{r}(\nu)$ of the Fig. 2 kernels. The Nyquist frequency is $\nu = \frac{1}{2}$ cycles/sample.

as the performance metric. This section derives a fundamental formula for the error as a function of image power spectrum and kernel tap weights, $r(s + n)$. But first, several preliminary results are required.

A lemma central to the method used here is proved with distinct approaches in [19] and [20]:

$$\text{comb}(x) \Leftrightarrow \text{comb}(\nu).$$ (6)

The correspondence $\Leftrightarrow$ relates a function to its Fourier transform. The comb function is formally defined by

$$\text{comb}(x) = \sum_n \delta(x - n),$$ (7)

with δ the Dirac delta function. Equation (6) says that the comb is its own Fourier transform. The formal transform of (7)

$$\text{comb}(\nu) = \sum_n e^{-2\pi i n\nu}$$ (8)

will also prove useful.

Other correspondences for arbitrary functions g, h, are the Fourier shift theorems

$$g(x + s) \Leftrightarrow e^{2\pi i s\nu}\hat{g}(\nu)$$ (9)

$$e^{-2\pi i x\nu_0}g(x) \Leftrightarrow \hat{g}(\nu + \nu_0),$$ (10)

and the convolution theorem

$$g(x)h(x) \Leftrightarrow (\hat{g} * \hat{h})(\nu)$$
$$(g * h)(x) \Leftrightarrow \hat{g}(\nu)\hat{h}(\nu).$$ (11)

The definitions of convolution and of the comb (7) can be used to prove:

$$\sum_n g(x - n)h(n) = [g * (\text{comb} \cdot h)](x).$$ (12)

A special case of (12) is also useful:

$$\sum_n g(x - n) = (g * \text{comb})(x).$$ (13)

These relations can be used to express a discrete sum of squared sample values of a function g in terms of its Fourier transform $\hat{g}$. Letting $g \rightarrow g^2$, $x \rightarrow 0$ in (13), we can write

$$\sum g^2(n) = (g^2 * \text{comb})(0).$$ (14)

Letting $\mathcal{F}$ denote the Fourier transform operator, we successively apply (11) and (6), the definition of $\mathcal{F}^{-1}$, and (11) again to simplify (14) further:

$$\sum g^2(n) = [\mathscr{F}^{-1}(\mathscr{F}(g^2 * \mathrm{comb}))](0)$$
$$= [\mathscr{F}^{-1}(\mathscr{F}(g^2)\,\mathrm{comb})](0)$$
$$= \int dv'\,[\mathscr{F}(g^2)(v')\,\mathrm{comb}(v')] \tag{15}$$
$$= \int dv'\,[\hat{g} * \hat{g}](v')\,\mathrm{comb}(v').$$

Next the frequency-space version of (7) is used in (15), together with the definition of convolution:

$$\sum g^2(n) = \sum_n \int dv'\,\delta(v' - n) \int dv\,\hat{g}(v' - v)\hat{g}(v)$$
$$= \sum_n \int dv\,\hat{g}(n - v)\hat{g}(v). \tag{16}$$

If g is real, then

$$\hat{g}(v) = \hat{g}^*(-v) \quad \forall v, \tag{17}$$

and then from (16),

$$\sum g^2(n) = \sum_n \int dv\,\hat{g}^*(v - n)\hat{g}(v). \tag{18}$$

If, furthermore, the sequence $g(n)$ represents an oversampled version of g, so that $\hat{g}(v) = 0$ for $|v| \geq \frac{1}{2}$, then (18) simplifies to

$$\sum g^2(n) = \int dv\,|\hat{g}(v)|^2, \tag{19}$$

which is closely related to Parseval's Theorem [18]. Note the distinction, however: the integration in (19) extends over all frequencies, and $\hat{g}$ is the Fourier transform of the underlying continuous function g, not of the discrete Fourier series $\{g(n)\}$. These two transforms are equal for $|v| < \frac{1}{2}$ if g is oversampled, but even if it is not, Eq. (19), which we call the generalized Parseval's Theorem, still holds in an average sense described below.

Equation (19) implies that as long as g is oversampled, $\sum g^2(n)$ is independent of where the sampling grid is laid down relative to the image. Shifting the grid by an amount t in one direction is equivalent to shifting $g(x)$ in the other:

$$g(x) \rightarrow g(x + t). \tag{20}$$

According to (9), in Fourier space (20) is equivalent to

$$\hat{g}(v) \rightarrow \hat{g}(v)e^{2\pi itv}, \tag{21}$$

under which change (19) is invariant. Because the value of (19) is independent of grid placement, we call it the strong form of the generalized Parseval's Theorem. When g is not oversampled, a weaker version of (19), described below, still holds.

Under the translation of the function f by the amount $t \in [0, 1]$, according to (21), (18) becomes

$$\sum g^2(n) \rightarrow \sum_n e^{2\pi itn} \int dv\,\hat{g}^*(v - n)\hat{g}(v). \tag{22}$$

Note that the result of averaging (22) over t: $0 \rightarrow 1$ is (19), because only the $n = 0$ term survives on the right side. Therefore, the weaker form of the theorem is also expressed by (19), if the left-hand side is understood to be averaged over all placements of the sampling grid.

In our application of (19), $\sum g^2(n)$ will be d_s^2, which is a measure of interpolation error. The mean over t amounts to an average over grid translations. Therefore, it is appropriate whenever the imagery of interest contains no preferred features relative to the sampling grid, which is the usual practical case.

To identify g with the interpolation problem, we let

$$g(x) = f_i(x + s) - f(x + s). \tag{23}$$

Then (5) allows the identification

$$d_s^2 = \sum g^2(n). \tag{24}$$

If g is oversampled, then (18) becomes (19), so that

$$d_s^2 = \int dv\,|\hat{g}(v)|^2, \tag{25}$$

in which, from (23) and (9),

$$\hat{g}(v) = [\hat{f}_i(v) - \hat{f}(v)]e^{2\pi ivs}. \tag{26}$$

As is the case for (19), Eq. (25) also has a strong version and a weak. If, for example, g is oversampled, then (25) is exact, without any averaging over grid translations. Unfortunately, for the interpolation problem f_i, and hence g, is almost never bandlimited. Therefore, g cannot be oversampled, even if f is. So the condition under which (25) has been shown to hold in the strong sense is not usually satisfied when g is given by (23).

Nevertheless, when (23) defines g, we can show that (25) is still exact whenever f is oversampled, even if g is not.

Using (12) in (1) with $h = f$ and $g = r$, taking the transform, and applying (9) and (11) results in

$$\hat{f}_i(v) = \hat{r}(v)[\mathrm{comb} * \hat{f}](v), \tag{27}$$

which is usually not bandlimited because neither factor is, as shown below. Consequently, $\hat{g}$ in (26) is usually not bandlimited.

The frequency version of (13) allows (27) to be rewritten as

$$\hat{f}_i(\nu) = \hat{r}(\nu) \sum_m \hat{f}(\nu - m). \tag{28}$$

The second factor is nonzero for arbitrarily large ν, as is the first whenever the kernel r is of finite support, i.e., for all practical interpolators. Consequently, f_i is not bandlimited, as claimed.

Nevertheless, using (18) together with (26) and (28) results in

$$\sum_n g^2(n)$$

$$= \sum_n \int d\nu \left[\hat{r}^*(\nu - n) \sum_m \hat{f}^*(\nu - n - m) - \hat{f}^*(\nu - n)\right] \tag{29}$$

$$\times\, e^{2\pi i s n} \left[\hat{r}(\nu) \sum_j \hat{f}(\nu - j) - \hat{f}(\nu)\right],$$

which we now show to be independent of grid placement whenever f is oversampled.

Assuming then that

$$\hat{f}(\nu) = 0 \quad for \ |\nu| \geq \tfrac{1}{2}, \tag{30}$$

we examine the most complicated cross term of (29):

$$\sum_n e^{2\pi i s n} \int d\nu\, \hat{r}^*(\nu - n) \sum_m \hat{f}^*(\nu - n - m)\hat{r}(\nu) \sum_j \hat{f}(\nu - j). \tag{31}$$

Because of (30), the integration in (31) kills all summand terms except when $n + m = j$. So (31) becomes

$$\sum_n e^{2\pi i s n} \int d\nu\, \hat{r}^*(\nu - n) \sum_m \hat{f}^*(\nu - n - m)\hat{r}(\nu)\hat{f}(\nu - n - m). \tag{32}$$

Changing variables, $\nu \to \nu + n + m$, in (32) produces

$$\int d\nu\, |\hat{f}(\nu)|^2 \left\{ \sum_{n,m} e^{2\pi i s n}\hat{r}^*(\nu + m)\hat{r}(\nu + n + m)\right\}. \tag{33}$$

Each of the other cross terms of (29) produces a result similar to that of (33), but with a simpler expression within the braces. The final result,

$$d_s^2 = \int |\hat{f}(\nu)|^2\, e_s^2(\nu)\, d\nu, \tag{34}$$

with

$$e_s^2(\nu) = \sum_n e^{2\pi i s n} \left\{ \delta_0^n - \hat{r}^*(\nu - n) - \hat{r}(\nu + n) \right.$$
$$\left. + \sum_m \hat{r}^*(\nu + m)\hat{r}(\nu + n + m)\right\}, \tag{35}$$

is manifestly independent of grid placement. The quantity $|\hat{f}(\nu)|^2\, e_s^2(\nu)$ is called the error spectrum. The error factor $e_s^2(\nu)$ depends on only the interpolating kernel, not on the image being interpolated. Equation (35) can be expressed more simply as

$$e_s^2(\nu) = |E_s(\nu)|^2, \tag{36}$$

with

$$E_s(\nu) \equiv \sum_n [e^{2\pi i n s}\hat{r}(\nu + n)] - 1, \tag{37}$$

a "complex error" factor.

Summarizing, for any fixed shift s, the interpolation error is independent of an oversampled image's location relative to the sampling grid. Furthermore, the error is linear in the power spectrum. Power spectrum here simply means the squared modulus of the Fourier transform. It is not necessary to assume that $\{f(x)\}$ describes a stationary stochastic process, in which case the power spectrum equals the Fourier transform of the autocorrelation function. Equation (34) holds regardless of the stochastic character of $\{f(n)\}$, including deterministic.

Note that if s is a random variable that can be described only probabilistically, then the Fourier series in (35) can be used to express the moments of the error, because all the s dependence is contained in the first factor. For example, in the interpolation (as opposed to the resampling) problem, for which s can be considered uniformly distributed on [0, 1], the mean value of the error is just the $n = 0$ term of (35).

Also note that if f is undersampled, then (34) is still true in the average sense discussed earlier. The proof parallels that used above for the weak form of the generalized Parseval's Theorem. That is, the mean of d_s^2 over all grid placements depends only on the power spectrum of f and the error factor (35).

At this point, the dependence of the interpolation error on the image power spectrum has been isolated in (34), but the error factor remains expressed as an infinite sum over sidebands in (36) and (37). These are impractical forms because realistic interpolators have infinite frequency-support. But note that the first term of (37) can be written as

$$\sum_n e^{2\pi i n s}\hat{r}(\nu + n) = \sum_n e^{2\pi i n s} \int r(x)e^{-2\pi i x(\nu + n)}\, dx. \tag{38}$$

The real-space version of (8) together with (7) then permits (38) to be rewritten as

$$\int r(x)e^{-2\pi i x \nu} \operatorname{comb}(s - x)\, dx = \sum_n e^{-2\pi i \nu(s+n)} r(s + n). \quad (39)$$

Therefore,

$$E_s(\nu) = \sum_n [e^{-i\omega(s+n)} r(s + n)] - 1, \quad (40)$$

which is the analog of (37) in terms of r instead of $\hat{r}$. Note that the sum in (40) is finite whenever the interpolator is of finite support.

When expressed in real quantities only, (36) and (40) become ($\omega \equiv 2\pi\nu$)

$$e_s^2(\nu) = 1 - 2 \sum_n \cos(\omega[n + s]) r(n + s)$$
$$+ \sum_m \cos(\omega m) \sum_n r(n + m + s) r(n + s), \quad (41)$$

which is the analog of (35).

Equations (40) and (36), together with the strong and weak interpretations of (34), permit the evaluation of any interpolator at shift s in terms of a finite number of kernel values whenever r has finite support. Should $\hat{r}$ have finite support, (37) plays a role similar to that of (40).

4. RELATION TO PRIOR WORK

The papers by Park and Schowengerdt [9, 10] are seminal to this work. They considered an error measure,

$$d^2 = \int_{-\infty}^{+\infty} [f_i(x) - f(x)]^2\, dx. \quad (42)$$

Note from (5) that

$$d^2 = \langle d_s^2 \rangle, \quad (43a)$$

where $\langle\ \rangle$ denotes a mean value over an assumed uniformly distributed random shift $s: 0 \to 1$. The error d^2 is thus appropriate when a (quasi) continuum of resampled values is of interest as, for example, in some image display applications.

Park and Schowengerdt [10] also derived an error factor $e^2(\nu)$, which can be interpreted similarly as a mean value, now of the error factor defined in this paper. That is,

$$e^2(\nu) = \langle e_s^2(\nu) \rangle. \quad (43b)$$

The mean value of the fundamental equation (34) is

$$d^2 = \int |\hat{f}(\nu)|^2 e^2(\nu)\, d\nu, \quad (44)$$

and the mean value of the error (41) can be written as

$$e^2(\nu) = 1 - 2\hat{r}(\nu) + \sum_m \cos(\omega m)(r * r)(m), \quad (45)$$

in which r is assumed to be symmetric (as in Ref. [10]). Equation (45) can be called the fundamental formula of Ref. [10]. It is accompanied by a fundamental theorem, namely the fundamental theorem that was derived in this paper, but with the substitutions

$$e_s^2(\nu) \to e^2(\nu) \quad \text{and} \quad d_s^2 \to d^2, \quad (46)$$

in (34).

This theorem was less surprising in Ref. [10] than it is here, because there it was shown to hold only in the above sense of an average over the shift s of the two sampling grids (shown, for example, in Figs. 1b and 1c) relative to each other. This should be distinguished from the grid placement averaging considered in the weak versions of the theorems discussed previously. That operation held the shift of one grid relative to the other fixed at the value s; the pair of grids was shifted as a unit relative to the image (the continuous function $f(x)$ in Fig. 1a), or vice versa.

Also, in applications for which some performance criterion is defined in terms of d^2 or $e^2(\nu)$, Eq. (45) is normally useful only if a parameterized form r_α is used to define a class of kernels (and hence also of transforms $\hat{r}_\alpha$). For example, d^2 can be minimized for a given power spectrum, or $e^2(\nu)$ can be constrained in some preferred way by adjusting the parameter α. In Ref. [10] the class of interpolators was PCC (Table 1). This choice allowed the parametric calculations of $\hat{r}_\alpha$ and $r_\alpha * r_\alpha$, which could then be inserted into (45), to be followed by the selection of that parameter α which minimized d^2.

On the other hand, when (41) is used to constrain the error—at some fixed s—the resulting equations depend on the N kernel values $\{r(s + n)\}$, not on the transform of r or its self-convolution. Typically this results in algebraic equations for these tap weights for any shift s, permitting the calculation of optimal kernels, rather than kernels that are optimal only in an averaged sense, and that are constrained within some class. As noted earlier, in all optimality contexts considered in Ref. [10], the solutions for PCC were suboptimal because of the restrictive smoothness constraint defining the PCC class. Notice from Table 1 that r for PCC is continuous, as is its first derivative, for all values of α.

5. PERFORMANCE EVALUATIONS

Examples

Figure 5a plots the error factors for some common interpolators for a .25 sample shift. The frequency 0.5 cycles/sample is the Nyquist frequency, above which no energy is present in an oversampled image. Generally at low frequencies, $e_s^2(\nu)$ varies as ν^2 for Nearest Neighbor, ν^4 for Linear, ν^6 for CC ("Cubic Convolution," i.e., PCC
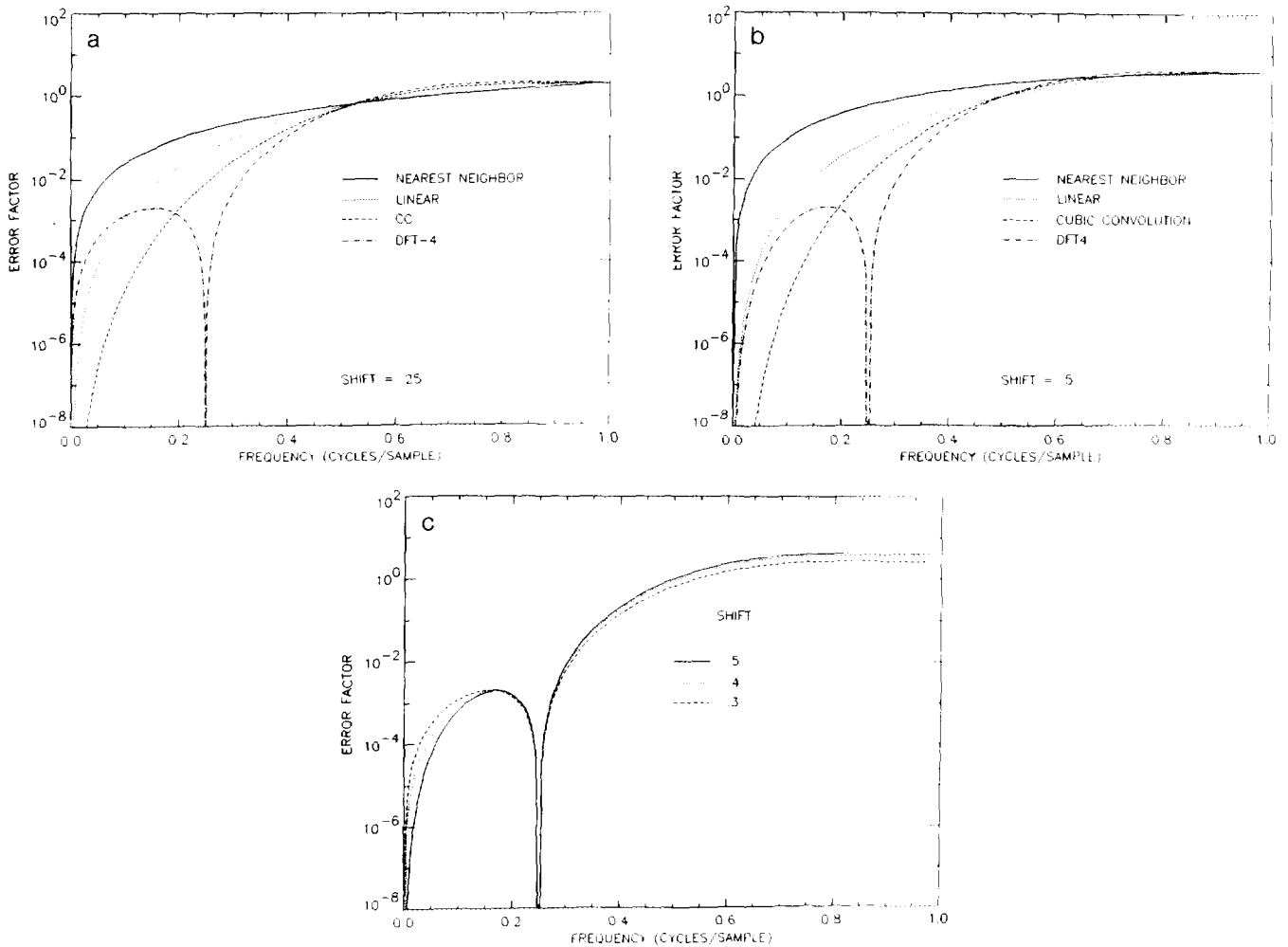
FIG. 5.   Error factors $e_i^2(\nu)$ for standard interpolators: (a) shift = .25 samples, (b) shift = .50 samples, and (c) DFT-4 at three shifts.

with $\alpha = -\frac{1}{2}$), and ν^2 for DFT-N interpolators. The small errors near dc are typical of polynomial methods. N-point interpolators that are designed optimally in this low-frequency region, called LF-N, have squared errors that vary as ν^{2N}. Nearest Neighbor interpolation is the same as LF-1, and LIN is LF-2. However, as remarked earlier, CC is not LF-4, for which $e_5^2(\nu)$ is smaller at low frequencies, varying as ν^8.

By contrast, a DFT interpolator of any order usually has a low-ν error that varies like ν^2, the same as for NN. An exception occurs at $s = \frac{1}{2}$ (Fig. 5b) at which $e_{1/2}^2(\nu)$ improves for DFT at low frequencies, varying as ν^4, like LIN. Associated with this fact is the unusual property of DFT interpolators that they can perform better at the normally most stressing shift, $s = \frac{1}{2}$, than at lower shifts. Figure 5c compares the performance of DFT-4 at three shifts. (For all s, DFT-4 has zero error at the DFT frequencies $\nu = 0$, $\frac{1}{2}$.) The plots show that, when applied to imagery with predominant energies below approximately

.15 cycles/sample, DFT-4 errors actually increase as the shift decreases from .5 to .3 samples.

Truncated SINC Interpolators

It is shown in Section 6 that SINC-N, the truncated sinc function, is the optimal kernel for a constant inband image spectrum. Error factors for SINC-2,4,6 are plotted in Fig. 6. For a constant image spectrum, these can also be interpreted as the error spectra. Note that the truncated SINCs reproduce certain frequencies almost perfectly. That is, the error plunges to near zero at certain frequencies for all shifts. This means that the continuous function f_i that results from interpolating from any set of sampled values of a sinusoid with one of these frequencies is nearly a copy of that sinusoid.

The oscillations in the Fourier transform of SINC-N that correspond to these dips in the error factor are often seen as undesirable. Figure 7 shows the result of applying
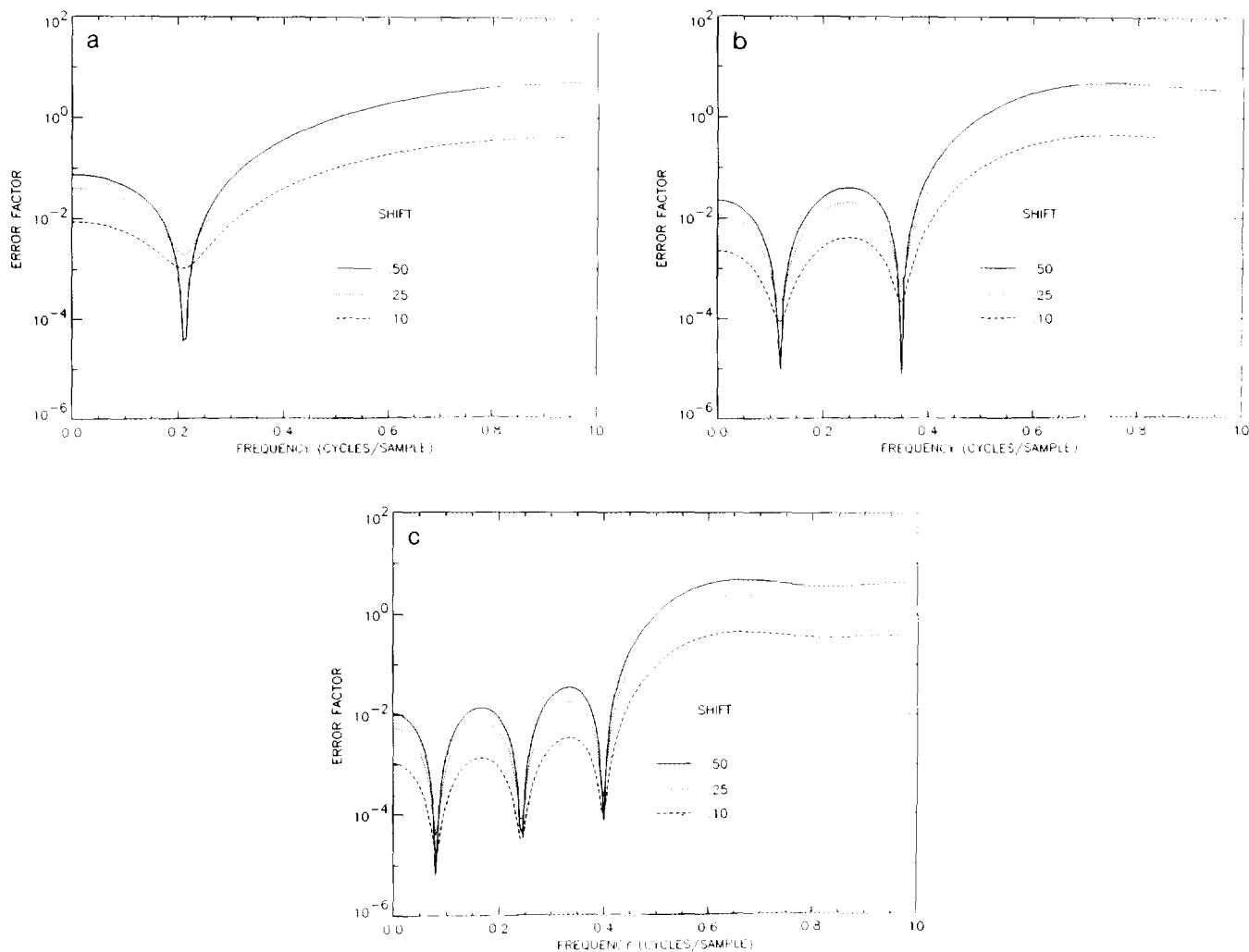
FIG. 6. $e_1^2(\nu)$ for truncated SINCs at three shifts: (a) SINC-2, (b) SINC-4, and (c) SINC-6.
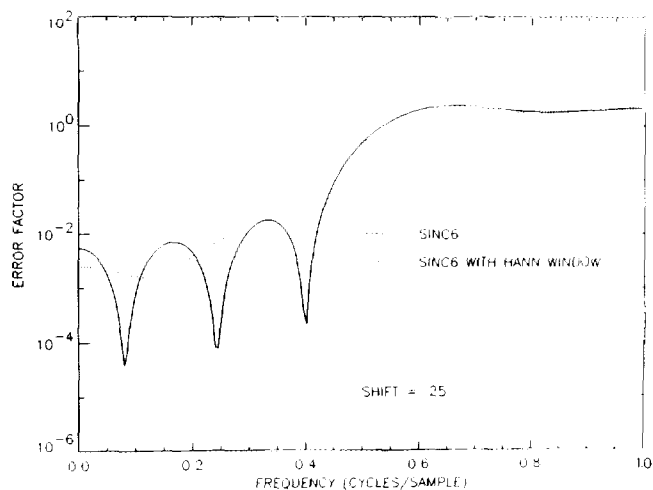

FIG. 7. Effect on $e_1^2(\nu)$ of windowing SINC-6 with a Hanning window.

a Hanning window to SINC-6, which is a standard technique for reducing sidelobe oscillations. This windowing is just the multiplication of the kernel by a raised cosine to taper its edges and remove Gibbs' overshoot in the transform [12]. The figure shows the effect on the error factor. It also is smoother after windowing, which effects a trade in excellent reproduction of certain frequencies for a more uniform performance.

Figures 6 and 7 contain instances of a general problem with truncated kernels—imperfect dc response. This is often an undesirable feature that is remedied by renormalizing the kernel. For SINC-N, at any given shift s the replacement

$$\mathrm{sinc}(n + s) \Rightarrow \frac{\mathrm{sinc}(n + s)}{\displaystyle\sum_{m=\lfloor -(N-1)/2\rfloor}^{\lfloor (N-1)/2\rfloor} \mathrm{sinc}(m + s)} \qquad (47)$$
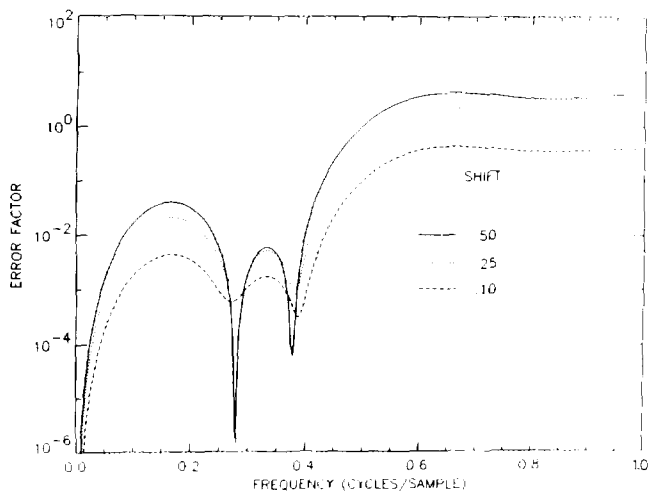
FIG. 8. $e_s^2(\nu)$ for SINC-6 after removal of dc error.

achieves the desired effect. Here [ ] means "integer part of." Note that for integer q and $\varepsilon \geq 0$, $[q + \varepsilon] = q$, even if q is negative.

The errors for $N = 6$ renormalized SINCs are shown in Fig. 8. They should be compared with Fig. 6c. The lowest-frequency local minimum has been shifted to dc, the second minimum has been shifted to the right, and the third's location is almost unchanged. The latter two minima are now also somewhat shift dependent, meaning that no longer is a continuous sinusoid at these frequencies especially well reproduced.

Figure 6 through 8 demonstrate the ease with which the performance of common modifications of SINC-N, or of any other interpolator, can be predicted by using the formula (41). However, this equation can also be
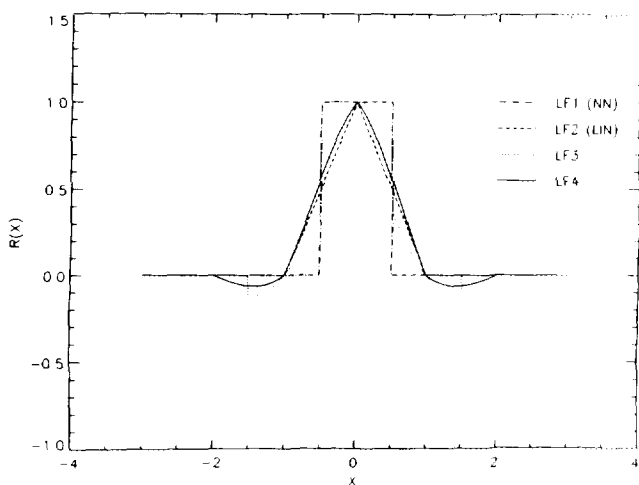

FIG. 9.   Kernels for Low-Frequency (LF) optimal interpolators.

used to define desired properties of an interpolator, whose kernel then becomes the solution of a system of equations. The Low-Frequency optimal problem is a good example.

LF Interpolators

Energy in imagery is usually concentrated at low spatial frequencies, and the LF interpolators are designed to work well generally for image-like spectra. For LF-N the error factor is constrained to be zero at dc, and the remaining degrees of design freedom, the number of which is determined by N, are used to make zero at $\nu = 0$ as many derivatives of the error as possible. This can be accomplished easily by requiring the complex error (40) to have zero derivatives. However, it is convenient first to multiply $E_s(\nu)$ by a phase factor $e^{i\omega s}$. Because of (36), this leaves unchanged the physically meaningful real error factor $e_s^2(\nu)$. Consequently, we can modify $E_s(\nu)$ of (40) to

$$E_s(\nu) \rightarrow \sum_n e^{-i\omega n} r(s + n) - e^{i\omega s}. \qquad (48)$$

The kth derivative of (48) at $\omega = 0$ is to be zero, meaning that

$$\sum_n (-n)^k r(s + n) = s^k. \qquad (49)$$

It is clear from (1) that (49) requires r to be that interpolator which perfectly reproduces the polynomials s^k ($k = 0, \ldots, N - 1$) from a nearby set of N sample points. (We generally restrict s to the values: $|s| \leq \frac{1}{2}$ (N odd) and $0 \leq s \leq 1$ (N even).) This is the well-known Lagrange interpolator, which fits a unique ($N - 1$)-order polynomial to N points. Note, however, that the Lagrangian polynomial so constructed is valid only on the central unit interval defined by the N sample points used to interpolate. As soon as the next unit interval is considered, a new sample point enters the finite sum in (49) and an old one leaves, so that the interpolated function f_i assumes the value of a new Lagrangian polynomial. This means that f_i, as well as its derivatives, can be discontinuous, a fact reflected in the Lagrangian kernels, $r(x)$. LF-1 is just NN, which has discontinuities at $x = \pm\frac{1}{2}$. LF-2 is identical to LIN, which has derivative discontinuities at $x = 0$, ±1. LF-3 has discontinuities at $x = 0$, $\pm\frac{1}{2}$, and $\pm\frac{3}{2}$, and LF-4 has derivative discontinuities at $x = 0$, ±1, and ±2.

Figure 9 illustrates these kernels for $N = 1$ through 4. The discontinuities at half-integers that appear in most interpolators with N odd usually make performance improvement marginal as $N \rightarrow N + 1$ with N even. Therefore, the remainder of this paper concentrates on comparisons with N even.
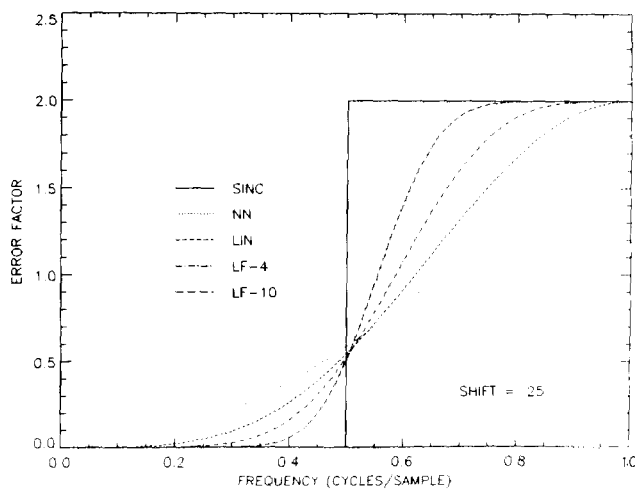
FIG. 10.   LF and SINC error factors $e_x^2(\nu)$ compared out to twice ν_{Nyq}.

For positive arguments the symmetric Lagrangian kernel is

$$r(n + s) = \prod_{\substack{[-(N-1)/2] \\ j \neq n}}^{[(N-1)/2]} (j + s) \Big/ \prod_{\substack{[-(N-1)/2] \\ j \neq n}}^{[(N-1)/2]} (j + n),$$

$$0 \leq n \leq \left[\frac{N-1}{2}\right], \left[\frac{N}{2}\right] - \frac{N}{2} \leq s \leq \left[\frac{N}{2}\right] - \frac{N}{2} + 1. \tag{50}$$

Again, [ ] means "integer part of."

Figure 10 compares the error factors (41) for LF interpolators to that of SINC at a standard shift $s = .25$. (Note that the SINC has finite support in frequency space, making (37) rather than (40) more convenient for computing the plotted error.) Although $e_x^2(\nu)$ can be made flatter at $\nu = 0$ by use of higher order interpolators, it becomes large beyond the Nyquist frequency. In the limit $N \to \infty$, LF-N approaches the "ideal" SINC interpolator. But clearly, the SINC is ideal only inband; for image energy beyond $\nu = \frac{1}{2}$, the polynomial interpolators are superior.

6. OPTIMAL INTERPOLATORS

If the form of the image power spectrum is known, the results of Section 3 can be used to find the N-point interpolator that minimizes the total squared error. Combining (41) and (34) results in

$$d_s^2 = R(0) - 2 \sum_n R(n + s)r(n + s)$$
$$+ \sum_m R(m) \sum_n r(n + m + s)r(n + s), \tag{51}$$

where

$$R(x) \equiv \int_{-\infty}^{\infty} \cos(x\omega)|f(\nu)|^2 \, d\nu. \tag{52}$$

Minimization of (51) with respect to the independent variables $r(n + s)$ results in an equation for the optimal N-point kernel.

$$R(n + s) = \sum_m r(n + s - m)R(m), \tag{53}$$

valid for those n for which $r(n + s)$ in (51) is nonzero, i.e., for the N values of n: $[-(N - 1)/2], \ldots, [(N - 1)/2]$.

A comparison of (53) with (1) shows that the former is equivalent to the requirement that r serve as a perfect interpolator for the function $R(x)$ for $x \in [-N/2, N/2]$, i.e., over the same support for which $r(x)$ is allowed to be nonzero. Equation (53) appeared in Ref. [2] as the condition minimizing a stochastic mean squared error. There the discrete sample values, here called $f(n)$, were assumed to represent an underlying stationary, ergodic, stochastic process. The function $R(x)$ was the autocorrelation function of that process, for which (52) is also valid if $|f(\nu)|^2$ is interpreted as the power spectrum of a stochastic process with the above properties. Equation (53) has now been shown to hold even when the $f(n)$ result from the sampling of a deterministic function, if $R(x)$ is interpreted according to (52) instead of more restrictively, as the autocorrelation function of a random process.

Because $R(m)$ is symmetric, (53) represents a particular class of matrix equations called Toeplitz. General procedures [13, 14], have been developed to solve these iteratively. Thus, in principle, $r(s + n)$ can always be found as a function of $2N$ numbers, $R(m)$ and $R(n + s)$, with $m = 0, \ldots, N - 1$ and $n = [-(N - 1)/2, \ldots, [-(N - 1)/2]$.

If these quantities are known, then (53) can be used to find the optimal $r(n + s)$. Notice that all the s-dependence in (53) is implicit, in the arguments of functions. This means that good interpolation might be possible without knowledge of s, as long as estimates of $R(m)$ and $R(n + s)$ are available. For example, if two images shifted by an unknown s are to be compared after one has been resampled, then R might be interpreted as an autocorrelation function, and periodograms [15] could be constructed from either image to estimate $R(m)$, and from both together to estimate $R(n + s)$. This technique introduces errors by using estimates of R, but it completely eliminates errors from a prior step that we have ignored: estimation of the shift s. The competition between these errors depends on the performance of registration methods and is not studied here.

Constant Spectrum

If $f(\nu)$ is constant (say, $= 1$) inband and zero otherwise, then the integral in (52) is calculable:

$$R(x) \to \text{sinc}(x). \tag{54}$$

Then (53) immediately yields

$$r(n + s) = \text{sinc}(n + s) \quad |n + s| \le N/2, \quad (55)$$

as claimed in the previous section. Thus, SINC-N, the truncated sinc interpolator, which is much eschewed in the literature [5, p. 35] because of the Gibbs' phenomenon in its Fourier transform, is in fact the optimal N-point interpolator for a flat inband image spectrum. SINC-N has been discussed in connection with Figs. 6–8.

Lorentzian Spectrum

Because the optimization equations for the stochastic problem are identical in form to (53), prior results can be applied immediately to the present problem. One particularly surprising result occurs for a Lorentzian profile, that is, a power spectrum of the form

$$|\hat{f}(\nu)|^2 \sim \frac{1}{\varepsilon^2 + \nu^2} \quad (\varepsilon \ge 0). \quad (56)$$

The transform of (56) is

$$R(x) \sim e^{-2\pi|x|\varepsilon}, \quad (57)$$

which corresponds to the autocorrelation function of a first-order autoregressive [AR(1)] stochastic process. Equation (56) is also a reasonable model for much natural imagery, independently of any probabilistic assumptions.

For the model (56), the optimal N-point ($N \ge 2$) solution can be verified by substitution into (53) as

$$r(s) = \frac{\rho^{s-1} - \rho^{1-s}}{\rho^{-1} - \rho}$$

$$r(s - 1) = \frac{\rho^{-s} - \rho^s}{\rho^{-1} - \rho} \quad (58)$$

otherwise $r(s + n) = 0$,

with $\rho \equiv e^{-2\pi\varepsilon}$ and $0 \le s \le 1$,

which has support two. That is, the optimal N-point solution is in fact a 2-point interpolator! This curious result, first shown in [2] for stochastic signals, also implies that for a $1/\nu^2$ power spectrum (the limit of (56) as $\varepsilon \to 0$), the optimal N-point interpolator is just linear interpolation, which can be verified as the limiting form of (58).

We now consider the general power-law spectrum:

$$|\hat{f}(\nu)|^2 \sim \frac{1}{\nu^p}. \quad (59)$$

Figure 11 shows $e_r^2(\nu) |\hat{f}(\nu)|^2$, which is the power spectrum of the error, that is, of the difference between interpolated
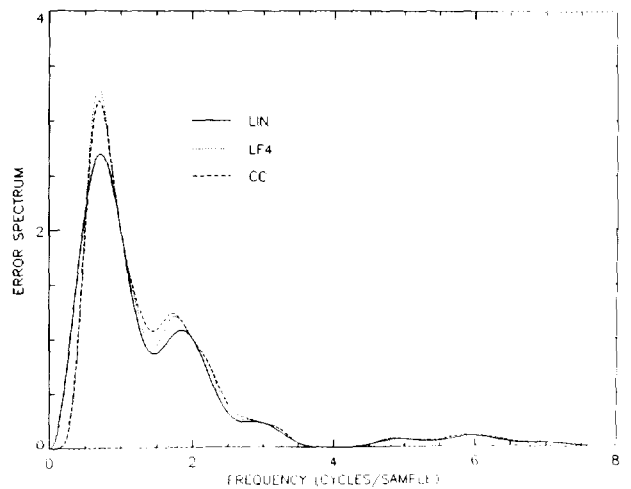

FIG. 11. Error spectra $e_r^2(\nu)|\hat{f}(\nu)|^2$ for LIN, LF-4, and Cubic Convolution (CC) for an image spectrum $|\hat{f}(\nu)|^2 = 1/\nu^2$. The shift s is 0.25 samples ($p = 2$ power law).

image and true value, for LIN, LF-4, and standard (smooth) Cubic Convolution (CC), for a $p = 2$ power spectrum, and at a shift of .25 samples. Although LIN is only a 2-point interpolator, it has been claimed to be optimal for $p = 2$. That is, according to (34) the integral of the error spectrum is smaller for LIN than for any other interpolator, regardless of the value of N. Figure 11 shows how LIN accomplishes this feat. When considered only out to the Nyquist frequency ($\nu = \frac{1}{2}$), LF-4 and CC are actually superior to LIN (by an rms factor of approximately 1.38). This remains true out to twice the Nyquist, over which frequency range CC and LF-4 each are computed to be a few percent better than LIN. It is only at higher frequencies that LIN recovers from its inband inferiority. If (59) with $p = 2$ is assumed to hold for all frequencies, then LIN is a few percent better than either LF-4 or CC.

Figure 11 explains a counterintuitive result implied by the fact that LIN is optimal for $p = 2$. For those signals, including imagery, for which the $p = 2$ power spectrum is a reasonable model, experience shows that many polynomial interpolators outperform LIN. The reason this does not contradict the optimality claim for LIN is that the model spectrum for which LIN produces the minimum error is usually invalid, for imagery, beyond the Nyquist frequency.

A power-law spectrum is commonly used to characterize one-dimensional slices of imagery, with p typically in the range $1 < p < 4$. Equation (59) is often assumed to hold in a piecewise sense, possibly with different values of p in different frequency ranges. Some low-frequency cutoff usually must be chosen to keep the total image energy (variance) finite, and commonly $\nu = \frac{1}{2}$ (Nyquist) is chosen as a practical high-frequency cutoff. The p that

is appropriate at low frequencies is often considered the most important because image energy tends to concentrate there. However, after interpolation, especially with polynomial kernels, the residual energy in a difference image often resides at higher frequencies, as exemplified in Fig. 11. Therefore, for extreme accuracy in interpolation, the high-frequency spectral content can be important also.

$p = 4$ Power Law Spectrum

As a third application of (53), we consider the case $p = 4$ in (59), which appears to be the limiting power of low-frequency divergence in natural imagery [16]. For $N = 2$, the solution to the Toeplitz equations is

$$r(s) = \frac{R(0)R(s) - R(1)R(1 - s)}{R^2(0) - R^2(1)}$$
$$\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (0 \le s \le 1). \quad (60)$$
$$r(s - 1) = \frac{R(0)R(1 - s) - R(1)R(s)}{R^2(0) - R^2(1)}$$

Although the integral in (52) defining R for the power spectrum of (59) diverges at low v for $p = 4$, the power spectrum of the error $e_v^2(v) \, |f(v)|^2$ is convergent at low frequency as long as

$$e_v^2(s) \sim v^{3+\delta}, \quad \text{with } \delta > 0. \quad (61)$$

This condition is violated by NN and (usually) DFT interpolators, for which $\delta = -1$. But even for LIN, the lowest nontrivial polynomial interpolator, $\delta = 1$. In practice, this means that the performance of NN and DFT interpolators can be highly dependent on the low-frequency cutoff that must exist in real imagery.

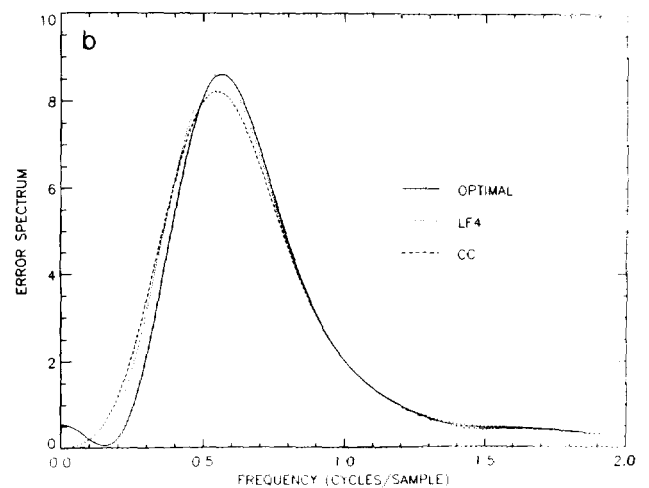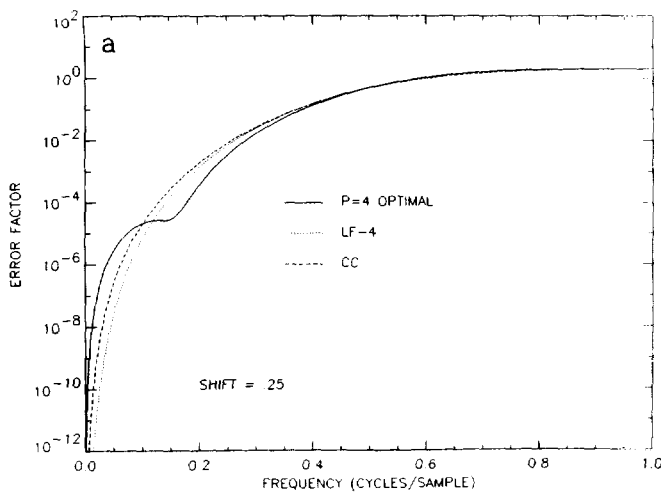The explicit calculations of the kernel in (60) need a

convergence mechanism for R, and so we solve instead for the power spectrum:

$$|\hat{f}(v)|^2 \sim \frac{1}{v^4 + \varepsilon^4}. \quad (62)$$

Then the limiting case $\varepsilon \to 0$ will give the $p = 4$ power-law result. For (62), $R(x)$ in (52) can be evaluated in closed form:

$$R(x) \sim \exp\left(-\frac{|\varepsilon x|}{\sqrt{2}}\right)\left[\cos\left(\frac{\varepsilon x}{\sqrt{2}}\right) + \sin\left(\frac{|\varepsilon x|}{\sqrt{2}}\right)\right]. \quad (63)$$

Then the minimum-error 2-point interpolator for the power spectrum in (62) can be calculated from (60) by using (63). In the limit $\varepsilon \to 0$, the solution (60) approaches linear, just as it does in the $p = 2$ case.

Similar calculations for $p = 4$ can be made for the 4-point optimum. The Toeplitz solution for the power spectrum (62) is quite complicated, but the limiting solution for $\varepsilon \to 0$ is simple:

$$r(s) = \frac{1}{5}(1 - s)(5 + 4s - 5s^2)$$

$$r(s + 1) = -\frac{1}{15}s(1 - s)(7 - 5s)$$

$$r(x) = 0 \quad |x| > 2 \quad\quad\quad (64)$$

$$r(x) = r(-x) \quad \forall x.$$

Figure 12a shows the error factors at $s = .25$ for three 4-point interpolators, and Fig. 12b shows the corresponding error spectra for a $p = 4$ input spectrum. Beyond $v = 1$ the curves are nearly identical, unlike the

FIG. 12. (a) $e_s^2(v)$ for three 4-point interpolators: the $(1/v^4)$-optimal, LF-4, and CC. (b) Error spectrum $e_s^2(v)|f(v)|^2$ for three 4-point interpolators and an image spectrum $|f(v)|^2 = 1/v^4$ ($N = 4$, $p = 4$ comparisons).

corresponding $p = 2$ curves. Nevertheless, it is again evident that the best interpolator (as measured by $\sqrt{d_s^2}$) depends on the frequency ν_t at which the $p = 4$ spectrum might be truncated to match realistic image data.

Integration of the Fig. 12b curves shows that over all frequencies, $\nu_t = +\infty$, the optimal solution (Eq. (64)) is only about 2% better than either LF-4 or CC; for $\nu_t = \frac{1}{2}$, it is 8 and 9% better, respectively. The optimal is still superior for $\nu_t = \frac{1}{4}$, but for $\nu_t = .1$, LF-4 produces 5.5 times less rms error than the $(\nu_t = +\infty)$ optimal. Evidently from Fig. 12b, $e_s^2(\nu)$ for the optimal varies as ν^4 for small ν, so that the error spectrum approaches a constant; for LF-4, $e_s^2(\nu)$ varies as ν^8.

Gaussian Spectrum

One final class of power spectrum is considered. A single point-like object or an assortment of such that are spatially separated by several times the blur imposed by a sensor's optics corresponds to a power spectrum proportional to the square of the sensor transfer function. Often a good model of this function is a Gaussian, and so our final example is

$$|f(\nu)|^2 \sim e^{-\omega^2 \sigma^2}. \tag{65}$$

This corresponds to a system point spread function of

$$p(x) \sim e^{-(x^2/2\sigma^2)}. \tag{66}$$

We consider three cases: $\sigma = 1$, $\frac{1}{2}$, and $\frac{1}{3}$, measured in samples. The first case represents an imaging system that samples at nearly the Nyquist rate; the third represents one that undersamples, such as a staring sensor with a 2-D array of focal plane detectors [20].
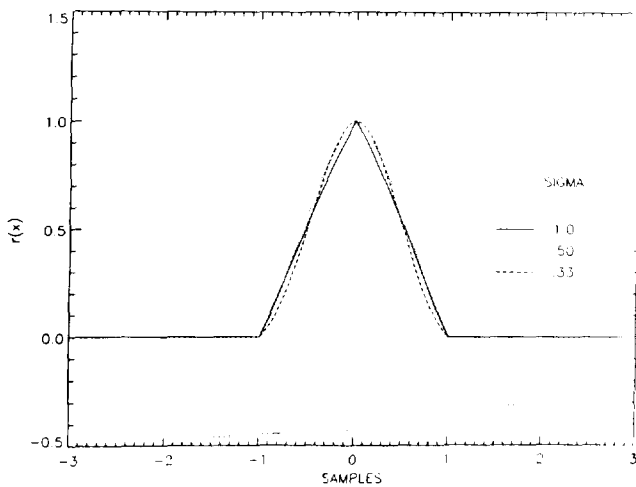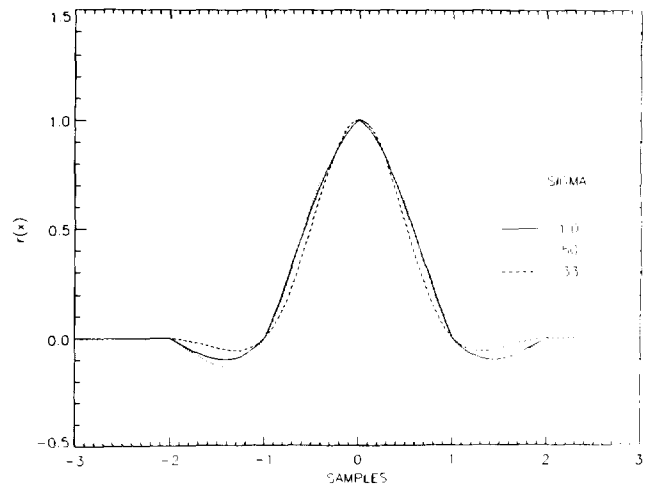

FIG. 14. Four-point ($N = 4$) Gaussian-optimal kernels.

The solution of (53), using (52) with the spectrum (65), is tedious but straightforward. For $N = 2$, the optimal kernel is

$$r(s) = \frac{\gamma^{s^2} - \gamma^{[1+(s-1)^2]}}{1 - \gamma^2} \quad \text{with } \gamma \equiv e^{-(1/2\sigma)^2} \quad (0 \le s \le 1), \tag{67}$$

which is plotted in Fig. 13 for the three values of σ. Like all kernels derived in this report, it is symmetric. Figure 14 plots the analytically complicated $N = 4$ results. Figure 15 shows the image power spectra corresponding to the three values of σ; these spectra have been scaled so that the total energies, i.e., image variances, are identical. Figures 16 and 17 show that $N = 2$ and $N = 4$ error factors, respectively, both for a shift of .25. Note that for $\sigma = .33$, the performance gain as N changes from 2 to 4

FIG. 13. Two-point ($N = 2$) Gaussian-optimal kernels for three values of the standard deviation α.
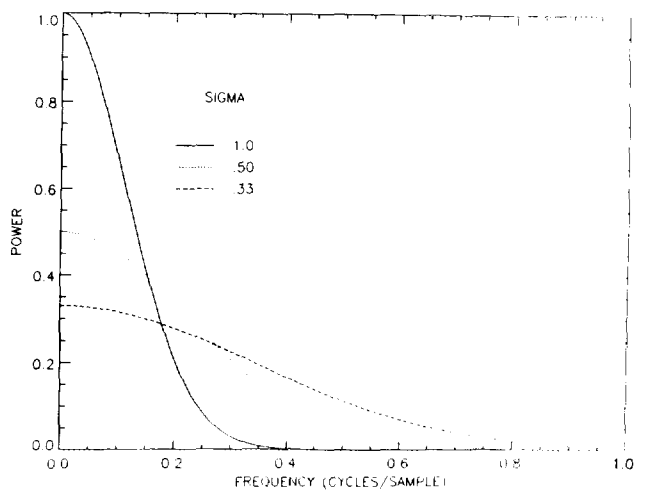

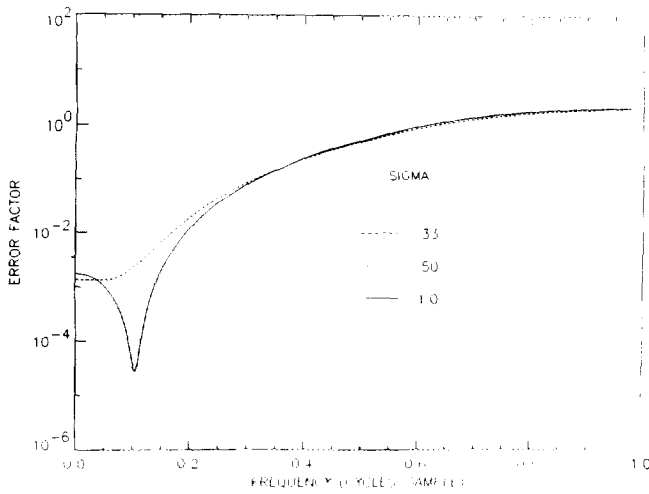FIG. 15. Gaussian power spectra $|f(\nu)|^2$ with identical total energies.

FIG. 16.   Error factors $e_s^2(\nu)$ for 2-point (N = 2) Gaussian-optimal kernels; s = ¼.
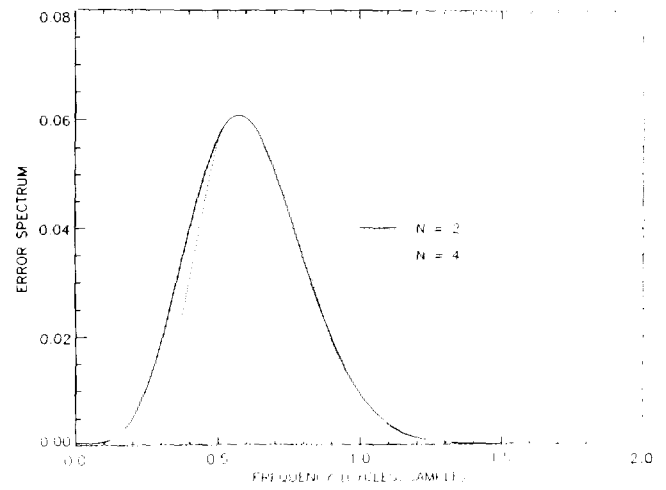

FIG. 18. $e_s^2(\nu)|f(\nu)|^2$ for 2- and 4-point Gaussian-optimal interpolators acting on Gaussian spectrum with σ = .33.

is small. The square roots of the integrals over the error spectra (Fig. 18) are proportional to the rms errors, and the difference of these proves to be less than 1 percent. On the other hand, if the spectra are truncated at the Nyquist frequency, the error reduction is 13%.

Increasing σ, which is measured in samples, can be thought of as either raising the sampling rate or compressing the spectrum toward dc, as shown in Fig. 15. For σ = .5 (Fig. 19) the rms error is reduced by approximately 16% as N changes from 2 to 4; for σ = 1 (Fig. 20) it is smaller by a factor of 3.3. The failure of the 4-point method to greatly outperform the 2-point for σ = .33 reflects the difficulty of designing an interpolator with a consistently small error factor over a wide range of frequencies.

Our final comparison is among the various values of σ for a given N. As noted above, the rms error (for ν_l = $+\infty$)

is nearly the same for N = 2 and N = 4 when σ = .33. Either case can, therefore, serve as a common baseline. For N = 2, the rms error is reduced by a factor of 1.9 as σ changes to .50, and by another factor of 3.7 as σ changes to 1.0. For N = 4, the corresponding reductions are 2.58 and 10.3, respectively. Optimal interpolation achieves its greatest gains for narrowband spectra.

Hybrid Designs

Figures 11 and 12b demonstrate the polynomial interpolators can be guilty of overkill at low frequencies. Optimal interpolators achieve superior performance by leaving some residual energy near dc. Similar results are evident in the Gaussian error spectra of Figs. 19 and 20. On the other hand, if the image spectrum actually ends short of
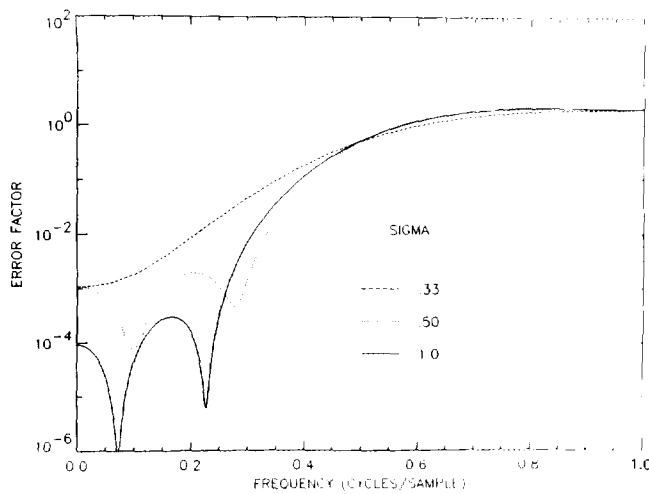

FIG. 17. $e_s^2(\nu)$ for 4-point (N = 4) Gaussian-optimal kernels.
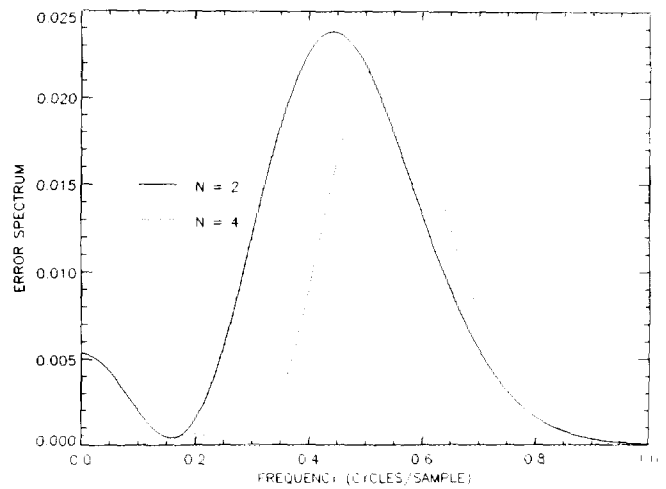

FIG. 19. $e_s^2(\nu)|f(\nu)|^2$ for Gaussian-optimal interpolators acting on Gaussian spectrum with σ = .50.
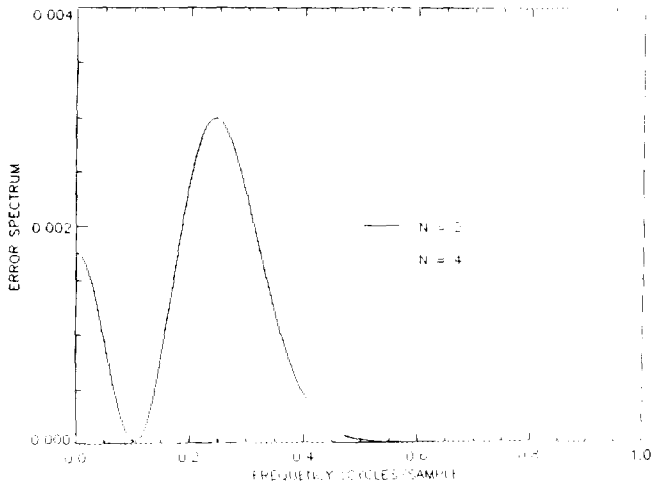
FIG. 20. $e_s^2(\nu)|f(\nu)|^2$ for Gaussian-optimal interpolators acting on Gaussian spectrum with $\sigma = 1.0$.
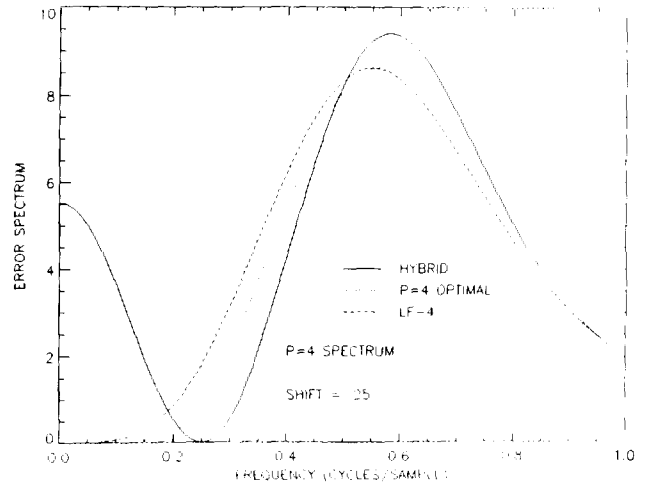

FIG. 22. Error spectra $e_s^2(\nu)|f(\nu)|^2$ for 4-point interpolators; $|f(\nu)|^2 = 1/\nu^4$.

the Nyquist frequency, then the $\nu_t = +\infty$ optimal kernels can perform poorly because their domain of superior performance is often at high frequencies. Therefore, both polynomial and $\nu_t = +\infty$ optimal solutions have defects when applied to truncated spectra. This problem could be remedied by solving (53) and (52) for a spectrum truncated at the Nyquist limit. Such solutions are quite complicated analytically.

An alternative approach uses (40) to define interpolators that sacrifice some of the low-frequency performance that is present in LF-4, for example, for smaller high-frequency errors. The error factor for LF-4 varies as ν^8 near dc, but image power spectra seldom diverge as fast as $1/\nu^4$. Therefore, the error spectrum is usually of order greater than ν^4, rapidly dying at dc.

By using (40), two degrees of design freedom in a 4-point

FIG. 21. $e_s^2(\nu)|f(\nu)|^2$ for 4-point interpolators; $|f(\nu)|^2 = 1/\nu^2$.

interpolator can be spent to set the error factor to zero at, say, half the Nyquist frequency, $\nu = \frac{1}{4}$. The remaining two degrees of freedom can be used to constrain to the value zero: (a) the error factor at dc, and (b) the first derivative of the complex error $E_s(\nu)$ at dc. Two degrees of freedom are required for the first step because two phases of a sinusoid can be present in an image for $\nu = \frac{1}{4}$. The dc conditions (a) and (b) make $(d/d\nu)^n[e_s^2(\nu)]|_{\nu=0} = 0$ for $n = 0, \ldots, 3$, just as for linear interpolation.

Figure 21 shows error spectra out to the Nyquist limit that result from applying three interpolators to a $1/\nu^2$ spectrum. All these can be regarded as 4-point interpolators because, as was shown previously, LIN is actually the optimal N-point interpolator for a $1/\nu^p$ spectrum when $p = 2$. Above $\nu = .2$, benefits of the above hybrid design are evident.

Another set of error spectra, now for a $1/\nu^4$ image spectrum, is shown in Fig. 22, in which the optimal 4-point $p = 2$ interpolator LIN has been replaced by the optimal 4-point $p = 4$ interpolator (64). Here the sacrifice in low-frequency performance required to construct the hybrid is severe. The image power is too divergent near dc, and the hybrid solution is inferior.

Figure 23 shows the final hybrid example. The image spectrum is Gaussian with $\sigma = .33$, and the $\nu_t = +\infty$ optimal is compared with LF-4 and the above hybrid. LF-4 overkills low frequencies, the optimal achieves its largest gain beyond the Nyquist limit, and the hybrid is the best choice for an image spectrum truncated at the Nyquist.

Nulls in the error or its derivatives could be forced at other frequencies besides $\nu = \frac{1}{4}$, or even instead of at $\nu = 0$. For higher N, larger numbers of such constraints can be imposed, all implemented by setting (40) or its derivatives to zero at the selected frequencies and solving
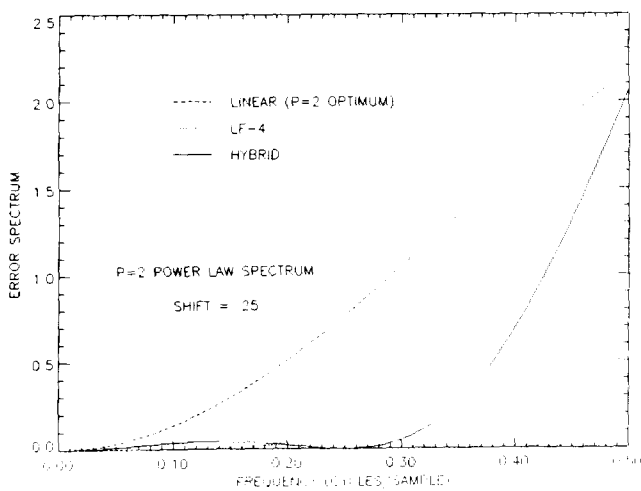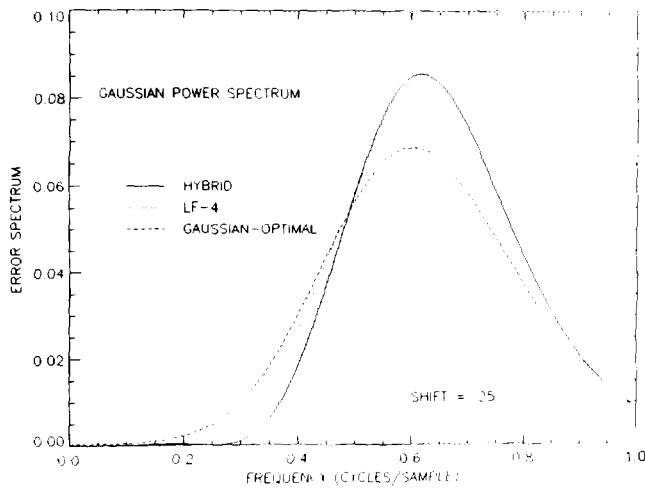
FIG. 23. Error spectra as in Fig. 22, but with $\sigma = .33$ Gaussian input spectrum $|f(\nu)|^2$.

for the $r(n + s)$. For example, DFT-N can be defined by setting (40) to zero at $N/2$ equally-spaced ($\Delta\nu = 1/N$) frequencies, starting at dc. ($\nu = 0, \frac{1}{2}$ are counted as "half" a frequency each, because only one phase of a sinusoid at either of these frequencies is visible to the sampling process.) In practice the selection of such constraints is dictated by a knowledge of the types of power spectra over which the interpolation methods are to be applied.

7. BEYOND INTERPOLATION

The optimal interpolators described above for orders $N = 2$ and $N = 4$ usually reduce errors by factors of less than two when compared to conventional interpolators of the same order. Larger reductions are possible by increasing N, and the above methods provide a means for designing such interpolators. For example, the error factor for LF-10 was shown in Fig. 10, and Fig. 24 compares the error spectra after using LF-2, -4, -10 on a $1/\nu^3$ spectrum truncated at $\nu = \frac{1}{2}$. LF-10 produces a smaller rms error than LIN by a factor of 2.3. Reductions with increasing N generally occur because N equals N_d, the number of design degrees of freedom for the interpolator. However, for certain applications, N_d can be increased more effectively, without increasing N.

In change detection, two digital images are often compared by subtraction after one has been overlaid on the other, i.e., resampled. Our analyses are well suited to this technique. The mean squared error d_x^2 becomes a measure of residual clutter after subtraction.

The change to be detected can be an extensive one, such as a region of blood flow on an angiogram, or large-scale agricultural evolution as seen from space. It can be a local change, such as the motion of a small object. The

general goal is to combine the two images so that only such changes persist.

The N-point interpolators described above are actually digital filters operating on an image. To implement change detection with interpolation, a second image is used, but only in the subtraction step. It can be thought of as a template for gauging interpolation accuracy. However, by digitally filtering both images before subtraction, N_d can be increased to nearly $2N$. This dual filtering introduces a new element into the analysis, consisting of constraints on the individual filters to preserve the change that is to be detected. For example, for extensive changes an appropriate constraint can be to require perfect dc response of each filter. For detection of small moving targets of known shape, the filters can be constrained to conserve target energy or peak amplitude.

Figure 25 illustrates performance results for two examples of this dual difference filtering. These should be compared to Fig. 24 (note the scale change). The dual LF-4 is the analog of the interpolating LF-4. The constraint imposed on its single-image filters was preservation of dc, i.e., of local mean values. The dual hybrid is analogous to the interpolating hybrid described above, with the important new feature that errors at $\nu = \frac{1}{2}$ are also perfectly suppressed, an impossibility with interpolation. This particular dual hybrid was constrained to preserve the peak amplitude of a small Gaussian target (corresponding to $\sigma = 1$ in the Section 6 analysis). The dual LF-4 and hybrid yield smaller rms errors than even LF-10 (Fig. 24), by factors of 5.9 and 12.1, respectively.

Because the second image is now also filtered before subtraction, the difference image is no longer a measure of interpolation error. Dual filtering methods relinquish the goal of interpolation for the sake of maximizing a signal-to-clutter ratio. These ideas are developed further in a companion paper [17].
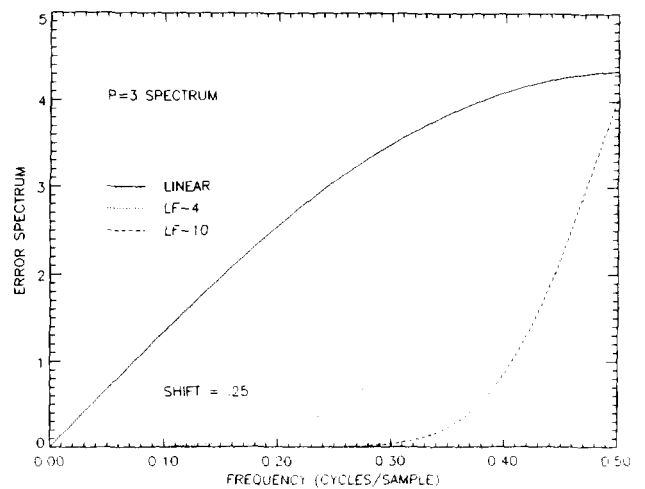

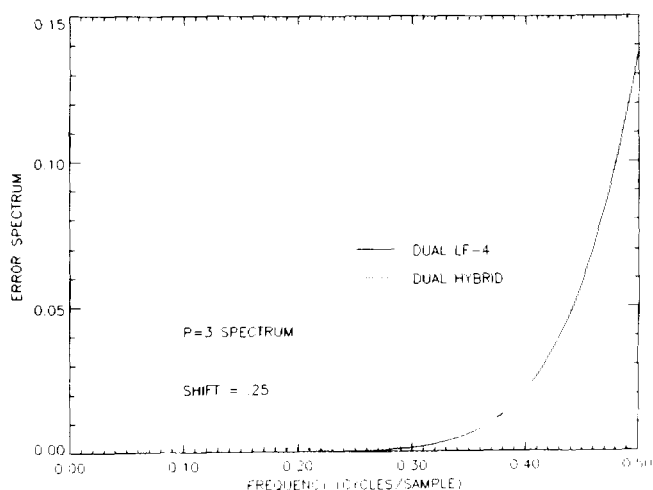FIG. 24. Error spectra for LF-2, -4, -10, and $|f(\nu)|^2 = 1/\nu^3$.

FIG. 25. Error spectra for Dual Difference Filtering methods and image spectrum $|f(\nu)|^2 = 1/\nu^3$.

8. SUMMARY

This paper provides tools for evaluating and designing local interpolators. Mathematical theorems are derived that express the spectral dependence of the root mean squared interpolation error. The formula is applicable even in the presence of image aliasing, if the proper averaging is understood. The error spectrum can be written as a product of the continuous image power spectrum and an error factor that depends only on the interpolation kernel and the shift. Forms for the error factor are derived that are useful for interpolators that are of finite extent either in space or in frequency.

A general optimization equation for the minimum-error interpolator is derived and solved for various model spectra: constant inband, Lorentzian, power law, and Gaussian. The interpretation of the equation for stochastic signals is also provided. When image spectra are modeled with analytic forms extending over infinite frequency range, the optimal solutions often result in residual error spectra that are concentrated around the Nyquist limit. Such solutions are sometimes inferior to polynomial methods when applied to Nyquist-truncated versions of the model spectra.

Methods are described for customizing the performance of an N-point interpolator when the precise form of the image spectrum is unknown. Performance error can be made zero at selected frequencies, the number of which is limited only by the order of the interpolator. Such hybrid methods can also be designed to keep the error small in selected regions of the spectrum by concentrating many zeros of the error factor into a small range, such as for LF-N, in which all zeros are collapsed to $\nu = 0$. The hybrids are also valuable suboptimal methods when the form of the ideal interpolator is cumbersome, as often happens with image power spectra that are modeled as truncated forms of analytic functions.

Finally, a generalization of interpolation is described for the change detection application. By sacrificing image fidelity, this dual difference filtering can achieve much higher target-to-noise ratios than interpolation.

REFERENCES

1. A. Schaum and M. McHugh, *Analytic Methods of Image Registration: Displacement Estimation and Resampling*, NRL Report 9298, Naval Research Laboratory, Washington DC. 1991.
2. G. W. Wornell, D. S. Kauffman, and B. Sharpe, Minimum mean-square error resampling for remotely sensed imagery, in *12th Canadian Symposium on Remote Sensing*, IEEE No. **89CH2768-0**, IEEE, New York, 1989.
3. A. J. Mord, N. H. Endsley, E. Ramberg, and H. J. Reitsema, Electronic image alignment: implementation and applications to imaging system design. *Opt. Engrg.* **24**(3), 1985, 507–515.
4. J. Owczarczyk, W. J. Welsh, and S. Searby, Performance analysis of image registration techniques, in *Third Int. Conf. on Im. Proc. and Applic.*, 1989, pp. 10–13.
5. J. A. Parker, R. V. Kenyon, and D. E. Troxel, Comparison of interpolating methods for image resampling, *IEEE Trans. Med. Imaging* **MI-2**(1), 1983, 31–39.
6. A. Ferraioli, Review of Clinical Applications of Cardiac Digital Angiography, *IEEE Trans. Med. Imaging* **9**(1), 1990, 113.
7. K. P. Prasad and P. Satyanrayana, Fast interpolation algorithm using the FFT, *Electron. Lett.* **22**, 1986, 185–187.
8. S. Shlien, Geometric correction, registration, and resampling of LANDSAT imagery, *Canad. J. Remote Sensing* **5**, 1979, 74–89.
9. S. K. Park and R. A. Schowengerdt, Image sampling, reconstruction, and the effect of sample-scene phasing, *J. Appl. Opt.* **21**(17), 1982, 3142–3151.
10. S. K. Park and R. A. Schowengerdt, Image reconstruction by parametric cubic convolution, *Comput. Vision Graphics Image Process.* **23**, 1983, 258–272.
11. D. Fraser, Interpolation by the FFT revisited—An experimental investigation, *IEEE Trans. Acoust. Speech Signal Process.* **37**(5), May 1989, 665–675.
12. G. Wolberg, *Digital Image Warping*, pp. 139–140, IEEE Computer Society Press, Los Alamitos, CA, 1990.
13. E. A. Robinson and S. Treitel, *Geophysical Signal Analysis*, p. 163, Prentice–Hall, Englewood Cliffs, NJ, 1980.
14. W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical Recipes*, pp. 47–51, Cambridge Univ. Press, Cambridge, 1986.
15. S. M. Kay, *Modern Spectral Estimation: Theory and Application*, Prentice–Hall, Englewood Cliffs, NJ, 1988.
16. M. S. Farber, S. J. Hemple, and B. A. Eckstein, Characterization of IRAMMP data using scene registration, in *Proc. IRIS Specialty Group on Targets, Backgrounds, and Discrimination, March 1991*, Vol. 1, pp. 89–114, Environmental Research Inst. of Ann Arbor, Michigan.
17. A. Schaum, *Dual Difference Filtering: A Replacement for Interpolation and Subtraction to Detect Changes in Misregistered Signals*, NRL Report 9522, Naval Research Laboratory, Washington DC, Nov. 30, 1992.
18. W. Rogosinski, *Fourier Series*, Chelsea, New York, 1959.
19. R. N. Bracewell, *The Fourier Transform and Its Applications*, McGraw–Hill, New York, 1978.
20. A. Schaum, *Principles of Interpolator Evaluation and Design*, NRL Report 9356, Naval Research Laboratory, Washington DC, Nov. 15, 1991.