Naval Research Laboratory
Washington, DC 20375-5000

NRL Report 9298

Analytic Methods of Image Registration: Displacement Estimation and Resampling

A. SCHAUM AND M. MCHUGH

Advanced Concepts Branch
Optical Sciences Division

February 28, 1991

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	February 28, 1991	Final 1/85 to 1/90

4. TITLE AND SUBTITLE

Analytic Methods of Image Registration: Displacement Estimation and Resampling

5. FUNDING NUMBERS

PE - 0602111
PR - RA11W53

6. AUTHOR(S)

A. Schaum and M. McHugh

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
Washington, DC 20375-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

NRL Report 9298

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Technology
Arlington, VA 22217-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Registration algorithms are developed and evaluated by using oversampled scanning imagery directly and staring imagery with a subpixel autocorrelation model. A Fourier transform-based method, Phase Correlation, is enhanced to remove edge effects, to accommodate nonintegral shifts, and to resolve ambiguity in the interpretation of its output. The result is shown to be accurate over a wide range of misregistration and to be capable of detecting cloud parallax and other relative-motion effects. A family of gradient-based methods is also derived and is shown to include older methods: the Image Displacement Estimation Algorithm (IDEA) and the Gradient Estimation Method (GEMS). One member of the family, the Canonical Gradient Estimate (CAGRE), proves to be generally superior, as long as the noise-to-clutter ratio is not unusually large compared to typical values for Earth backgrounds. Resampling methods are also tested: linear, spline, phase-shifting, and cubic convolution; cubic convolution performs marginally better for all values of shift. The net result here is a substantial improvement in the state of the art of analytic image registration, which may permit the use of frame differencing as a moving-target indicator in the most cluttered natural backgrounds.

14. SUBJECT TERMS			15. NUMBER OF PAGES
Image registration	Phase correlation	Resampling	37
IR target detection	Frame differencing	Velocity estimation	16. PRICE CODE
Displacement estimation			

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std 239-18
298-102

CONTENTS

ANALYTIC METHODS OF IMAGE REGISTRATION: DISPLACEMENT ESTIMATION AND RESAMPLING

1. BACKGROUND

The problem motivating this study is the autonomous detection and tracking of unresolved targets in images collected by an IR sensor. With a low error rate, a target must be detected and tracked within a few frame times of its appearance in the sensor's field of view.

Historically, the limiting factor in the performance of IR detection systems has been the high rate of false alarms induced by background clutter. A natural approach to reducing these detection errors is to exploit the temporal redundancy of natural imagery, which is large if interframe times are short. In practice, often these times *can* be made short enough, while still allowing targets of interest to traverse at least one sample. This permits the use of frame differencing as a target detection algorithm.

After frame differencing, the signature of a moving target is a dipole, and the residual signal from a pair of well-registered backgrounds is sensor/photon noise. However, if drift or jitter is present, background clutter leaks through a differencer and degrades performance. This report discusses methods for removing the residual misregistration that persists after hardware-based methods have been exhausted, or are not used.

2. REPORT OUTLINE

We first consider two methods of estimating the amount of misregistration: Phase Correlation (PC) and the Image Displacement Estimation Algorithm (IDEA) [1]. These appear to be the leading software-based candidates meeting the constraints imposed by the mission described above. The concern here is not with registration algorithms that require a large number of images — frame differencing requires only two. Therefore, the previously reported Pseudoregistration [2], which is a polynomial interpolation method with IDEA as a component, is not considered; nor is the Subspace Projection Algorithm [3] which is a geometric construction that maps all the measurements into a lower dimensional subspace [4]. PC and IDEA require only a pair of images for estimating the displacement, and they are appropriate for a wide variety of applications where the number of available images is small. Phase Correlation is discussed in detail in Section 3.

Section 4 is devoted to gradient-based methods, including IDEA. We derived the general analytical form of all such estimates and, in the process, we explain the past deficiencies apparent in some experimental applications of IDEA. This leads naturally to the defining and analysis of an expanded family of gradient-based estimates. The performance of all these algorithms is predicted as a function of the image autocorrelation function, with image-independent additive white noise allowed. As a corollary application of the analytical approach, we consider previously reported

Manuscript approved August 20, 1990.

efforts to register data from the Air Force's High-Resolution Calibrated Airborne Measurements Program (HiCAMP) by using a method called GEMS. This earlier work had met with only partial success, but no explanation was ever offered for its imperfect performance. Appendix A interprets its results in terms of the Section 4 model.

Section 5 introduces a method that eliminates the need to interpolate digital imagery in order to test displacement estimates, such as PC and IDEA, for subpixel shifts. It is based on the use of oversampled scanner data.

Finally, Section 6 examines the second half of the registration problem, resampling, which is interpolation between integral samples after estimation of a (perhaps) nonintegral displacement. Several schemes are compared and the leading overall candidate is found to be a method called Cubic Convolution (CC). However, the margin of its superiority is not great, and the best resampler choice depends importantly on other factors, such as ease of integration with the estimation stage of registration.

3. THE PHASE CORRELATION METHOD

The Phase Correlation method (PC) appeared in the open literature as early as 1975 [5], where its superiority to conventional correlation was noted. It has been referenced occasionally in the literature and recently was used by the authors, working for the Navy's Infrared Analysis Modeling and Measurements Program (IRAMMP) as a sensor diagnostic and in target tracking work [6]. However, its full potential apparently has never been realized. In particular, its capability for recognizing parallax effects is demonstrated here. Cloud parallax as seen from a satellite platform has been shown through simulation at the Naval Research Laboratory (NRL) to be capable of significantly degrading the performance of a frame-differencing signal processor. It is also the principal temporal processing problem faced by airborne infrared search and track systems.

Phase Correlation is based on the principle that a translation of the coordinate frame used to define a mathematical function is reflected in the Fourier domain purely as a linear shift in the phase. That is, if x_n describes the intensity at the n^{th} sample in the first image, and

$$y_n = x_{n-s} \tag{1}$$

represents the same sample from the next image frame, which is shifted by s samples,* then

$$Y_k = X_k \, e^{-i2\pi sk/N} \tag{2}$$

where X_k (Y_k) is the discrete Fourier transform (DFT) of x_n (y_n). For example

$$X_k = \sum_{n=0}^{N-1} x_n \, e^{-i2\pi nk/N} \tag{3}$$

with N the number of samples in the image.

*The simpler one-dimensional notation is used, but the results generalize naturally to two image dimensions.

Equation (2) holds generally as $N \to \infty$ (with s/N, the absolute shift, held constant) because the DFT becomes the continuous Fourier transform. For finite N, Eqs. (1) and (2) hold only if s is an integer and x_n is periodic:

$$x_{n+N} = x_n. \tag{4}$$

This condition means that Eqs. (1) and (2) are strictly true only if the images have been derived from a scene that is periodic. Although real imagery seldom has this property, PC, which was inspired by Eq. (2), is still a surprisingly accurate method, even when $|s|$ is a large fraction of N.

Guided by Eq. (2), the Fourier phase difference ϕ_k could, in principle, be found by computing the complex logarithm of the ratio:

$$\frac{X_k}{Y_k} = e^{i\phi_k}. \tag{5}$$

In the ideal case, $\phi_k = 2\pi sk/N$. Therefore, the shift s could be estimated from the slope of a straight-line fit of ϕ_k vs k. However, Eq. (5) admits 2π ambiguities in the value of ϕ_k, so s can be determined at any frequency k only up to an additive multiple of N/k. Methods to "unwrap" the phase difference have been developed [7], but they are generally computationally intensive procedures that, furthermore, have not been extended to two dimensions.

An elegant way to remove the ambiguity in phase is to exponentiate it, i.e., maintain the phase information in a form like the right-hand side of Eq. (5). More precisely, one forms the "phase-correlation function" p defined as

$$p_n = \frac{1}{N} \sum_{k=0}^{N-1} \hat{X}_k \hat{Y}_k^* e^{i2\pi nk/N}, \tag{6}$$

where * means complex conjugation and ^ means "normalized," e.g.,

$$\hat{X}_k = \frac{X_k}{|X_k|}. \tag{7}$$

The phase-correlation function in Eq. (6) is the cross-correlation function of the "whitened" scenes $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ that result from normalizing their Fourier transforms $\boldsymbol{X}$ and $\boldsymbol{Y}$ to unit magnitude. The standard cross-correlation function c_n, which equals the right-hand side of Eq. (6) if the ^s are removed, has long been used to estimate the relative image shift s because the peak value of c_n occurs theoretically at $n = s$. However, the peak in $\boldsymbol{c}$ has an intrinsic, scene-dependent breadth that usually extends for many samples. The width of the peak is related to some typical "correlation length" in the image.

Just as for the function $\boldsymbol{c}$, the maximum of $\boldsymbol{p}$ also occurs at position s, but its peak-width is small (approximately one pixel*) and is independent of scene statistics. Furthermore, if s is an

*The width can be greater if the data are undersampled.

integer, then $p_n = \delta_{n-s}$ (Kronecker delta function). These idealized properties of both $\boldsymbol{c}$ and $\boldsymbol{p}$ are degraded for real imagery because of edge effects, noise, temporal background evolution, distortions, etc., but $\boldsymbol{p}$ proves to degrade much more gracefully.

In Fig. 1(a), c_n is plotted for a pair of subimages with relative displacement $s = 35$ columns (1/3 of the subimage size), extracted from a single infrared image collected by a high-flying aircraft. Several local maxima correspond to accidental alignment of similar features from different parts of the terrain images. These peaks are sometimes larger than the peak at the true shift.

Figure 1(b) shows p_n for the same images. Because the energy in natural background imagery typically is concentrated in the lower spatial frequencies, whitening (Eq. (7)) produces a relative increase in the high-frequency content of the images. This explains the noisiness of $\boldsymbol{p}$ compared to $\boldsymbol{c}$. However, the prominence of the peak in $\boldsymbol{p}$ persists, and this is the hallmark of the PC method: surplus signal-to-noise is available for coping both with imagery that is far from the theoretical ideal and with shifts so large that only small portions of the two images are common.

As an example of the robustness of PC, Fig. 2 shows the result of applying it to a pair of images collected with an airborne sensor (HiCAMP) aimed at terrain through a hole in a cloud deck. The ground, represented by the smaller peak, appears to move because of motion of the sensor platform. High clouds have a different relative motion that is reflected by the appearance of another distinct peak. For such imagery, cross-correlation always broadens and merges the two peaks, hiding parallax effects.

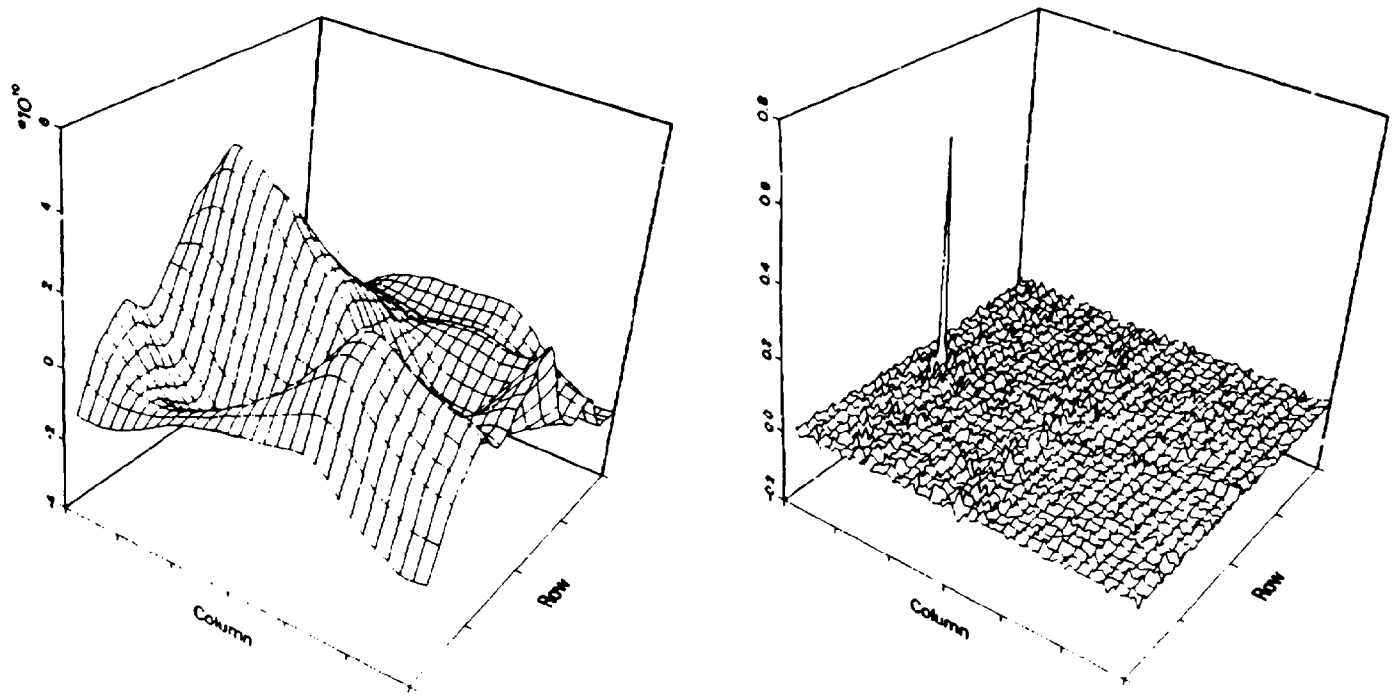
The double-peak effects can also be seen when the two coherently moving parts of an image are the background and a target. Figures 3(a) and 3(b) show reproductions of imagery from a ground-based Background Measurements and Analysis Program (BMAP)* experiment [8] conducted at the Naval Research Laboratory in which a commercial aircraft appears along with a few background clouds. Figures 3(c) and (d) depict the resulting $\boldsymbol{c}$ and $\boldsymbol{p}$. The central peak of $\boldsymbol{p}$ informs us that a large fraction of the scene energy is stationary (the background), while the displaced peak indicates the presence of a moving object (the aircraft) and simultaneously permits an estimate of its velocity.

Properties of the Phase-Correlation Function

Note that in Eq. (6) $\hat{X}_k \hat{Y}_k^*$ is computed, rather than X_k / Y_k, as suggested by Eq. (5). In the ideal case (noiseless periodic scene and s = integer) the two expressions are equivalent, but for real data, the former has the advantage in that it remains normalized to a constant (unit) magnitude, i.e., whitened. The whitening process gives the phase-correlation function $\boldsymbol{p}$ several advantages over the cross-correlation function $\boldsymbol{c}$.

The most striking property of $\boldsymbol{p}$ is the sharp, scene-independent peak discussed earlier. However, $\boldsymbol{p}$ is also invariant with respect to DC shift in image intensity, as well as any intensity scaling that depends only on spatial frequency. This implies, for example, that for imagery produced by different detectors, PC is insensitive to offset and (fixed) gain errors in calibration. Also, PC is unaffected by differences in the illumination of a pair of images that are seen only in reflected light. Registration in different spectral bands is also feasible because of the scaling invariance.

*The Navy's Background Measurements and Analysis Program was sponsored by the Office of Naval Technology.



(a) (b)

Fig. 1 — Comparison of (a) the cross-correlation function with (b) the phase correlation function for a pair of HiCAMP images shifted by 35 pixels. The phase correlation function has a sharp peak at the correct location (35,0), but the cross-correlation function has several broad peaks and a maximum at (34,0), which is one full pixel removed from the correct shift

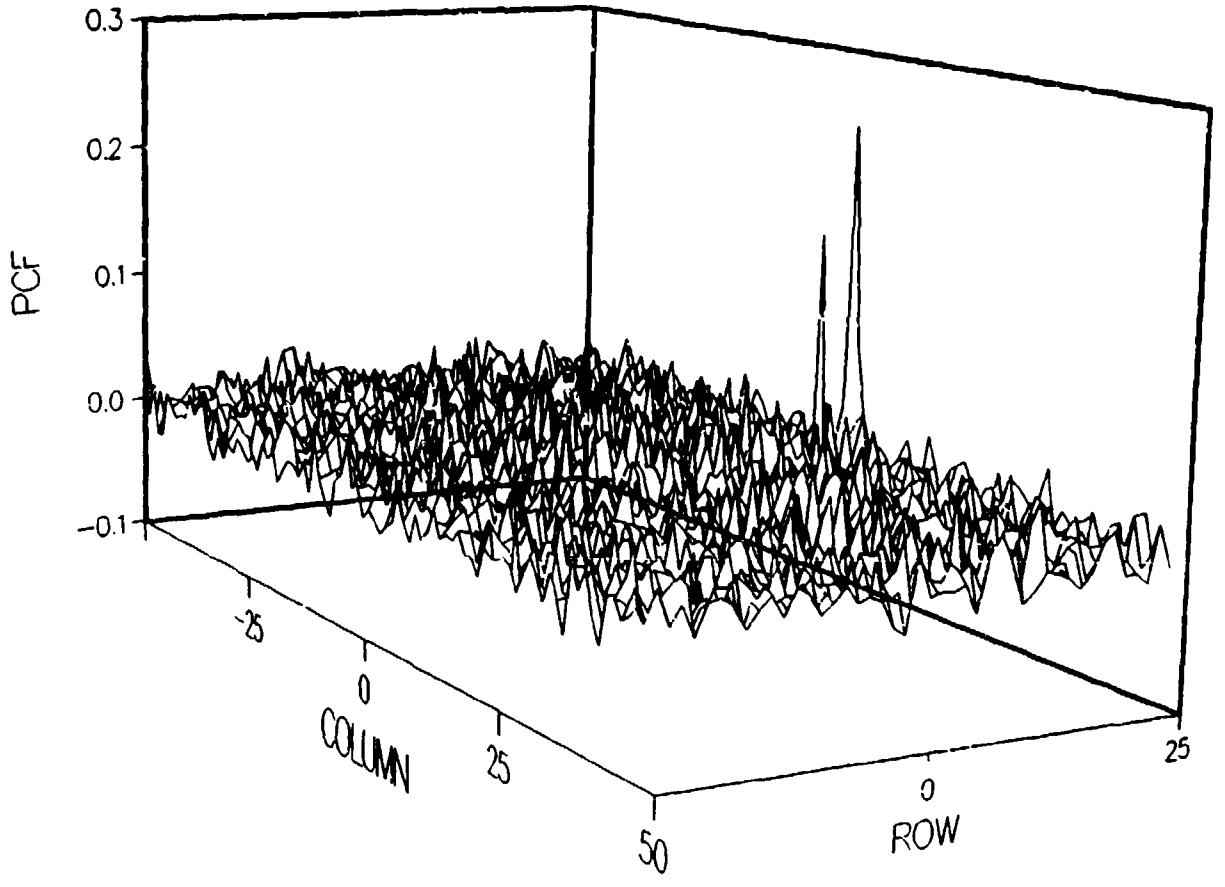


Fig. 2 — "Superresolution" of image motions with PC. The Phase Correlation function is shown for a pair of HiCAMP images in which cloud parallax is present. Because the clouds are closer to the sensor, their displacement (after 1 second) differs slightly from that of the ground and a second peak appears. Cross-correlation always merges such peaks into a broad blur.

Fig. 3 — Two BMAP images (a) and (b) of an aircraft moving across a thin cloud bank. (c) The cross-correlation function gives only a single broad peak at the origin, indicating a stationary background. (d) the Phase Correlation function distinguishes between the background (peak and the origin) and the target (peak displaced by amount of aircraft motion).

Furthermore, two useful conserved quantities can be derived: the total signal and the total signal energy. Equation (6) and the fact that x_n, y_n are real imply that

$$\sum_{n=0}^{N-1} p_n = \sum_{n=0}^{N-1} p_n^2 = 1. \tag{8}$$

It follows that the mean and variance, μ_p and σ_p^2, are independent of the image content:

$$\mu_p = \frac{1}{N}, \quad \sigma_p^2 = \frac{N-1}{N^2}. \tag{9}$$

Equations (8) and (9) hold also in two dimensions, with N the total number of samples.

When the maximum value of p is near unity, Eq. (8) constrains the rest of the function to small values. The height of the peak thus serves as an indicator of the reliability of PC. (This is not true

of the cross-correlation function.) If p_s, the value at the true shift, is considered to be the "signal," and the rest of the function is modeled as random noise, then a signal-to-noise ratio (SNR),

$$\text{SNR} = \left(N \frac{p_s^2}{1 - p_s^2} \right)^{1/2}, \tag{10}$$

can be calculated, valid for large N. It measures the likelihood of locating the correct peak in p_n, i.e., the reliability of the method for estimating displacement. Figure 4 shows a typical distribution of p_n, with $N = 5760$ and $p_{\max} = p_s = 0.18$. Although the peak is far from its theoretical maximum value of one, even with such a modest frame size the SNR is 13.6.

Notice that in Fig. 4 the amplitudes of the residual samples are approximately Gaussian-distributed. If they were also independent, then we could associate with a given value of $p_{\max}$ the probability that it represented the true displacement. This probability depends on the marginal distribution of the Gaussian. However, in practice this method usually fails.

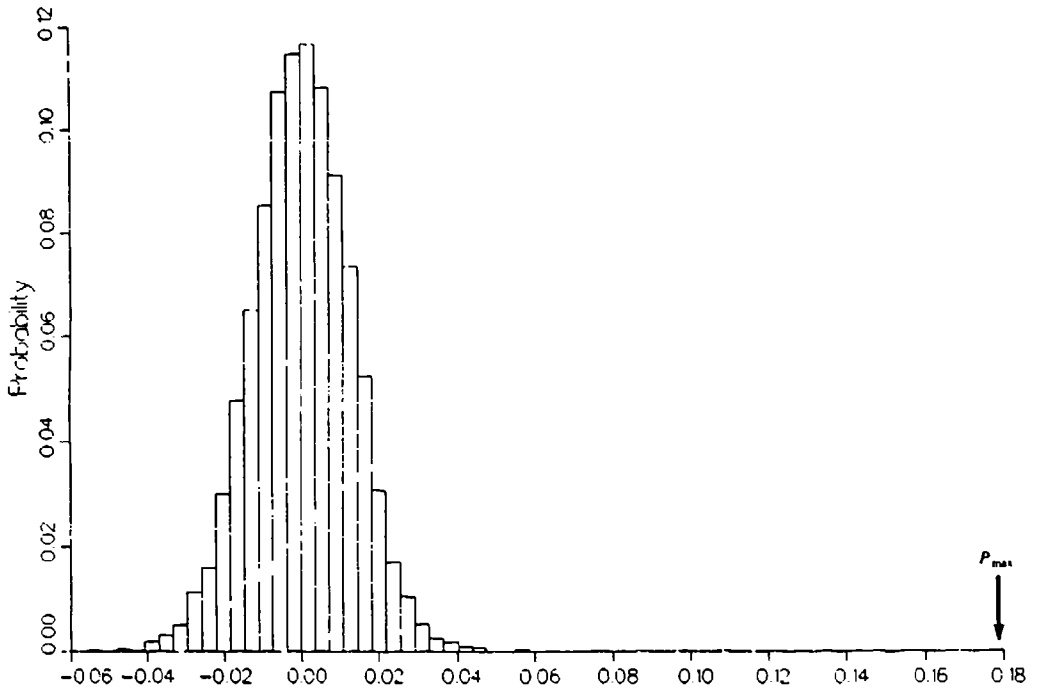


Fig. 4 — Histogram of the Phase Correlation function p_n from a pair of BMAP images. Ideally the peak value p_{max} is 1, and the remainder of the function is zero. But sensor noise, nonperiodicity, subpixel shifts, distortions, etc., all broaden the histogram and decrease p_{max}, subject to the constraints of Eq. (8). In this example, p_{max} is only 0.18, but the signal-to-noise ratio is still greater than 13.

Nevertheless, for any pair of natural background images, there is a threshold for p_{max}, above which value PC is reliable. This performance metric, the magnitude of p_{max}, also lets one quantify the degradation of PC as a result of noise, image distortions, nonoverlap of images, etc. Figure 5 describes an example, showing the effects of pure rotation on PC's accuracy. The value of p_{max} is plotted for a given image size. For small rotations, this maximum value occurs at (0,0), indicating the correct translation of 0 pixels (for HiCAMP, a pixel is identical to a sample). At a large enough rotation, PC fails (shown by the dots on the curves), because p_{max} merges with the "noise." The maximum of p no longer occurs at the point (0,0). In practice, a threshold can be set at several times the value of p_{max} at which PC fails.

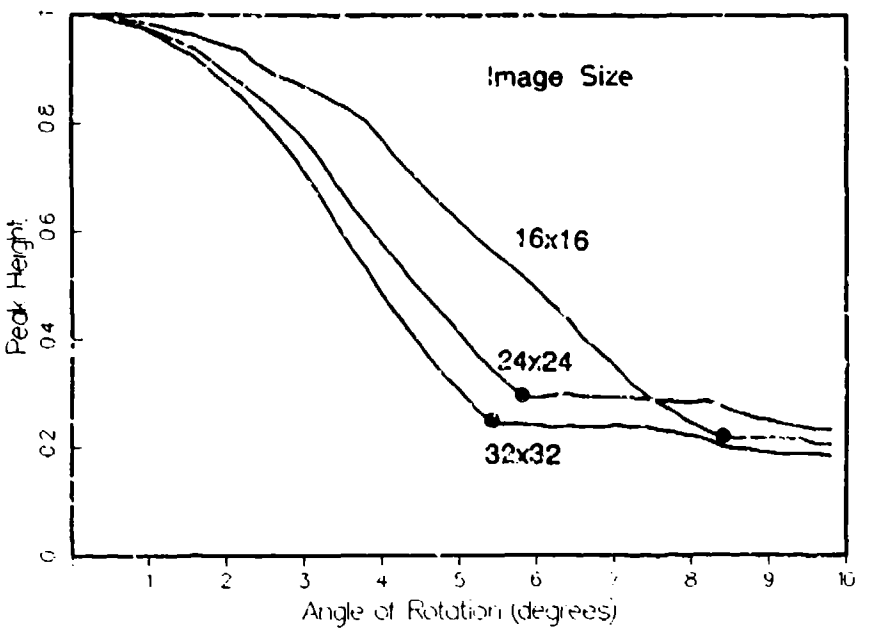


Fig. 5 — The value of the peak of the Phase Correlation function p_{max} as a function of rotation angle for several image sizes. Rotations of HiCAMP imagery were approximated by using Cubic Convolution (see Section 6). For pure, small rotations, the peak occurs at the origin of a correlation plot (not shown), indicating the correct translation: 0 pixels. As the rotation is increased, p_{max} decreases until it merges with the "noise." The dots on the curves indicate those angles where noise overcomes the desired peak at the origin.

Enhancements and Technical Modifications of PC

For ideal pairs of images (periodic scenes and integral shift) the Phase Correlation function p_n reduces to the Kronecker delta function. But real imagery has noise, nonrigid distortions, noninteger sample displacements, and finite extent. Our modifications to the basic PC method enhance its performance in the presence of these degrading factors.

The first modification reduces edge effects. The representation of a finite image by the discrete Fourier transform implicitly imparts to it a periodicity equal to the image size, which often creates a strong virtual discontinuity across the edges. When the high frequencies associated with these discontinuities are combined in p_n, they usually wash out and result in noise and a reduced peak value.

However, whenever the discontinuities are coincident, phases in the two images are partially correlated, and the values of p_n can rise. Therefore, local peaks can occur at all points corresponding to even a partial overlapping of the edges of the two shifted images. To suppress these false peaks, which take the shape of a cross through the origin of a PC plot, we taper the images at the edges to their means by using, for example, a von Hann window, i.e., multiplication by a sinusoid. (The results are insensitive to the type of tapering.)

Another problem caused by the implicit imposition of periodicity occurs in the interpretation of a peak in $\boldsymbol{p}$. As are images, $\boldsymbol{p}$ is also made periodic by the use of DFTs so that, for example, a peak at position s is always accompanied by a peak at position $s-N$. This ambiguity can be eliminated by the following procedure: Choose one shift hypothesis and apply PC again, but this time to cropped versions of the two images, corresponding to those subimages that would be in common were that hypothesis correct. The peak in the new $\boldsymbol{p}$ is sharper than in the old if the hypothesis is correct; otherwise, the peak disappears because the cropped images have no common points. In two dimensions, the ambiguity allows all combinations of s and $N \pm s$ in both directions to be confused, but the same cropping principle can be used. In practice, the displacements are usually known to an accuracy better than one-half frame ($N/2$ samples), so that no iterations are required. Cropping, however, is still beneficial.

Another modification of PC concerns shifts by a nonintegral number of samples. Because the finite DFT must be used to produce results in finite time, all transformed signals, including p_n, are effectively bandlimited. In the limit for large N, p_n therefore approaches a sinc function:

$$p_n \simeq \frac{\sin \pi(n-s)}{\pi(n-s)}, \tag{11}$$

i.e., a bandlimited δ-function. Note that when the shift s is an integer, Eq. (11) reduces to the Kronecker delta function.

Therefore, after locating the peak value $p_{\max}$, several points in its neighborhood can be fit to Eq. (11) to produce a subsample estimate for s. In practice, using only $p_{\max}$ and its highest neighbor gives excellent results. This method is used in the comparisons with IDEA performance in Section 4.

In many imaging systems, distortions such as zoom, rotation, and skew are introduced by sensor motion. These distortions all degrade the performance of PC. However, their effects can be minimized by partitioning the image into small, overlapping subframes and applying PC to each of these separately. (Alternatively, one may apply PC to only a select number of subframes—for example the corners and the center—and deduce all local motions by interpolation.) The advantage of using smaller subframes is reduced computation; the penalty is the decreased SNR implied by Eq. (10). The size of the penalty can depend on many other factors besides size of the subframe: pixel resolution, sampling rate, frame time, platform altitude, and the statistics of the backgrounds of interest relative to sensor noise.

Finally, we report on an improvement to PC based on the original idea of a linear phase shift (see Eq. (2)). The location of the maximum of p_n gives an excellent estimate of the displacement. This estimate can be used as a guideline for phase-unwrapping. The straight line in Fig. 6 has a slope corresponding to the shift (0.4 samples) predicted by PC for a BMAP image pair. The data points are the phases of the DFT unwrapped around the straight line. Because the data are highly

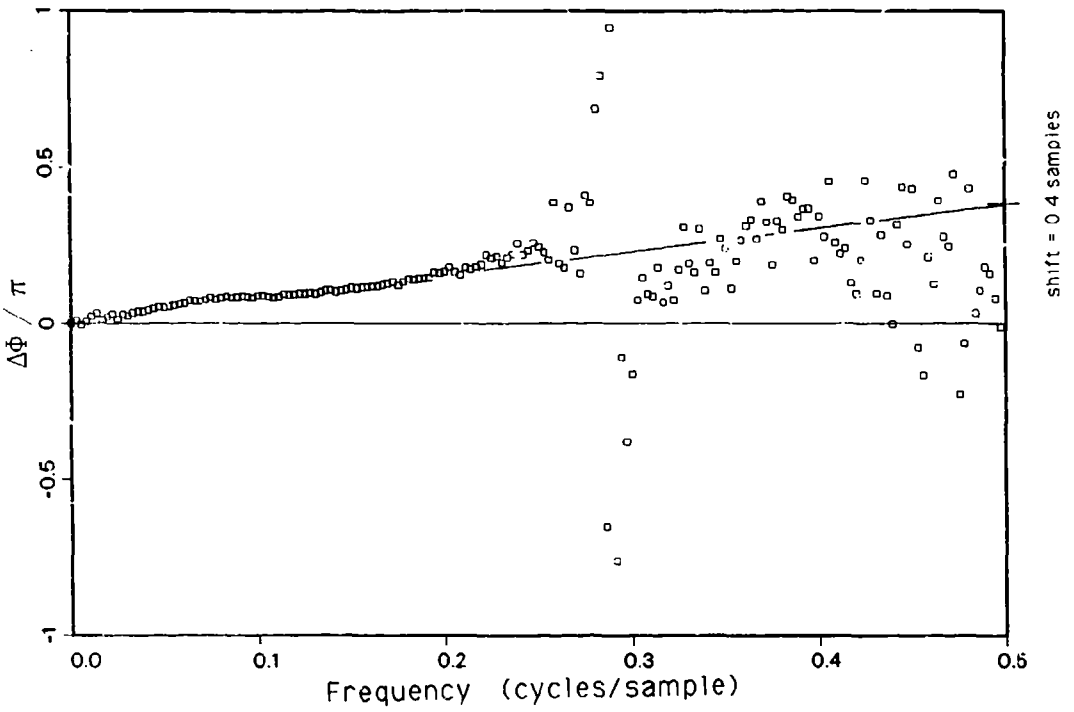


Fig. 6 — The Fourier phase differences between two BMAP images. The straight line ($\phi = 2\pi ks/N$) indicates a shift s of 0.4 samples, which is known from the Phase Correlation peak location. Oversampling (3.5 samples per dwell) causes high-frequency noise that can be removed with low-pass filtering. A subsequent lms Phase Fit then can produce a more accurate shift estimate without the need to interpolate the PC function.

oversampled, in the Nyquist sense, the points at higher frequencies correspond to temporal noise. They can be discarded, and a linear fit can be made to the low-frequency points only, to yield a more accurate estimate of the slope and hence of the true image displacement. Without the original line as a guide, the determination of slope is as ambiguous as the phase, and this Phase Fitting method cannot be realized. Estimating the slope by this method not only reduces an obvious source of error, temporal noise — it also obviates the need for interpolation of the PC function. Furthermore, if the data are undersampled, aliasing causes a negative bias in PC. If the undersampling is not too severe, all aliasing occurs at high frequencies, so that bandpass filtering is again appropriate.

4. GRADIENT-BASED DISPLACEMENT ESTIMATION

Gradient-based displacement estimation is based on a first-order Taylor Series expansion of image intensity. This section first describes one such older method, IDEA, and applies it to calibrated imagery. For comparison, PC is applied to the same data. The theoretical underpinnings of IDEA's effectiveness are analyzed, and the insights gained are used to generate a family of gradient-based estimates. Their experimental performance is then compared to theoretical predictions by use of a model autocorrelation function.

IDEA [1] was originally developed by the Optical Sciences Company (tOSC) as a heuristic method. It was improved through a suggestion of E. Rauch of Lockheed Corporation, who recognized the original version as an approximation to a least-mean-squares (lms) solution to the displacement estimation problem. He then extended this concept to a larger class of image distortions, calling

the resulting version the Gradient Estimated Motion Suppression (GEMS) algorithm [9]. The motivation in these earlier studies, as here, was minimizing residual clutter in differenced frames.

Instead of repeating tOSC's intuitive rationale for the form of IDEA, which is a particular gradient-based method, we describe the general principles of gradient-based displacement estimation. (The gradient-based approach was introduced originally by Limb and Murphy [10].) Then IDEA, as well as a new family of related algorithms, becomes understandable in a larger context. For simplicity, the analysis is restricted to one dimension. Appendix B discusses the extension to two dimensions.

As before, x_n and y_n are defined to be corresponding pixel intensities from an image pair shifted by s samples. These are taken to be samples of an underlying continuous image intensity, r, which is the convolution of scene radiance with a sensor point-response. Because the interest here is in gradient-based estimation, it may be assumed without loss of generality that the sequence $\{r_n\}$ has had its mean removed.

The position labels are chosen so that

$$x_n = r_n,$$

and then

$$y_n = r_{n+s}. \tag{12}$$

For small shifts s, the gradient approximation is

$$y_n \simeq x_n + s\, g_n, \tag{13}$$

with $g_n \equiv dr/dn$. If approximate values of g_n are available, an estimate $\hat{s}$ of the value of s may be found by minimizing the squared differences,

$$\frac{1}{N} \sum_{n=0}^{N-1} \left(y_n - x_n - \hat{s}\, g_n \right)^2, \tag{14}$$

between measured and estimated values of y. The answer to the minimization problem is

$$\hat{s} = \frac{\sum (y_n - x_n) g_n}{\sum (g_n)^2}. \tag{15}$$

Unfortunately, this estimate requires the values of the slope g_n, information that is not generally available. Some approximation for g_n is necessary, and its form can introduce bias.

The IDEA formulation originated without the benefit of Eq. (15) as a guide, but is related to it. IDEA is given by:

$$\hat{s}_{IDEA} = \frac{1}{4} \frac{\sum (y_{n+1} - x_{n+1} + y_n - x_n)(y_{n+1} - y_n + x_{n+1} - x_n)}{\sum (x_{n+1} - x_n)^2}. \tag{16}$$

This estimate differs from the form in Eq. (15) in that: (1) the difference in the numerator's first term has been replaced by a difference of two averages, and (2) different estimates for the derivatives are used in the numerator and denominator. Each of these differences causes a bias* in the expected value of $\hat{s}_{IDEA}$, which is illustrated in Fig. 7. The histograms, which result from applying IDEA to 203 BMAP images, show an expected dispersion about a mean estimate, but the mean is consistently less than the true shift by an amount that is comparable to the standard deviation.

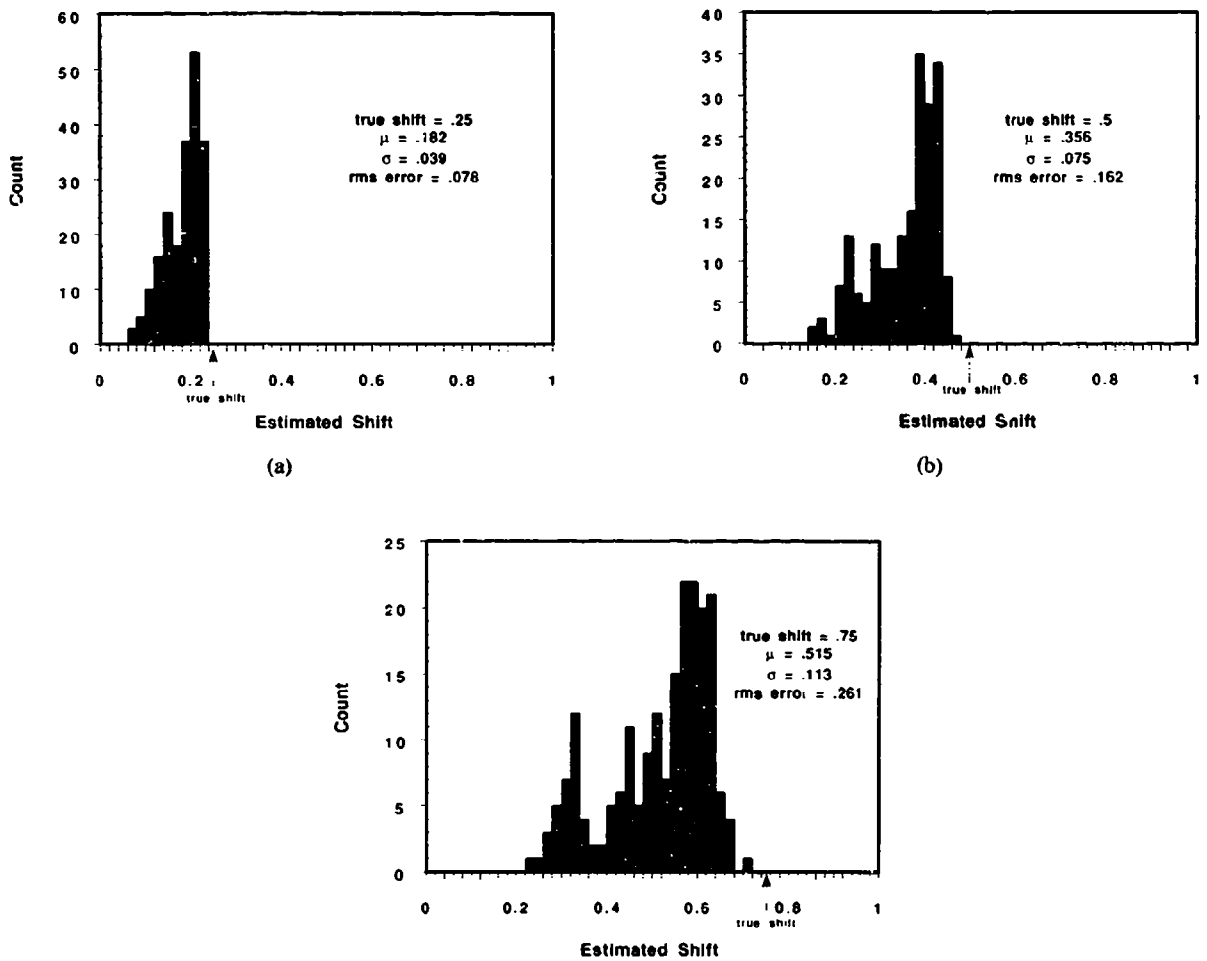


Fig. 7 — Performance of IDEA applied to 203 BMAP image pairs with known shifts of (a) 1/4, (b) 1/2, and (c) 3/4 pixel. The images were extracted with the technique illustrated in Fig. 14 and described in Section 5. The shift estimate is negatively biased in all cases, confirming analytical predictions that use the model autocorrelation function.

*Bias was apparently anticipated by D. Hench in Ref. 11, p. 18, Eq. (11).

Compare these results with Fig. 8, which shows the results of estimating the shifts using Phase Correlation. Being a gradient-based method, IDEA is expected to be an excellent estimate for small shifts, but it is clearly lacking, because of bias.

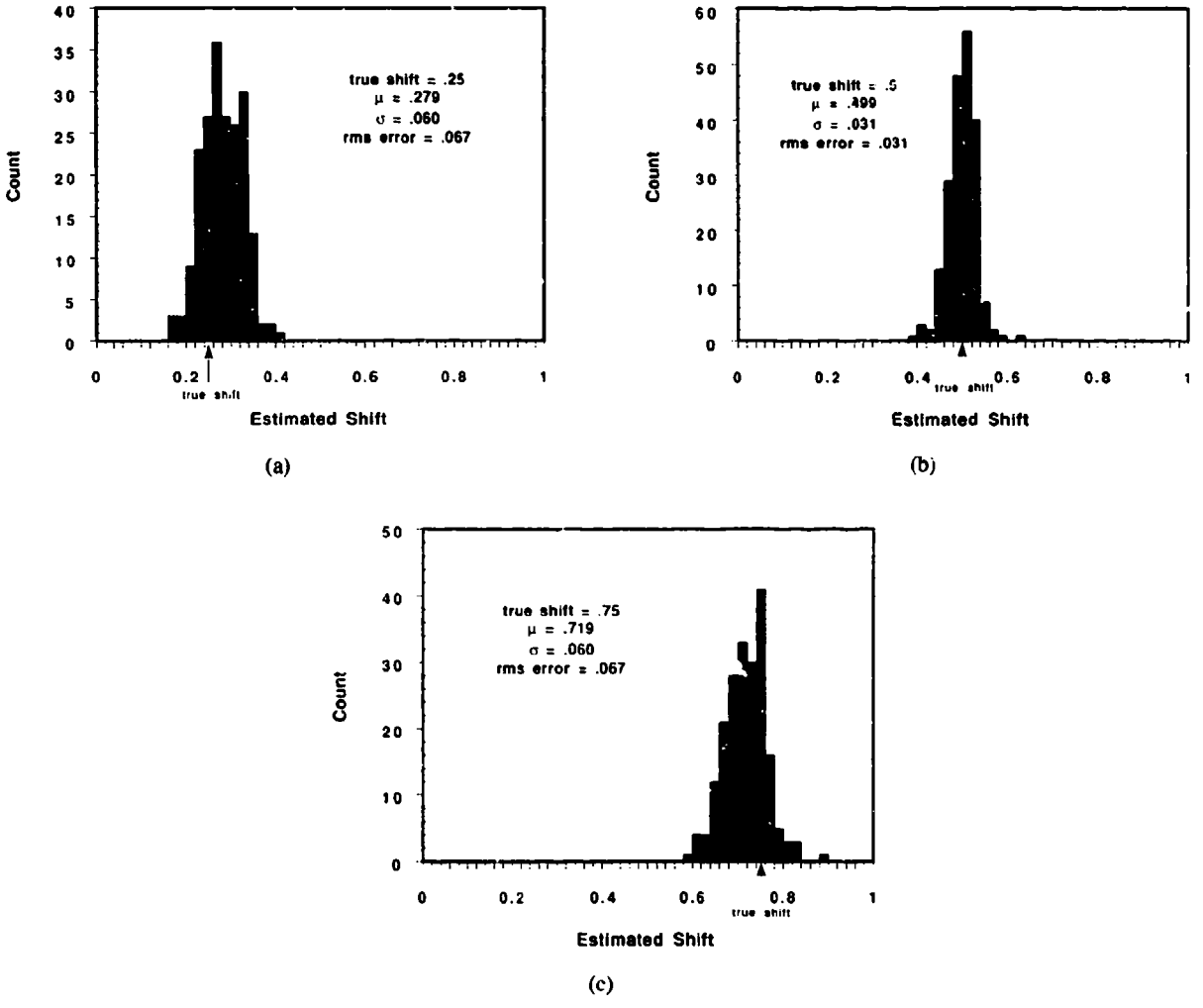


Fig. 8 — Subpixel performance of Phase Correlation. Applied to the same database of imagery as in Fig. 7, Phase Correlation produces smaller rms errors than IDEA, primarily because of a smaller bias. Because it is based on a gradient approximation, IDEA is expected to perform better, at least for small displacements.

The statistics of $\hat{s}_{\text{IDEA}}$ can be analyzed by using the following expansion of the expected value of the ratio of random variables A and B:

$$\langle \frac{A}{B} \rangle = \langle \frac{<A> + \Delta A}{<B> + \Delta B} \rangle = \frac{<A>}{<B>} - \frac{<(\Delta A)(\Delta B)>}{<B>^2} + \ldots \tag{17}$$

where $\Delta X \equiv X - <X>$, $X = A,B$. Using the leading terms in Eq. (17) is reasonable whenever using the Taylor Series expansion of 1/B is reasonable — that is, as long as B never gets too small and doesn't fluctuate much. The denominator in Eq. (16), being a sum of many squares, usually conforms to these constraints. The higher-order terms in Eq. (17) are small relative to the first terms in our application. The ultimate justification for this claim lies in the agreement with empirical results, which will be demonstrated. Generally, the higher-order terms are of order (at least) $\sigma_x / <X>$, relative to the first, so Eq. (17) is usually a good approximation when the coefficient of variation of X is small.

The first-order bias of $\hat{s}_{\text{IDEA}}$ can be expressed in terms of the normalized autocorrelation function of the underlying noiseless image:

$$\rho_s \equiv <r_n r_{n+s}> / \sigma_r^2 . \tag{18}$$

The variable r is assumed to be stationary (in the stochastic sense), with mean zero and variance σ_r^2.

Along with the measured values x_n and y_n, additive (sensor) noise terms are included and are assumed to be independent of the intrinsic scene statistics. Equation (12) becomes

$$x_n = r_n + \xi_n$$

and

$$y_n = r_{n+s} + \nu_n , \tag{19}$$

with ξ_n and ν_n zero-mean, uncorrelated white noise processes, with common variance:

$$\sigma_N^2 \equiv <\xi_n^2> = <\nu_n^2> .$$

By substituting Eq. (19) into Eq. (16) and using Eq. (17), we can write the leading term for $<\hat{s}_{\text{IDEA}}>$ as

$$<\hat{s}_{IDEA}> = \frac{1}{4} \frac{\rho_{1-s} - \rho_{1+s}}{1 - \rho_1 + \sigma_N^2/\sigma_r^2} . \tag{20}$$

The bias associated with this estimate is just the expected estimated shift minus the true shift, that is : $<\hat{s}_{IDEA}> - s$.

According to Eq. (20), further characterization of $<\hat{s}_{\text{IDEA}}>$ depends on the form of the image autocorrelation function.

Model Autocorrelation Function

The autocorrelation function has been normalized to the value 1 at $s = 0$, where it assumes its theoretically maximum value. It can be shown that if $<(g_n)^2>$ $(= <(r'_n)^2>)$ is finite, which can be assumed for imagery blurred by optics, then ρ_s must be flat at $s = 0$, i.e., $\rho_0' = 0$. Therefore, for small s one expects:

$$\rho_s \simeq 1 - \alpha s^2 \ (\alpha \geq 0). \tag{21}$$

Substituting Eq. (21) into Eq. (20) yields, in the limit of zero noise,

$$\hat{s}_{\text{IDEA}} \simeq s. \tag{22}$$

That is, for an autocorrelation function of the form in Eq. (21), $\hat{s}_{\text{IDEA}}$ is unbiased. All the other gradient-based estimates for s that we analyze share this property. Because $\hat{s}_{\text{IDEA}}$ is biased empirically (Fig. 7), the extent to which our other estimates are less biased than $\hat{s}_{\text{IDEA}}$ is determined in part by the extent to which Eq. (21) fails to model the real autocorrelation function.

For digital imagery, often the only directly measurable values of ρ_s are for s integral, and for our purposes, $s = 1, 2$ are the most important data points (along with the image variance σ_r^2, which has been used to scale ρ_s). Because, in addition, ρ_s should be an even function of s, here we assume

$$\rho_s \simeq 1 - \alpha s^2 + \beta s^4, \tag{23}$$

a quartic autocorrelation model. Then empirical values of σ_r^2, ρ_1, and ρ_2 can be used to define α and β. In all cases examined here, both α and β are positive.

Reducing Bias in $\hat{s}$

According to the discussion following Eq. (15), the art of generating a good gradient-based estimate $\hat{s}$ reduces to finding a good approximation for g_n. The simplest choice is $g_n \simeq x_{n+1} - x_n$. We expect this to be more accurate for positive shifts than for negative; the choice $x_n - x_{n-1}$ should be better for negative shifts. The two can be combined to form a mean difference, $1/2\ (x_{n+1} - x_{n-1})$. Another possibility is to average estimates from the two images; for example, $g_n \simeq 1/4\ (x_{n+1} - x_{n-1} + y_{n+1} - y_{n-1})$.

Four approximations for g_n are considered:

(a) $g_n \simeq x_{n+1} - x_n$ **(CAGRE)**

(b) $g_n \simeq 1/2\ (x_{n+1} - x_{n-1})$ **(GEMS)**

(c) $g_n \simeq 1/2\ (x_{n+1} - x_n + y_{n+1} - y_n)$

(d) $g_n \simeq 1/4\ (x_{n+1} - x_{n-1} + y_{n+1} - y_{n-1})$. (24)

The acronym **CAGRE** (**CA**nonical **GR**adient **E**stimate) is used to describe the estimate of Eq. (15) when the natural substitution of Eq. (24a) is used. The **GEMS** [9] formulation apparently used Eq. (24b).

The value of $<\hat{s}>$ (and, hence, the bias) can be estimated for each estimate generated by approximations (a) through (d) in Eq. (24) by using the same method used to evaluate $\hat{s}_{IDEA}$. The results are:

$$(a)\ <\hat{s}_a> = \frac{(1 - \rho_1) - (\rho_s - \rho_{1-s}) + \sigma_N^2/\sigma_r^2}{2(1 - \rho_1 + \sigma_N^2/\sigma_r^2)}$$

$$(b)\ <\hat{s}_b> = \frac{\rho_{1-s} - \rho_{s+1}}{1 - \rho_2 + \sigma_N^2/\sigma_r^2}$$

$$(c)\ <\hat{s}_c> = \frac{\rho_{1-s} - \rho_{s+1}}{2\,(1 - \rho_1) - (\rho_{1-s} - 2\rho_s + \rho_{s+1}) + 2\sigma_N^2/\sigma_r^2}$$

$$(d)\ <\hat{s}_d> = \frac{2\,(\rho_{1-s} - \rho_{s+1})}{(1 - \rho_2) - 1/2\,(\rho_{2-s} - 2\rho_s + \rho_{s+2}) + \sigma_N^2/\sigma_r^2}. \qquad (25)$$

Figure 9 graphs $<\hat{s}>$ for each approximation by using the quartic polynomial model, Eq. (23), which is fit to measured values of ρ_1 and ρ_2 from HiCAMP imagery selected to emphasize the differences in performance of the estimates. Noise-to-clutter ratios (σ_N/σ_r) of 0.00, 0.05, and 0.10 are illustrated. The results for IDEA are also shown.

Note that for an IR scanning sensor with state-of-the-art sensitivity ($\simeq 0.2\ \mu W/(cm^2\text{-ster-}\mu m)$ LW), (σ_N/σ_r) is often less than 0.01 for land or cloud backgrounds. The corresponding values of $<\hat{s}>$ are indistinguishable from those in Fig. 9 for a noise-to-clutter ratio of zero. The most benign Earth backgrounds tend to be ocean, for which the corresponding value of (σ_N/σ_r) is approximately 0.10. However, frame-to-frame subtraction, whose primary purpose is clutter removal, is an unlikely method of detecting moving targets in benign backgrounds. In fact, it *increases* noise by a factor of $\sqrt{2}$.) Consequently, for values of (σ_N/σ_r) much greater than 0.10 (i.e., uncluttered scenes), displacement estimation is probably unnecessary.

Figure 10 plots sample means (based on 10 or more trials) of $\hat{s}$; these correspond to the theoretical curves of Fig. 9. A single image is used, with different trials corresponding to independent simulations of additive noise. Nonintegral sample values are generated for the HiCAMP imagery by using Cubic Convolution (CC) (see Section 6). The CC method interpolates between measured values with a cubic polynomial, making the polynomial model (Eq. (23)) for ρ_s particularly tenable.

The curves of Figs. 9 and 10 are in qualitative agreement. The greatest divergence, which occurs for estimate (d), can be traced to the dependence of $<\hat{s}_d>$ on $\rho_{s\pm2}$ (Eq. (25)). Among all the expected values of estimates considered, only $<\hat{s}_d>$ depends on ρ at such a large value of its argument. Recall that the model Eq. (23) is fit only to smaller values of s, viz. $s = 1,2$.

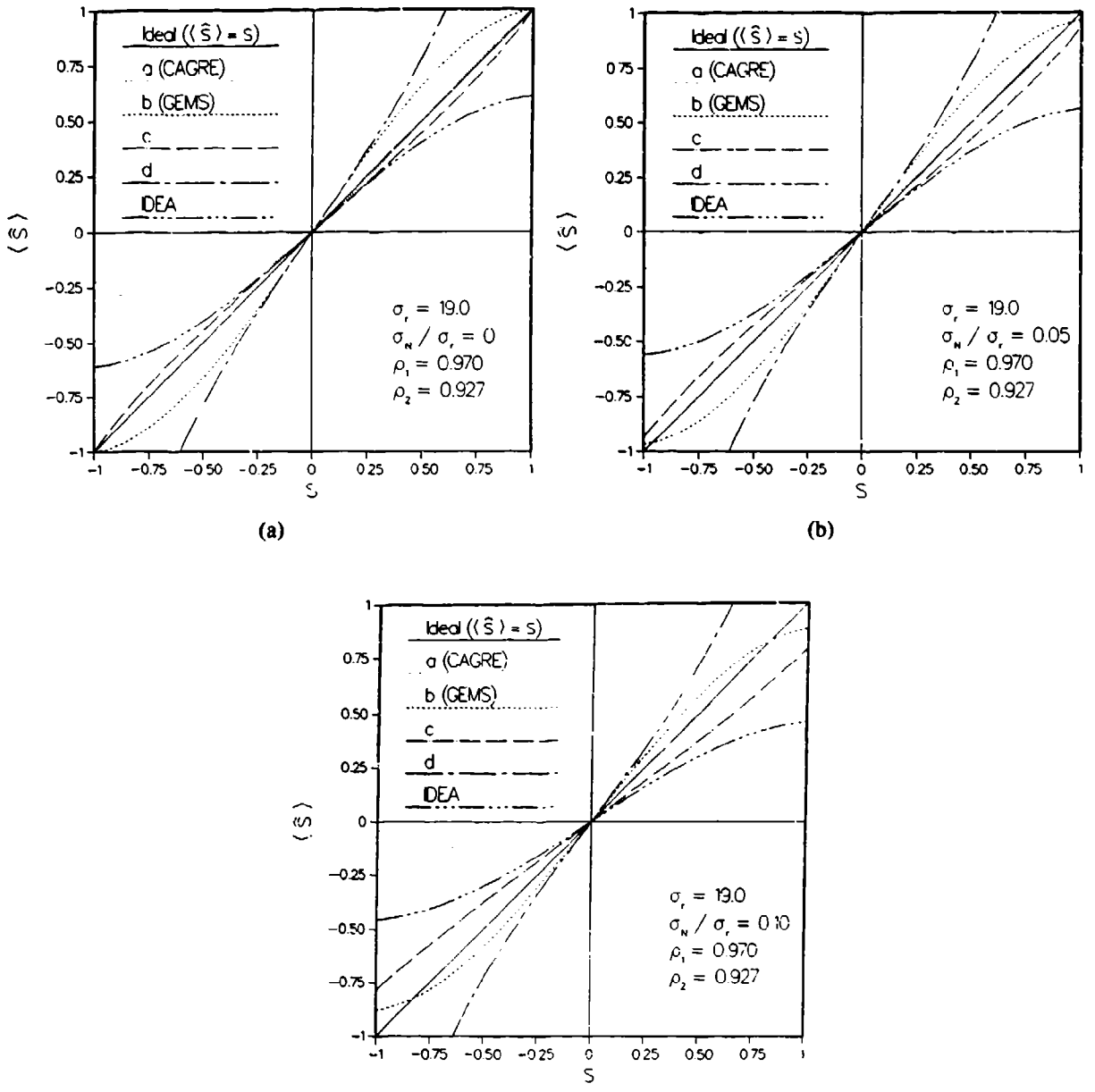


Fig. 9 -- Theoretical performance of IDEA and a family of gradient-based estimates. Constants of the model autocorrelation function are fit to measured values of ρ_1 and ρ_2 (= .970 and = .927) for a single HiCAMP image (of broken clouds and terrain) treated as noiseless, with $\sigma_r = 19\ \mu W/(cm^2\text{-ster-}\mu m)$, while actual sensor noise was $\simeq 1\ \mu W/(cm^2\text{-ster-}\mu m)$. (a) For zero added noise, method (a) is superior, as long as s is nonnegative (it can always be so chosen, by proper labelling convention). (b) Noise-to-clutter ratio of .05 corresponds to a cluttered background with HiCAMP noise. Method (a) is still superior for $s > .05$, but for $|s| < .05$, (c) is the best. (c) Noisy, cluttered imagery (noise-to-clutter = .10). The choice of best estimator depends strongly on the range of expected values of the true displacement.

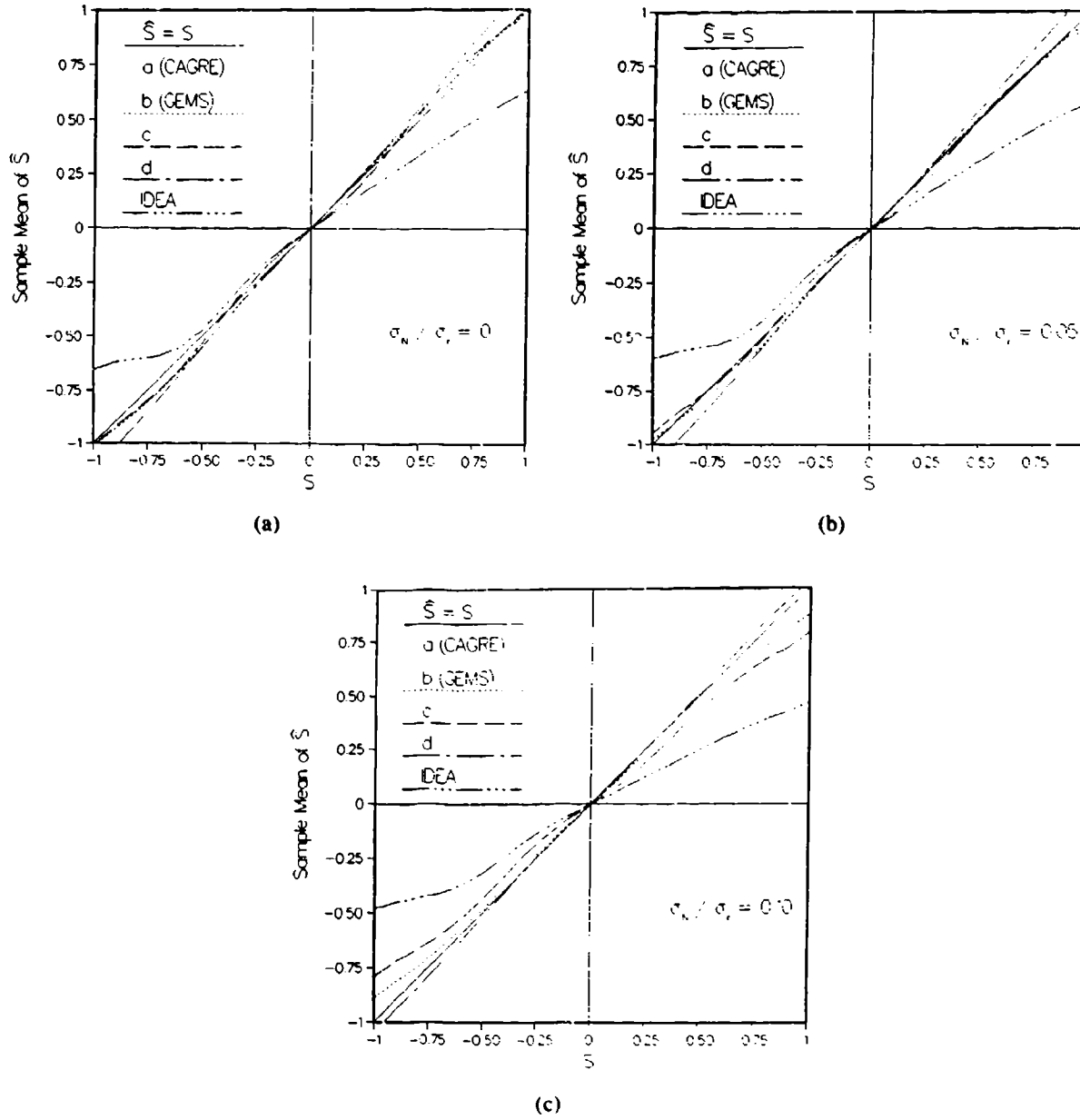


Fig. 10 — Experimental sample means corresponding to the predictions of Fig. 9. In (b) and (c), the estimators were applied after random noise was added to the HiCAMP image. Agreement with theory is generally good, except for method (d), whose mean performance depends more heavily on large values of the shift s through the model autocorrelation function (viz., ρ_{s+2}) than either (a), (b), (c), or IDEA.

Notice in Figs. 9 and 10 that all graphs are antisymmetrical, except for (a) (CAGRE). Also, Figs. 9(a) and 10(a) imply that for positive s, (a) performs best, with (c) the best of the remaining methods. For negative s, (a) can be the worst method, but this defect is inconsequential because s can always be forced to lie between 0 and 1 by a mere relabelling of the coordinates. For the two noise levels illustrated in Figs. 9 and 10, the best estimate depends on the range of typical values of the true shift s. In Fig. 9(b) for $s > .05$, estimate (a) is again generally superior, but for small shifts it becomes the most biased. For $|s| < .05$, estimate (c) is the best, although IDEA is sometimes as good. In Fig. 9(c), the best choice has a more complicated dependence on s. Note that $\hat{s}_{IDEA}$, the estimate that instigated our study of gradient-based methods, is usually one of the more biased choices.

The performance of each estimate depends on the image statistics. Figure 11 shows theoretical results, as in Fig. 9, but for another HiCAMP scene, with ten times the clutter. Except for IDEA, the differences in performance are much less pronounced than for the scene used to generate Fig. 9. For zero noise, again (a) is superior (for positive s), but now (b) and (c) are close competitors. As the noise increases, (d), which uses the most averaging in approximating g_n and hence is expected to be less sensitive to noise, emerges as superior. Note that here a noise-to-clutter ratio of 0.1 corresponds to an artificially high (HiCAMP) sensor noise (i.e., about 20 times its actual value for this scene).

In general, for small s all the estimates except (a) satisfy $<\hat{s}> \rightarrow (1 + \epsilon)s$; if the noise-to-clutter ratio is small enough, then $<\hat{s}_a>$ is also of this form. Also, for zero noise and at $s = 1$, the formulae in Eq. (25a-c) give zero bias, whereas for (d) and IDEA, the amount of bias depends on scene statistics. These $s = 1$ properties depend only on the first-order approximation of Eq. (17), not on the model of the autocorrelation function, Eq. (23).

Note that the expected values from Eqs. (20) and (25) could be used, in principle, to estimate the bias of any given estimate. Then the bias could be subtracted from $\hat{s}$, in the hope of forming an even better estimate. However, this procedure would require independent knowledge, not only of expected noise, but of scene statistics through σ_r and ρ_s.

Finally, these results can be used to interpret some previously published work. In a study [9] of GEMS performance, Rauch and Zele found the rms residual of the difference frame, $D \equiv <(y_n - x_n)^2>^{1/2}$, as a function of the GEMS estimate of the jitter, $|<\hat{s}_{GEMS}>|$, for 200 HiCAMP scenes and a range of jitter from 0.00 to approximately 0.25 pixels. Appendix A explains their results as a consequence of bias in GEMS, of a particular (linear) form that is identical to that predicted for estimate (b) at small values of s. This demonstrates the consistency of our model with independent experimental results.

Section Summary

This analysis began by noting large biases, listed in Fig. 7, in the performance of a gradient-inspired displacement estimate, IDEA, as applied to a database of BMAP imagery for induced shifts of 0.25, 0.5, and 0.75 pixels. For comparison, Fig. 8 showed the improved results of applying Phase Correlation, a more global method, to the BMAP images.

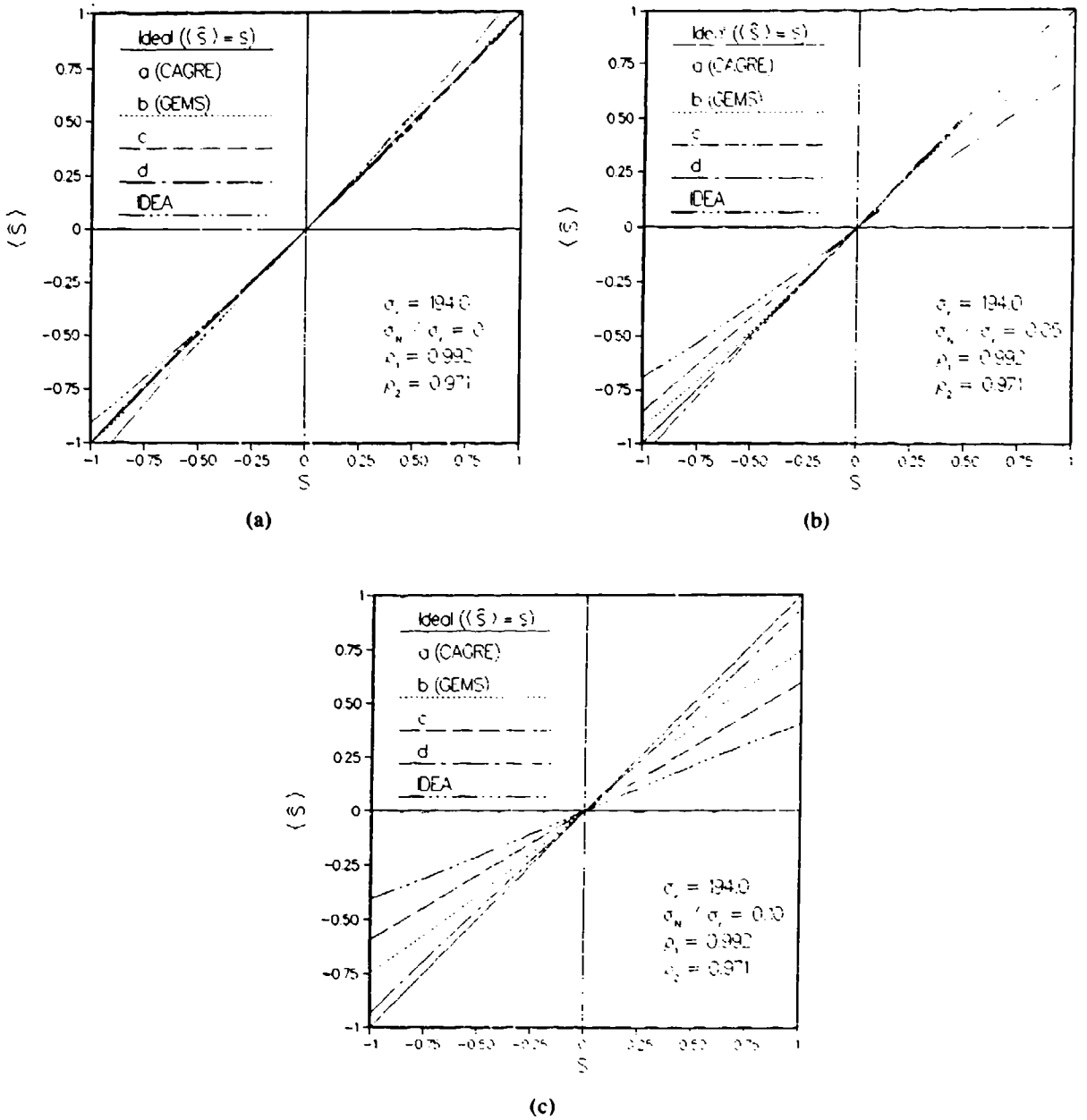


Fig. 11 — HiCAMP results as in Fig. 9, but for an image with 10 times the clutter (σ_c = 194 μW/(cm^2-ster-μm), ρ_1 = .992, ρ_2 = .971). Generally there is less difference in performance than for the case shown in Fig. 9. For zero added noise, (Fig. 11(a)), estimate a) (CAGRE) is again superior. But now a substantial relative noise, (Fig. 11(b)) .05, which corresponds to 10 times HiCAMP sensor noise, is required to make the best choice dependent on s. With twice the noise in Fig. 11(b), Fig. 11(c), estimate d) emerges as the best choice.

Analysis of the theoretical and empirical performance of gradient-based methods led us to a systematic investigation of a family of algorithms that have a generally reduced bias, a contention validated by the results in Table 1 and illustrated in Fig. 12 for the BMAP data. We found that:

- Compared to IDEA, bias and error in the new gradient estimates are reduced, particularly for CAGRE.
- In low noise-to-clutter environments, typical of state-of-the-art IR sensors, the theoretical analysis and HiCAMP validation (Appendix A) imply that the simplest of the gradient estimates, $\hat{s}_a$ (CAGRE), is usually the least biased, if s is simply forced by labelling convention to lie between the values 0 and 1.
- CAGRE can become biased for small displacements if the clutter-to-noise ratio is small, but in such cases displacement estimation is less important to the target-detection problem.

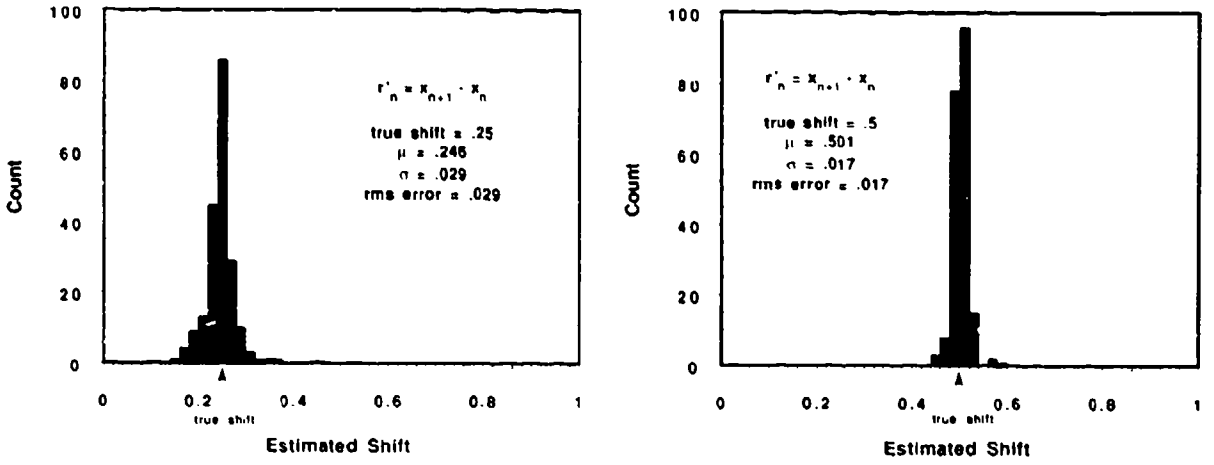
Table 1 — Mean, Variance, and Root-Mean-Square Error (rmse) for Displacement Estimators Applied to the BMAP Database for Three Values of Shift

METHOD	s = 0.25			s = 0.5			s = 0.75		
	μ	σ	rmse	μ	σ	rmse	μ	σ	rmse
(a) GAGRE	0.246	0.029	0.029	0.501	0.017	0.017	0.757	0.030	0.031
(b) GEMS	0.278	0.031	0.042	0.545	0.049	0.067	0.785	0.051	0.061
(c)	0.194	0.037	0.067	0.413	0.057	0.105	0.674	0.054	0.094
(d)	0.287	0.032	0.049	0.590	0.070	0.114	0.908	0.125	0.202
IDEA	0.182	0.039	0.078	0.356	0.075	0.162	0.515	0.113	0.261
PHASE CORRELATION	0.279	0.060	0.067	0.499	0.031	0.031	0.719	0.060	0.067

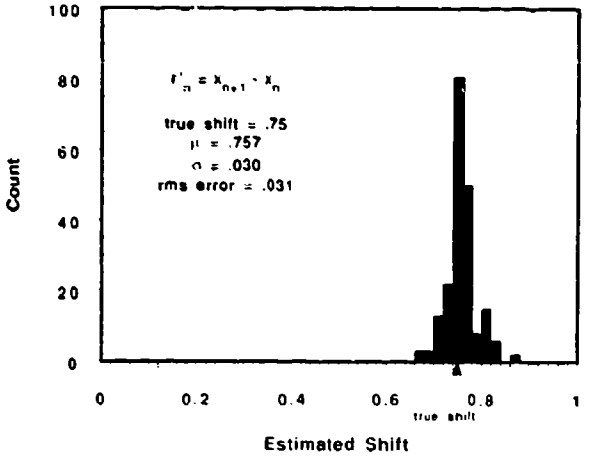
5. METHODS OF PERFORMANCE ASSESSMENT

One way of testing the accuracy of a displacement estimate is to control the relative displacement between images and then compare the estimate to the known value. However, this control was absent at the time of measurement for most appropriate IR data. Nevertheless, often a virtual shift is achieved by creating a shifted copy of an image; for integer displacement, one simply relabels the samples by the desired shift and crops the spurious edges. If desired, additive noise can also be simulated by using a random number generator.

Fractional shifts often require interpolation to produce a second frame, as was done with HiCAMP imagery. However, most registration methods also adopt, at least implicitly, some interpolation assumption (linearity, for gradient-based estimates). Whatever the form, this approximation may be intimately related to the procedure chosen to create the shifted frame—or at least the choice may be strongly biased toward a similar technique. One then faces the prospect of testing registration accuracy on a pair of frames whose relative shift was **simulated** by using a method similar to the one being **tested**. Impressive apparent performance can then result merely because of a kinship between registration technique and testing method.



(a) (b)



(c)

Fig. 12 — Histograms for CAGRE applied to BMAP data, for comparison with IDEA in Fig. 7 and PC in Fig. 8

For example, Fig. 13(a) shows the results of applying PC to the same simulated HiCAMP shifts as in Fig. 9(a). By contrast, in Fig. 13(b) the same simulated displacement was achieved by using a Fourier shift method instead of Cubic Convolution. Because PC is itself a Fourier method, its performance is wonderfully accurate in Fig. 13(b). It would be perfect except for a small error caused by the von Hann window, which is used to taper the edges of the image.

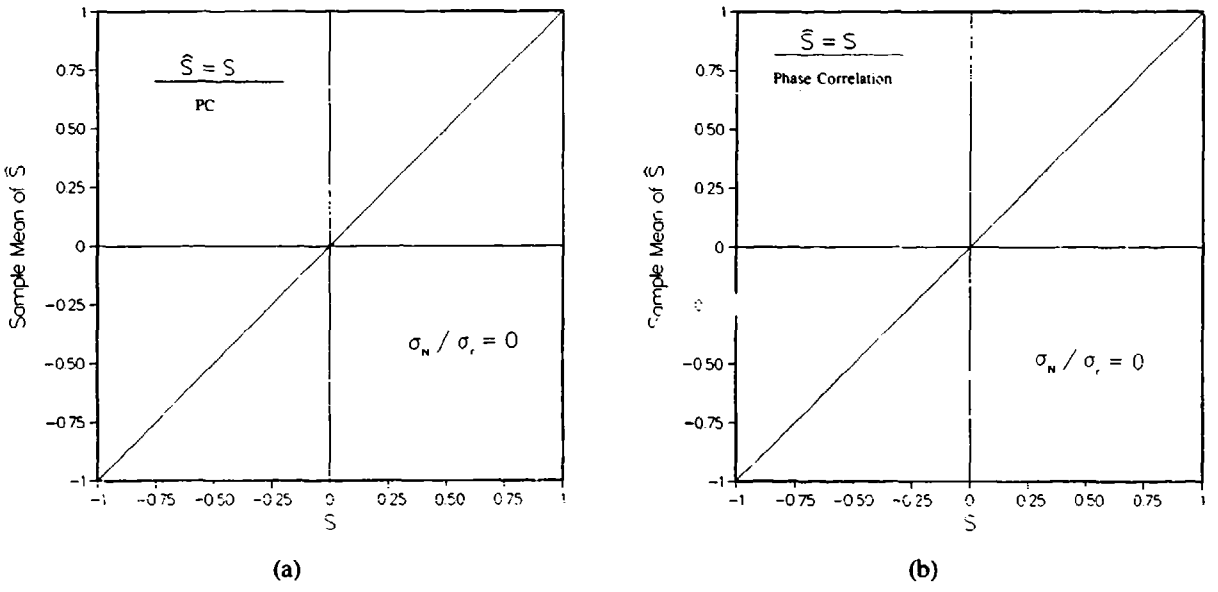


Fig. 13 — Effect of choice of subpixel simulation method on accuracy of displacement estimate. (a) Phase Correlation applied to the same images as in Fig. 9; (b) the displaced images are now produced with Phase Shifting instead of Cubic Convolution. Performance is greatly enhanced because of the similarity between the method of shift simulation and the method of displacement estimation.

One way of ensuring independence of the simulation and registration methods is to use imagery from a scanning sensor that samples more than once per pixel. In this way, analytic interpolation can be avoided altogether. For example, if the sampling rate is four times per dwell time,* then from a single frame of scanning data two new frames can be created. One frame contains every fourth sample (e.g., samples 1, 5, 9,...), and the other frame contains every adjacent fourth sample (e.g., samples 2, 6, 10,...) (Fig. 14). The two images so constructed are effectively misregistered by 1/4 pixel. Shifts of 1/2 and 3/4 may also be created in the same manner. Section 4 used this principle in the comparisons of Phase Correlation and IDEA.

The primary testbed used for comparing the displacement estimates was a database of 203 independent BMAP images of various types of background. The BMAP sensor is a dual-band Midwave Infrared and Longwave Infrared (MWIR/LWIR) scanner that samples an analog signal 3.5 times per dwell time. The backgrounds represented in the database are in either MWIR or LWIR wavebands and include clear sky, various cloud types, terrain, ocean, and urban scenes. The images were culled from a much larger database and represent independent scenes, with no overlap. Because the registration of low-clutter imagery is irrelevant to our problem, a lower limit was imposed on the

*The dwell time is the time the sensor requires to scan the geometrical image of a point object.

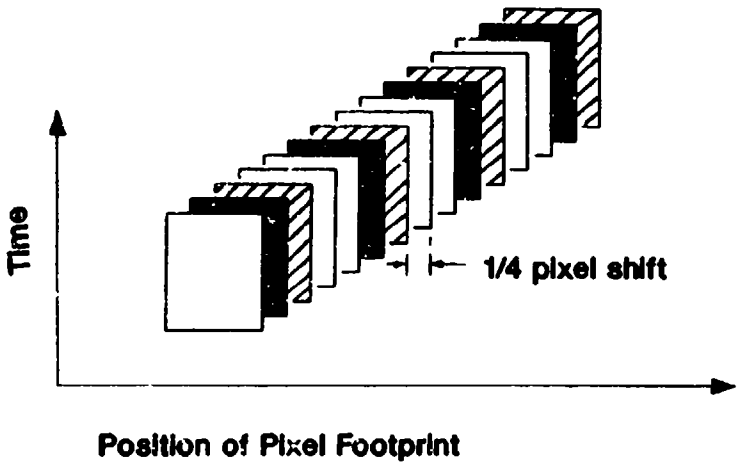


Fig. 14 — Use of oversampled BMAP data to create images with exactly known subpixel shifts. Forming one image from samples 1,5,9,... (shaded pixels) and a second image from samples 2,6,10,... (crosshatched), gives an image pair with a shift of exactly 1/4 pixel with real (sensor-produced) noise. Image pairs with shifts of 1/2, 1/3, 3/4, etc., can be extracted in a similar manner.

clutter content of imagery in the BMAP database ($\sigma_r \geq .05\,\sigma_N^*$). The major effect was to eliminate scenes that are mostly blank with no discernible natural features.

Because this method of simulating misregistration limits the fractional shifts to rather large values,† the more traditional method (described in the text) of interpolation between pixels (or samples) has also been used. In these cases, the imagery came from the HiCAMP database, which was derived from a staring sensor. The effective "sampling rate" is once per dwell time. Results from using this imagery are not as reliable as for BMAP data because they can depend on the particular interpolation procedure (resampler) chosen as described earlier. Subpixel shifts were simulated by using Cubic Convolution, which is described in Section 6.

6. RESAMPLING METHODS

The final product of any registration algorithm is a pair of spatially aligned images. With an estimate of the relative displacement between two images in hand, the final task is to resample one of the images at the estimated shift. The success of the registration process depends not only on the accuracy of the displacement estimate but on the fidelity of the resampling method. This section compares four resampling techniques applied to BMAP IR background data.

Because resampling an image at a displacement of an integer number of samples requires only reindexing the data (along with the usual cropping of nonoverlapping portions), the problem of resampling for an arbitrary displacement is easily reduced to the case of a subsample shift. Two general methods are used for retrieving continuous data from discrete measurements: interpolation and approximation. Interpolation methods reproduce the data exactly at the integer-sample values, while approximation methods, such as an lms fit of some low-order polynomial to neighboring samples,

*For the BMAP sensor, σ_N was approximately 0.20 $\mu W/(cm^2\text{-ster-}\mu m)$.

†As soon as a shift less than 1/3.5 pixels is simulated, coverage gaps between successive pixels develop.

need not. In situations of interest here, the shift can be small, and for such cases the limiting behavior of a resampler should correspond to no resampling. Therefore, only interpolation methods are considered. Also, the discussion is limited to shifts in one direction only—two-dimensional shifts can be effected simply by successive one-dimensional shifts.

Interpolators

Linear Interpolation

This method, which is the simplest, is expected to give reasonable performance and will serve as a baseline for comparison with more elaborate schemes.

Spline Interpolation

Spline interpolation algorithms are perhaps the most commonly used, with an extensive literature devoted to their development. Here **Cubic Spline** interpolation was chosen to represent this class of algorithms [12]. In particular, the CSAKM routine from the International Mathematical and Statistical Libraries, Inc. (IMSL, Inc.) library [13] was selected; this program is based on a method developed by Akima [14]. Optical systems remove high spatial frequencies, so an interpolator that is devoid of large artifactual oscillations is preferable. The Akima interpolator exhibited the least intrasample fluctuations behavior of all the IMSL algorithms.

Phase Shifting

Just as the Fourier Phase can be used to estimate the displacement between a pair of images, alteration of the phase can be used to produce a shifted copy of a single image. This method has been used to simulate sensor jitter [15]. (See also Fig. 13(b).) It is exact for integer-sample shifts and can be adapted to subsample shifts. If X_k is the DFT of the image x_n given by Eq. (3), a shifted image x_{n+s} is defined by its Fourier transform:

$$X_k^s \equiv \begin{cases} X_k\, e^{2\pi iks/N} & , \quad 0 \le k < 1/2(N-1) \\ X_k\, e^{-2\pi iks/N} & , \quad 1/2\,(N-1) \le k < N\,, \end{cases} \tag{26}$$

where N is assumed to be odd. This form imposes conjugate symmetry ($X_k^s = (X_{N-k}^s)^*$) to ensure that the new sequence x_{n+s} remains real. Choosing N to be odd avoids a problem at the Nyquist frequency, $k = 1/2N$ cycles/frame; a sinusoid of this frequency that is in phase with the sampling process cannot be detected. Ignoring this effect can lead to surprisingly poor interpolation, with large oscillations between the data samples.

Cubic Convolution

This technique is particularly well-suited to interpolation of optical imagery [16] and has been used to resample Landsat data. Between any pair of integer sample values, Cubic Convolution constructs a cubic polynomial that is continuous and has a continuous derivative across the integers. After these conditions are satisfied, the one remaining degree of freedom is used to minimize errors

between the interpolated value and a presumed Taylor-series expansion of the underlying continuous image. The procedure is implemented by simple convolution with a four-point kernel:

$$x_{n+s} = \sum_{k=-1}^{2} w_k(s)\, x_{n+k}, \quad (0 \le s \le 1) \tag{27}$$

with

$$w_{-1}(s) = 1/2\,(-s + 2s^2 - s^3)$$

$$w_0(s) = 1/2\,(2 - 5s^2 + 3s^3)$$

$$w_1(s) = 1/2\,(s + 4s^2 - 3s^3)$$

$$w_2(s) = 1/2\,(-s^2 + s^3).$$

An attractive feature of this method is its purely local dependence on the data (four adjacent sample values), as opposed to Cubic Spline and Phase Shifting, which depend on global image values. It is ideal for implementation in a massively parallel computing architecture.

Evaluation and Results

To assess the performance of resampling techniques, oversampled BMAP data are used to create image pairs with known subpixel displacement. Previous studies typically have relied on either synthetic data or on an artificial shift of real data. By using the method illustrated in Fig. 14 for shifts of 1/4, image pairs are constructed with shifts of 1/2, 1/4, and 1/5 pixel. Then the resampling algorithms can be applied to real, shifted imagery; one is not forced to analyze data with unknown jitter, which would require a preliminary displacement estimate contributing extra, unknown error.

Performance of the resampling techniques is measured by the value of the clutter reduction factor Γ defined by

$$\Gamma \equiv \left[\frac{<(y_n - x_n)^2>}{<(\hat{y}_n - x_n)^2>} \right]^{1/2} \tag{28}$$

where $\hat{y}_n$ is the resampled version of y_n, and x_n is the "unshifted" image. Γ is the amount that resampling reduces clutter leakage for a frame differencing signal processor. If perfect registration is achieved and the background varies linearly over the range of the shift, then

$$\Gamma \to \Gamma_0 = \left[\frac{\sigma_g^2}{2\sigma_N^2}\, s^2 + 1 \right]^{1/2} \tag{29}$$

where* $\sigma_g^2 >> <g_n^2>$. For images with high clutter ($\sigma_g >> \sigma_N$) and/or large shifts, clearly the potential for accurate resampling to decrease residual clutter after differencing is greater.

Figure 15 plots the experimental results for Γ vs s. The performance ranking is nearly constant, with Cubic Convolution being marginally superior. The theoretical value Γ_0, given in Eq. (29), is plotted as the dashed line for $\sigma_g/\sigma_N = 6$, which is typical for BMAP data [17]. Linear interpolation, although consistently worst, gives results comparable to the other, more elaborate interpolations. Because all the routines perform similarly, other factors—such as ease of integration with a displacement estimate, or computational speed—may dominate in the choice of resampling technique.

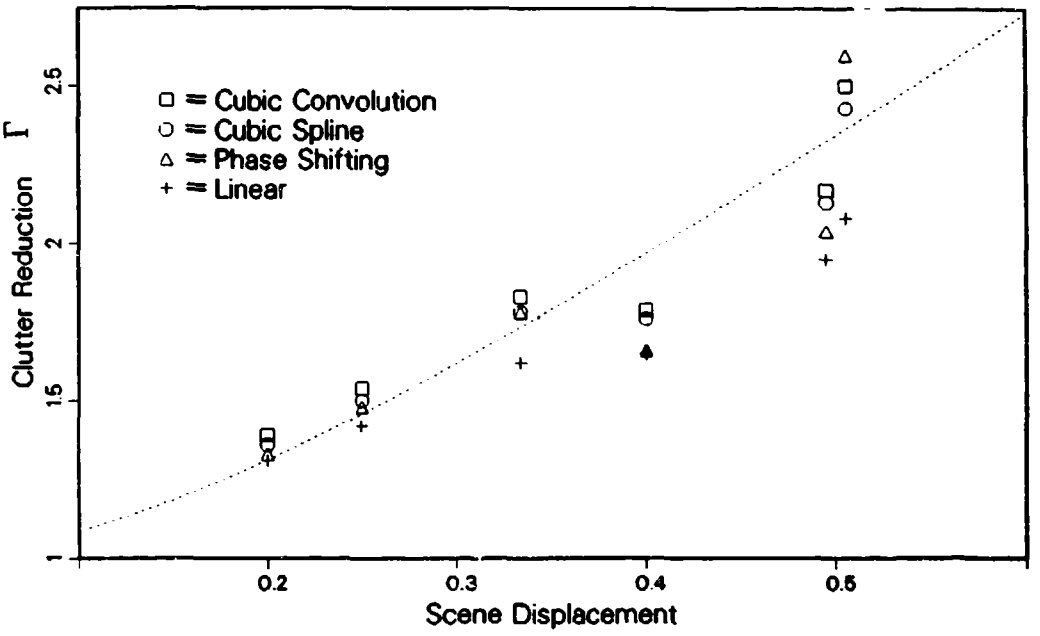


Fig. 15 — Evaluation of four interpolation algorithms. The clutter-reduction factor Γ, averaged over the BMAP database, is plotted as a function of displacement. Cubic Convolution is superior at all displacements except one, but it provides only marginal improvement over simple linear interpolation. The dashed line is the theoretical value Γ_0, which assumes perfect realignment and a σ_g to σ_N ratio of 6, a typical value of BMAP data.

7. SUMMARY AND CONCLUSIONS

We have shown that PC (Phase Correlation) is a robust estimator of image displacement. It is accurate for large as well as small displacements; it is superior to conventional cross-correlation; and it has a natural ability to detect parallax effects. More generally, it can be used to recognize the independent motion of features within an image and so may have important applications directly in the area of autonomous target detection.

Motivated by the early success of IDEA (Image Displacement Estimation Algorithm) to investigate gradient-based methods of shift estimation, we developed a family of such methods that differ only in the choice of approximation for image gradient. In the process, we explained the bias

*The derivation of Eq. (29) follows along the lines of the Section 4 calculations. See especially Eq. (13).

apparent in experimental applications of IDEA and GEMS (Gradient Estimation Method) to HiCAMP and BMAP imagery. The optimal choice among the family depends on the particular image statistics. However, as long as shifts are small and the noise-to-clutter ratio is not unusually large ($>>0.1$), CAGRE is expected to perform best.

Phase Correlation is generally as effective for the HiCAMP and BMAP imagery as most of the gradient-based methods we examined. CAGRE performs better for the BMAP database, but the Phase Fitting method, or a similar modification for removing high-frequency noise or bias, will probably make this phase-based method perform at least as well as any gradient-based method, while maintaining its superior capabilities in the nonidealized circumstances described. For space-based applications with sampling rates of three per dwell time, this method has achieved error levels of a few millipixels with negligible bias. Of the interpolating methods evaluated, Cubic Convolution is the most accurate for almost all shifts studied. It is also easy to implement, requiring only a few "multiplies" and "adds" of local sample values.

One of the methods used here for evaluating registration methods is new. It exploits oversampled data from an IR scanner to avoid synthetic generation of displacements or use of unknown jitter in real imagery. Synthetic methods, however, were part of our HiCAMP analyses.

The improved methods of IR image registration developed here may permit the use of frame differencing as a moving target indicator in more cluttered environments than has previously been considered practical. This possibility motivated the present work, but our results do not depend on wave band and should prove valuable in any application requiring algorithmic image registration.

REFERENCES

1. D. Hench and D. Fried, "Status of Pseudoregistration Development," BC-276, the Optical Sciences Company (tOSC), Placentia, CA, Feb. 1985.

2. D. Hench and D. Fried, "Pseudoregistration — A Clutter Rejection Technique for Staring Sensors," Proc. Eleventh DARPA Strategic Systems Symposium, Vol. 1, Oct. 22-25, 1985, Monterey, CA, pp. 226-246.

3. J. Evans, S. Yool, and D. Williams, "A Comparison of Two Jitter-Noise Reduction Algorithms," *Proc. IRIS* **31**(2), 349-358 (1986).

4. R. Gran, J. Hulsmann, and M. Rossi, "Clutter Rejection Processing of Hi-CAMP II Data Using Sub-Space Projection," Proc. Eleventh DARPA Strategic Systems Symposium, Vol. 1, Oct. 22-25, 1985, Monterey, CA, p. 249.

5. C. Kuglin and D. Hines, "The Phase Correlation Image Alignment Method," IEEE Proc. 1975 Conf. on Cybernetics and Society, Sept. 23-25, 1975, San Francisco, CA, pp. 163-165.

6. R. Srour, R. Lucke, and J. Kershenstein, "The BMAP Tracker Experiment," NRL Memorandum Report 6241, May 1988.

7. J. M. Tribolet, "A New Phase Unwrapping Algorithm," *IEEE Trans. Acoust. Speech Sig. Proc.* **ASSP-25**(2), 170-177 (1981).

8. R. L. Lucke, A. P. Schaum, J. C. Kershenstein, J. V. Michalowicz, B. V. Kessler, A. B. Blumenthal, I. Goldstein, and A. Krutchkoff, "The Navy's Infrared Backgrounds Measurements and Analysis Program," Proc. of the IRIS Target Background Discrimination Conference, Vol. 3, Feb. 13, 1986, Naval Ocean Systems Center, San Diego, CA, pp. 319-333.

9. H. E. Rauch and F. Zele, "HiCAMP Background Suppression in the Presence of Jitter," Proc. Eleventh DARPA Strategic Systems Symposium, Vol. 1, Oct. 22-25, 1985, Monterey, CA, pp. 182-201.

10. J. O. Limb and J. A. Murphy, "Estimating the Velocity of Moving Images in Television Signals," *Comput. Graph. Image Proc.* **4**, 311-325 (1975).

11. D. Hench, "Statistical Analysis of the Image Displacement Estimation Algorithm," TR-729, the Optical Sciences Company (tOSC), Placentia, CA, July 1986.

12. C. deBoor, *A Practical Guide to Splines* (Springer-Verlag, New York, 1978).

13. IMSL Math/Library, FORTRAN Subroutines for Mathematical Applications, Ver. 1. IMSL, Inc., Houston, TX, April 1987.

14. H. Akima, "A New Method of Interpolation and Smooth Curve Fitting Based on Local Procedures," *J. Assoc. Comput. Mach.* **17**, 589-602 (1970).

15. E. Winter, "Teal Ruby Planning and Diagnostic Tool Development Quarterly Report No. 1," Technical Research Associates, Inc., June 1986.

16. R. Keys, "Cubic Convolution Interpolation for Digital Image Processing," *IEEE Trans. Acoust. Speech Sig. Proc.* **ASSP-29(6)**, 1153-1160 (1981).

17. M. McHugh, A. Schaum, and J. Michalowicz, "Reduced Descriptions of Infrared Background by Clutter Content," IRIS Specialty Group on Targets, Backgrounds, and Discrimination, Vol. 1, Feb. 10-12, 1987, Orlando, FL, pp. 335-357.

Appendix A
EXPLANATION OF PRIOR WORK

The analysis of Section 4 can be used to interpret some previously published work. In a study of GEMS performance [A1], the rms residual of the difference frame, $D \equiv <(y_n - x_n)^2>^{1/2}$, was plotted vs the GEMS estimate of the jitter, $|<\hat{s}_{GEMS}>|$, for 200 HiCAMP scenes and a range of jitter from 0.00 to approximately 0.25 pixels. The resulting scatter plot (Fig. 5 of Ref. A1) shows a strong linear trend, especially in the range of 0.10 to 0.25 pixels. The upper curve of Fig. A1 shows a synopsis of the data, in the form of a fit to the model described below.

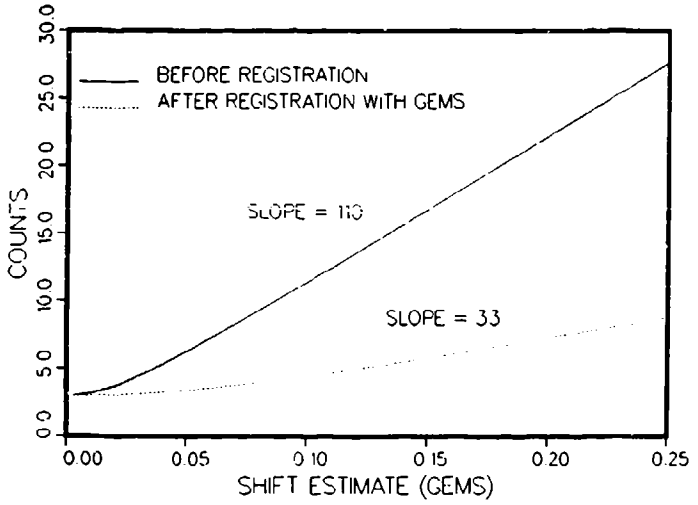


Fig. A1 — Synopsis of Rauch and Zele results: rms values of differenced frames vs GEMS estimate of shifts, before and after registration. The experimentally reported values have been fit to the model of Eq. (29) by matching: (1) the slope in the linear region and (2) the intercept. The value of the (common) intercept gives intrinsic image noise; the slopes of the two curves permit estimation of bias (~30%) in the GEMS algorithm.

The lower curve shows the corresponding fit to the values of D in the difference frame after GEMS registration (with linear interpolation) was applied. The data converge as $s \to 0$ to approximately the same ordinate value (three counts) for either curve, but for the second set, the slope was reduced from approximately 110 to 33 counts/pixel. Perfect registration would have reduced the slope to 0, the value of D then depending only on noise.

Using Eq. (19) to evaluate D as a function of the true jitter s yields

$$D = \left[2\sigma_r^2(1 - \rho_s) + 2\sigma_N^2\right]^{1/2}. \tag{A1}$$

Referring to the autocorrelation model of Eq. (23), we expect that the tested values of s are all small enough ($s^2 < .06$) that the quartic term can be ignored, so that Eq. (A1) becomes

$$D = \left[2\sigma_r^2 \alpha s^2 + 2\sigma_N^2\right]^{1/2}. \tag{A2}$$

Each fit of Fig. A1 is based on a match to this model of the slope in the region $0.1 < s < 0.25$ and the intercept. The residual linearity in the lower curve implies that $\hat{s}_{\text{GEMS}}$ is proportional to the true shift s:

$$\hat{s}_{\text{GEMS}} = ks. \tag{A3}$$

To see why, note that the measured value of D (for either curve) at the intercept can be used, along with Eq. (A2), to read off the noise:

$$2\sigma_N^2 = 9. \tag{A4}$$

Then, inserting the ansatz (Eq. (A3)) into Eq. (A2) and using the upper curve of Fig. A1 results in

$$[2\sigma_r^2 \alpha / k^2]^{1/2} = 110 \tag{A5}$$

in the region of linearity, $0.1 \leq \hat{s}_{\text{GEMS}} \leq 0.25$. Again assuming Eq. (A3), the residual error after registration is $s - \hat{s}_{\text{GEMS}} = \hat{s}_{\text{GEMS}}(1/k - 1)$. If this expression is substituted for the variable s in Eq. (A2), the form for the lower curve of Fig. A1 should be produced. Its slope value of 33 then yields

$$[2\sigma_r^2 \alpha (1/k - 1)^2]^{1/2} = 33. \tag{A6}$$

Equations (A5) and (A6) admit two solutions for k, either 0.70 or 1.30; the available data do not allow the distinction to be made. At any rate, the independent results of Rauch and Zele [A1] have been explained as resulting from bias in GEMS, of the particular linear form shown in Eq. (A3). This behavior is identical to the predictions derived for method (b) (as well as the other gradient-based methods) at small values of s, shown in Fig. 9 and 11.

REFERENCE

A1. H. E. Rauch and F. Zele, "HiCAMP Background Suppression in the Presence of Jitter," Proc. Eleventh DARPA Strategic Systems Symposium, Vol. 1, Oct. 22-25, 1985, Monterey, CA. pp. 181-201.

Appendix B
TWO-DIMENSIONAL SHIFT ESTIMATES

The one-dimensional gradient-based estimates discussed in this report have natural generalizations to two (or more) dimensions. Calling s the vector shift and $g_{m,n}$ the gradient at position (m, n), with components $g^i_{m,n}$, $i = 1, 2$, we minimize the mean square error:

$$\epsilon^2 = \sum_{m,n} (y_{m,n} - x_{m,n} - \mathbf{g}_{m,n} \cdot \mathbf{s})^2, \tag{B1}$$

the difference between the sample values in the second frame (**y**) and those predicted from the sample values (**x**) of the first. The result is the estimate

$$\hat{s} = \mathbf{M}^{-1} \mathbf{V} \tag{B2}$$

where the matrix **M** and the vector **V** have components:

$$M^{i,j} = \sum_{m,n} g^i_{m,n} g^j_{m,n} .$$

$$V_i = \sum_{m,n} (y_{m,n} - x_{m,n}) g^i_{m,n} . \tag{B3}$$

To solve Eq. (B2) we must approximate the gradients in Eq. (B3). Each of the approximations in Eq. (24) has a two-dimensional analogue. For example, the canonical substitution in Eq. (24a) becomes:

$$g^1_{m,n} \rightarrow x_{m+1,n} - x_{m,n} \tag{B4}$$

$$g^2_{m,n} \rightarrow x_{m,n+1} - x_{m,n} .$$

Substituting Eq. (B4) into Eq. (B3) and using the result in Fq. (B2) then produces the two dimensional version of CAGRE, estimate (a) (Eq. (24)).