Reprinted from

Symposium on

Machine Processing of

Remotely Sensed Data

June 3 - 5, 1975

The Laboratory for Applications of
Remote Sensing

Purdue University
West Lafayette
Indiana

IEEE Catalog No.
75CH1009-0 -C

DIGITAL IMAGE RECONSTRUCTION AND RESAMPLING FOR GEOMETRIC MANIPULATION

K. W. Simon

TRW Systems Group, Redondo Beach, Calif.

I. ABSTRACT

The problems of digital image registration and geometric correction can be subdivided into two parts: 1) determination of the warping function which will transform the geometry of the scene to the desired geometric coordinate system; and 2) processing of the digital image intensity samples, given the warping function, to produce image samples on the desired coordinate grid. The latter process, called "resampling", is a subset of the problems of image reconstruction, i.e., determination of the continuous (analog) image from a set of samples of the image, and is the subject of this paper.

This paper defines the process of image resampling in more detail in terms of general imager system models, the requirements of digital image geometric manipulation and constraints of available digital processing systems. The problem is then formulated as a constrained linear estimation problem with suitable image models and optimization criteria. The resulting reconstruction filters are compared to more heuristic approaches, such as nearest neighbor, bilinear interpolation, Lagrange interpolation, and cubic convolution (cubic and quartic spline interpolators). Finally, the various resampling techniques are compared against theoretical image models, synthetically generated imagery, and actual ERTS MSS data. Nearest neighbor, bilinear, and Lagrange interpolation resamplers are shown to give significantly poorer reconstruction accuracy than TRW Cubic Convolution and the optimal constrained linear estimator.

II. PROBLEM DEFINITION

For purposes of definition of the resampling or reconstruction process, consider the imaging system shown functionally in Figure 1. The scene $f(x)$ is assumed to be a random process of the two-dimensional spatial parameter x, observed through an aperture $a(x)$ as an image $g(x)$, which is sampled by the sampler $s(x)$. The resulting samples $\underline{g}$ are available to the digital processor. In general, the imager system contains geometric error sources which preclude specification of the ideal sampler phase at the time of imaging. Thus, image samples are not available at required locations, e.g., a given map projection grid system or at the same locations sampled on an earlier imaging pass. Assuming a function is available which describes the actually-sampled locations in terms of the desired grid locations (the distortion, or warp, function), an estimator, or reconstruction filter $w(x)$ can be derived which will estimate the continuous image $g(x)$ prior to the sampler. This continuous image estimate can then be effectively evaluated at the desired grid locations, hence, the terminology "resampling". (Alternatively, the estimator could be derived to estimate the original scene $f(x)$ prior to the imaging aperture, resulting in "aperture correction," or "image restoration," as contrasted to image reconstruction. Aperture correction suffers from noise sensitivity and is not always appropriate to the processing discussed here.)

Given an infinite number of sufficiently closely-spaced, uncorrupted samples of a band-limited image, it is well-known that the original unsampled (continuous) image can be reconstructed without error by using a two-dimensional sinc function for the interpolation kernel. However, an infinite number of contributions to each interpolated value requires an infinite amount of time to process. In reality, computation time and storage limitations restrict the estimate $\hat{r}(x)$ to be a function of at most $N^2 << \infty$ image samples. In addition, imagery is seldom perfectly band-limited to an extent compatible with realizable sampling rates. The specific problem of interest here can be stated as follows: "Given an image with specified spectral density and N^2_{TOTAL} samples of that image, perhaps corrupted by measurement noise, find the appropriate interpolation function which uses $N^2 (\leq N^2_{TOTAL})$ subscene samples to estimate the image value at each point in the image with zero mean error and minimum error variance."

Heretofore, because of the processing time limitations of general purpose digital computers, image resampling has generally been accomplished by "nearest neighbor" resampling for which N=1 (each point is a function of only one sample), or by bilinear interpolation. Nearest neighbor (so-called because the intensity of the sample nearest the desired location is ascribed to the desired location) is extremely fast to compute, but causes deletion or replication of image samples and position errors of up to $\pm$ 1/2 pixel (sample spacing), significantly degrading change detection performance and giving a blocked appearance to images with large warp functions. Bilinear interpolation of the four samples surrounding the desired location resolves difficulties of nearest neighbor (at increase in number of computer operations required), but causes noticeable resolution degradation in resampled

images due to straight-line truncation of intensity peaks in the image.

Hard-wired algorithm approaches to image resampling recently have made feasible interpolators with larger N, i.e., 4 or larger. Interpolators for N=4 have been studied extensively and results are reported here and elsewhere. (Rifman, 1974; Rifman, 1975; Taber, 1973; Caron, 1974.)

III. OPTIMAL LINEAR RECONSTRUCTION ESTIMATOR

Consider a signal $g(x)$ in one-dimension with a specified autocovariance, $C_g(x)$. (Extrapolation to two-dimensions is straightforward and avoided here for clarity.) The signal mean is assumed unknown. A number of equally-spaced samples are available:

$$\underline{g}^T = [g^*(x_0),\ g^*(x_1),\ \ldots,\ g^*(x_{N-1})]$$

where the measurements are corrupted by an uncorrelated zero-mean white noise sequence $\{v_k\}$ with variance σ_v^2

$$g^*(x_k) = g(x_k) + v_k$$

A linear unbiased estimator of $g(x)$ is desired such that the estimate error variance, $J(x) \equiv E[(g(x) - \hat{g}(x))^2]$, is minimized at all x. The form of the estimator is

$$\hat{g}(x) = \underline{W}^T(x)\underline{g} + u(x)$$

A constraint is added to the minimization problem requiring a constant input to the estimator to result in the same constant estimate, i.e., $g^*(x_{k+1}) = g^*(x_k)$ all $k \Rightarrow \hat{g}(x) = g^*(x_k)$. This is equivalent to requiring that

$$\underline{W}^T(x)\ \underline{1} + u(x) = 1$$

where $\underline{1}$ is an N-vector of all ones. Using the error variance as a cost functional to which we append the constraint with a Lagrange multiplier,

$$J(x) = \underline{W}^T\hat{R}\underline{W} + R_g(o) + \sigma_v^2 - 2\underline{W}^T\underline{G} + 2M_g(\underline{W}^T\underline{1}-1)\ u+u^2 + \lambda(\underline{W}^T\underline{1}+u-1)$$

With algebraic manipulation:

$$J(x) = \underline{W}^T\hat{C}\underline{W} + C_g(o) + \sigma_v^2 - 2\underline{W}^T\underline{H} + \lambda(\underline{W}^T\underline{1}-1 + u) + [u+M_g(\underline{W}^T\underline{1}-1)]^2$$

where $R_g(x)=C_g(x)+M^2_g$ and M_g is the mean of g. Also $\hat{C}$ is the autocovariance of $\underline{g}$ and $\hat{R}$ is the autocorrelation of g. But $u(x)=W^T(x)\underline{1}+1$ from the constraint, so

$$J = \underline{W}^T\hat{C}\underline{W} + C_g(o) + \sigma_v^2 - 2\underline{W}^T\underline{H} + \lambda(\underline{W}^T\underline{1}\ 1 + u) + (1-M_g)^2u^2$$

Note that the last additive term in J is the only term involving M_g and is non-negative. Since M_g is unspecified, J must be minimized over all M_g, implying that $u=0$. In this event, J becomes

$$J = \underline{W}^T\hat{C}\underline{W}+C_g(o)+\sigma_v^2 - 2\underline{W}^T\underline{H}+\lambda(\underline{W}^T\underline{1}-1)$$

and the constraint becomes

$$\underline{W}^T(x)\underline{1} = 1$$

Minimizing J with respect to $\underline{W}(x)$ yields

$$\underline{W}^o(x) = \hat{C}^{-1}[\underline{H}(x) - \tfrac{1}{2}\lambda(x)\ \underline{1}]$$

Substituting this into the constraint equation yields

$$\lambda = \frac{2(\underline{1}^T\hat{C}^{-1}\underline{H}-1)}{\underline{1}^T\hat{C}^{-1}\underline{1}}$$

The corresponding value of J is

$$J^o(x) = C_g(o)+\sigma_v^2-\underline{W}^{oT}(x)\ \underline{H}(x) - \tfrac{1}{2}\ \lambda(x)$$

or:

$$J^o(x) = C_g(o)+\sigma_v^2-\underline{H}^T(x)\hat{C}^{-1}\underline{H}(x) + \frac{(\underline{1}^T\hat{C}^{-1}\underline{H}-1)^2}{\underline{1}^T\hat{C}^{-1}\underline{1}}$$

If a suboptimal estimator $\underline{W}'(x)$, still subject to the constraint, were used, the estimate error variance would be

$$J(x) = \underline{W}'^T(x)\hat{C}\underline{W}'(x) + C_g(o)+\sigma_v^2-2\underline{W}'^T(x)\underline{H}(x)$$

From earlier, if the mean M_g is known, then

$$\underline{W}(x) = \hat{C}^{-1}\underline{H}(x)$$

$$u(x) = -M_g(\underline{1}^T\hat{C}^{-1}\underline{H}-1)$$

and

$$J^o(x)= C_g(o)+\sigma_v^2 - \underline{H}^T(x)\hat{C}^{-1}\underline{H}(x)$$

However, if M_g is erroneously estimated as $\hat{M}_g$, then

$$\hat{J}(x) = C_g(o)+\sigma_v^2-\underline{H}^T\hat{C}^{-1}\underline{H}+(M_g-\hat{M}_g)^2(\underline{1}^T\hat{C}^{-1}\underline{H}(x)-1)^2$$

If the mean were estimated by

$$\hat{M}_g = \frac{\underline{1}^T\hat{C}^{-1}\underline{g}}{\underline{1}^T\hat{C}^{-1}\underline{1}}$$

then both the estimators and the error variances for the two approaches would coincide.

In summary, the desired estimator is

$$\hat{g}(x) = \underline{W}^T(x)\,\underline{g} \qquad \text{estimate}$$

$$\underline{W}(x) = \hat{C}^{-1}\left[\underline{H}(x) - \frac{1}{2}\lambda(x)\underline{1}\right] \qquad \text{filter}$$

$$\lambda(x) = \frac{2(\underline{1}^T\hat{C}^{-1}\underline{H}(x) - 1)}{\underline{1}^T\hat{C}^{-1}\underline{1}} \qquad \text{Lagrange multiplier}$$

$$J^\circ(x) = C_g(0) + \sigma_v^2 - \underline{H}^T(x)\hat{C}^{-1}\underline{H}(x) + \frac{(\underline{1}^T\hat{C}^{-1}\underline{H} - 1)^2}{\underline{1}^T\hat{C}^{-1}\underline{1}} \qquad \text{error variance}$$

where $\hat{C}$ is the autocovariance of $\underline{g}$, and $\underline{H}(x)$ is the crossvariance of $\underline{g}$ and $g(x)$. (The problem of aperture correction can be handled similarly by replacing $\underline{H}(x)$ with the crossvariance of the original scene $f(x)$ and the samples $\underline{g}$.) The problem remaining is the determination of the signal and measurement covariance $\underline{H}(x)$ and the measurement autocovariance $\hat{C}$.

Utilization of theoretical autocovariances or those derived from test images with much greater resolution than the subject imager and convolved with theoretical sensor and electronics apertures generally result in filters with noticeable image resolution degradation. (For N=4 and the image model of Reference 2 (PoPP, 1972), the optimum estimator is very nearly linear interpolation.) The cause of this is the relatively low spectral power at high frequencies in images relative to low spatial frequencies, i.e., high frequencies, are sparse in images in spite of their importance to visual information content. Consequently, minimum rms filters for this type of image spectrum sacrifice accuracy at the high frequencies for slight improvements at low frequencies.

In order to give suitable emphasis to the higher frequencies in the image, an error criterion weighted by an appropriate function of image spatial frequency content at each point is required. Alternatively, the image can be prewhitened for derivation of the filter, thus resulting in a filter which emp-asizes all spatial frequencies equally (up to the Nyquist rate). In the latter case, the filter is designed for an autocovariance:

$$C_g(x) = \sigma_g^2\,\mathrm{sinc}(\pi a x)$$

For N=4, σ_v=0, and a=1, the resulting reconstruction filter is shown in Figure 2A.

In practice, the filter is used to estimate only points between the central two samples of the N samples, with other points being estimated from other appropriate sets of N samples.

IV. COMPARISON WITH HEURISTIC APPROACHES

Several heuristic approaches to image interpolation suggest themselves. As an example, for N=4-point reconstruction filters, the 4-point Lagrange interpolator is well-known (passes a cubic polynomial through the four points) and is shown in Figure 2B.

A more popular approach, developed at TRW Systems and called cubic convolution, utilizes a 4-section cubic spline function as the N=4-point interpolator kernel. The spline is chosen to satisfy the following boundary conditions:

$$w(0) = 1$$
$$w(\pm 1) = 0$$
$$w(x) = 0,\ |x| \geq 2$$
$$w(x) = w(-x)$$
$$w(x)+w(1+x)+w(1-x)+w(2-x) = 1$$

thus guaranteeing exact interpolation of constant intensity areas. The first derivative of w(x) is further constrained to be continuous, guaranteeing continuity of the first derivative of the interpolated signal. The resulting interpolator has one remaining degree of freedom, a. If the parameter a is chosen for exact constant slope interpolation, i.e., $-w(1+x)+w(1-x)+2W(1-x) = x$, then the resulting interpolator is as shown in Figure 2C. For continuity of second derivative of w(x) at $|x| = 1$, the resulting interpolator is as shown in Figure 2D. For the derivative of w(x) at x=1 to be the same as that of $\mathrm{sinc}\pi x$, the interpolator is as shown in Figure 2E. Alternatively, a quartic spline can be defined to satisfy the above boundary conditions plus the additional constraint of continuity of second derivative of w(x). This interpolation is shown in Figure 2F.

Reconstruction error for a Gaussian test function $e^{-x^2/2}$ was calculated for each of the above filters and several others for several sample phasings. Some of these errors are plotted in Figure 3 for two sampler phasings. The comparison using an error function as test signal gave similar results.

A second comparison was made using ERTS MSS data. The data samples were resampled on a grid shifted from the input sampled grid by 1/2 pixel along-scan using a high-order sinc interpolator (N=30, or 900 samples per output point). The same was then done for nearest neighbor (N=1), bilinear interpolation (N=2), cubic convolution (N=4), and a truncated sinc interpolator with N=10. The resampled images were differenced pixel-by-pixel with the 900-point sinc interpolation. Difference images and corresponding histograms are shown in Figure 4. Note that the 16-point cubic convolution yields lower error than the 100-point truncated sinc interpolator.

A third comparison was made by taking a high resolution digital image (3.4m sample spacing) and convolving it with an aperture similar to the EOS thematic mapper, i.e., a scanning square detector shape with an integrate-and-dump sampler. Samples were then extracted every 20m along-scan and every 28m across-scan and used to reconstruct the continuous image (as convolved with the aperture) using several of the above techniques. Difference images and error histograms are shown in Figure 5. The resolution degradation inherent in 2-point interpolators is apparent. Remaining errors in the 4-point cubic convolution are due primarily to insufficient sampling rate (aliasing).

A fourth comparison involved registration of two successive scenes of the same area (Baltimore, Md.) taken by ERTS MSS using nearest neighbor resampling and cubic convolution. The registered images were differenced and the difference images are shown in Figure 6. The errors inherent in low-order resampling are apparent.

V. REFERENCES

1. R.H. Caron, "Application of Advanced Signal Processing Techniques to the Rectification and Registration of Spaceborne Imagery," Proc. of First Houston Tech. Transfer Conference, Sept. 1974, pp 245-255.

2. D.J. Popp, D.S. McCormack, and J.L. Sedwich, "Imagery Correlation and Sampling Study," Report MDC A1740, McDonnell Aircraft Co., St. Louis, Mo., June 1972.

3. S.S. Rifman, "Digital Rectification of ERTS Multispectral Imagery," Symp. on Significant Results Obtained from ERTS-1, (NASA SP-327), Vol. I, Section B, pp. 1131-1142, Mar. 5-9, 1973.

4. S.S. Rifman and D.M. McKinnon, "Evaluation of Digital Correction Techniques for ERTS Images - Final Report," Report No. 20634-6003-TU-00, TRW Systems, Redondo Beach, Calif., July 1974.

5. S.S. Rifman, W.B. Allendoerfer, D.M. McKinnon, and K.W. Simon, "Experimental Study of Digital Image Processing Techniques for ERTS Data - Task I Final Report," Report No. 26232-6001-RU-01, TRW Systems, Redondo Beach, Cal., Jan. 1975.

6. J.E. Taber, "Evaluation of Digitally Corrected ERTS Images," Third ERTS-1 Symposium, Vol. I, Tech. Presentations, (NASA SP-351), Dec. 1973, pp. 1837-1843.

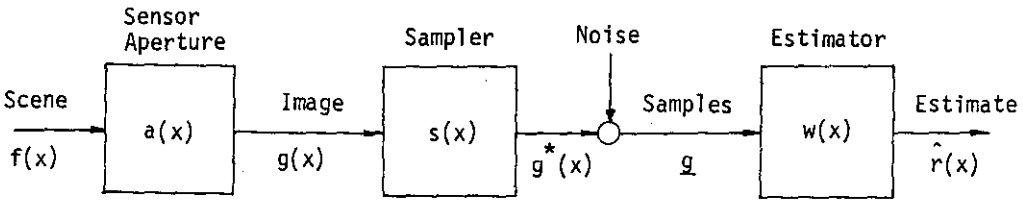

FIGURE 1. IMAGER MODEL

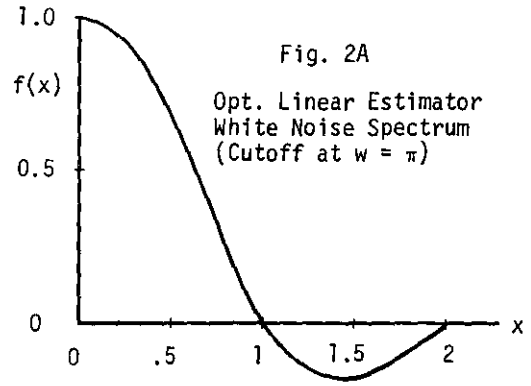

Fig. 2A

Opt. Linear Estimator
White Noise Spectrum
(Cutoff at w = π)

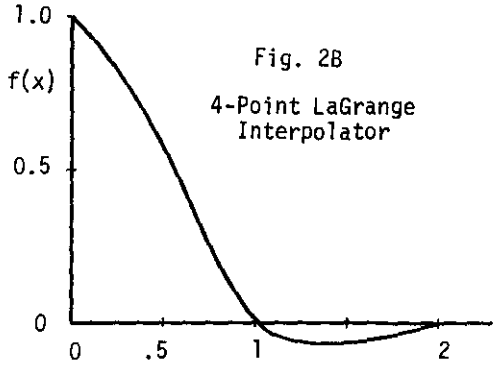

Fig. 2B

4-Point LaGrange
Interpolator

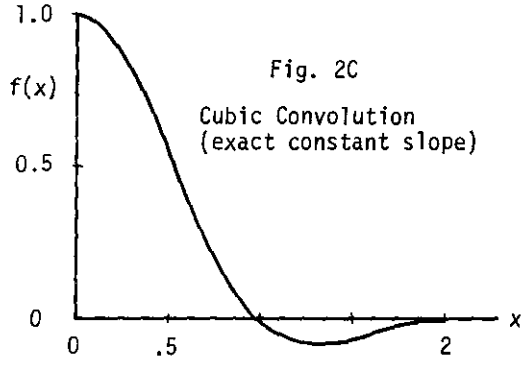

Fig. 2C

Cubic Convolution
(exact constant slope)

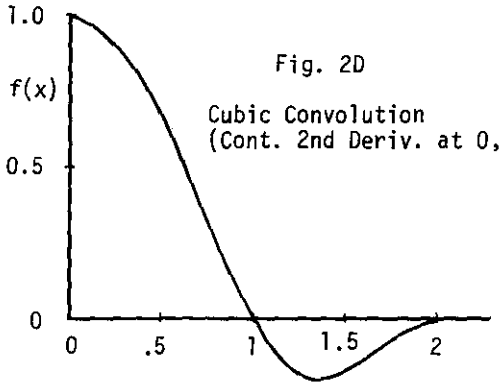

Fig. 2D

Cubic Convolution
(Cont. 2nd Deriv. at 0,1)

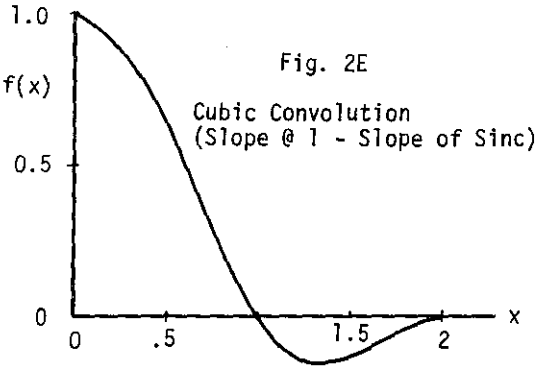

Fig. 2E

Cubic Convolution
(Slope @ 1 - Slope of Sinc)

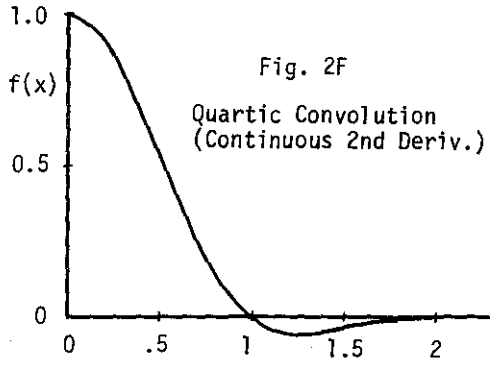

Fig. 2F

Quartic Convolution
(Continuous 2nd Deriv.)

FIGURE 2. INTERPOLATION KERNELS

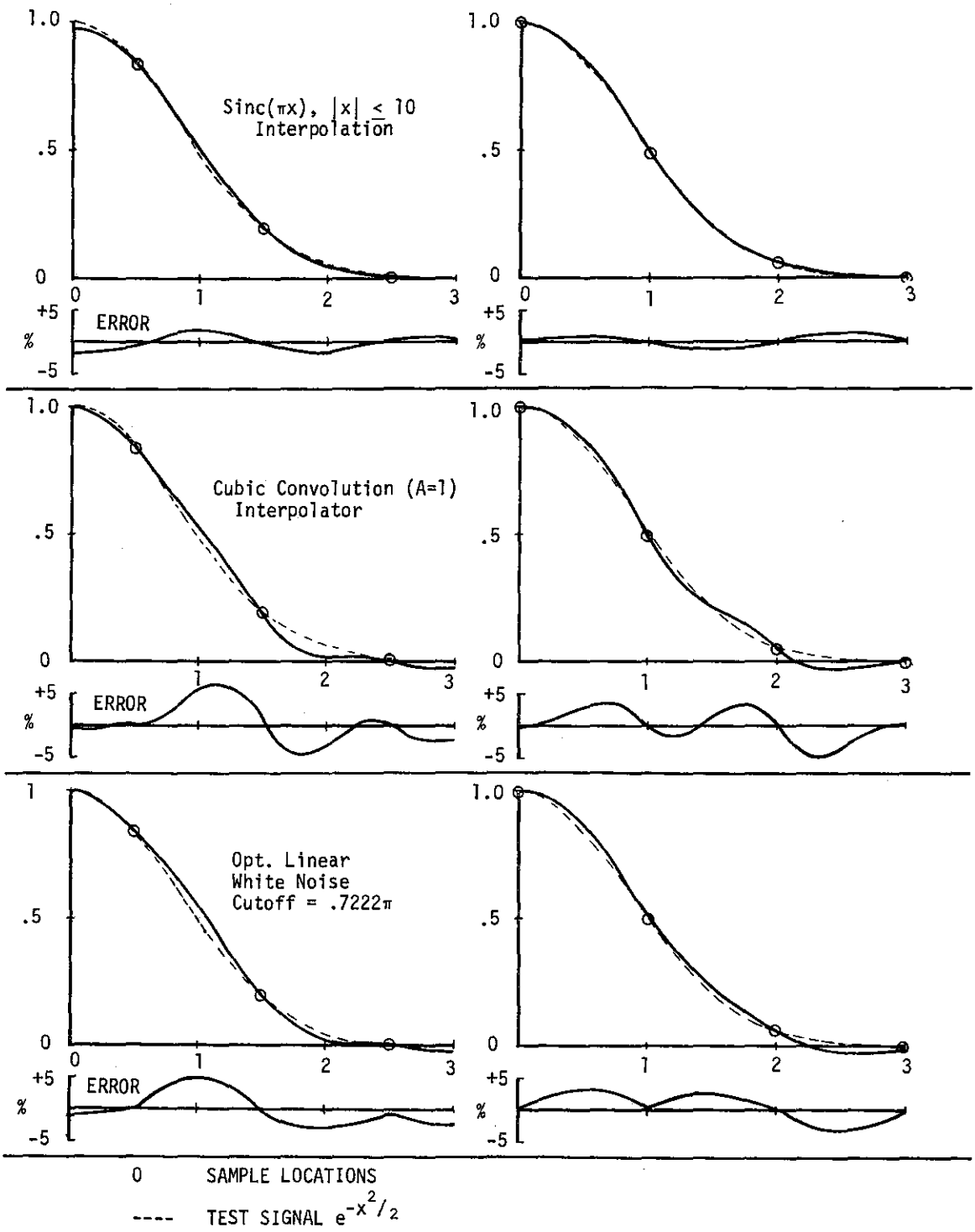

FIGURE 3. INTERPOLATION ERRORS

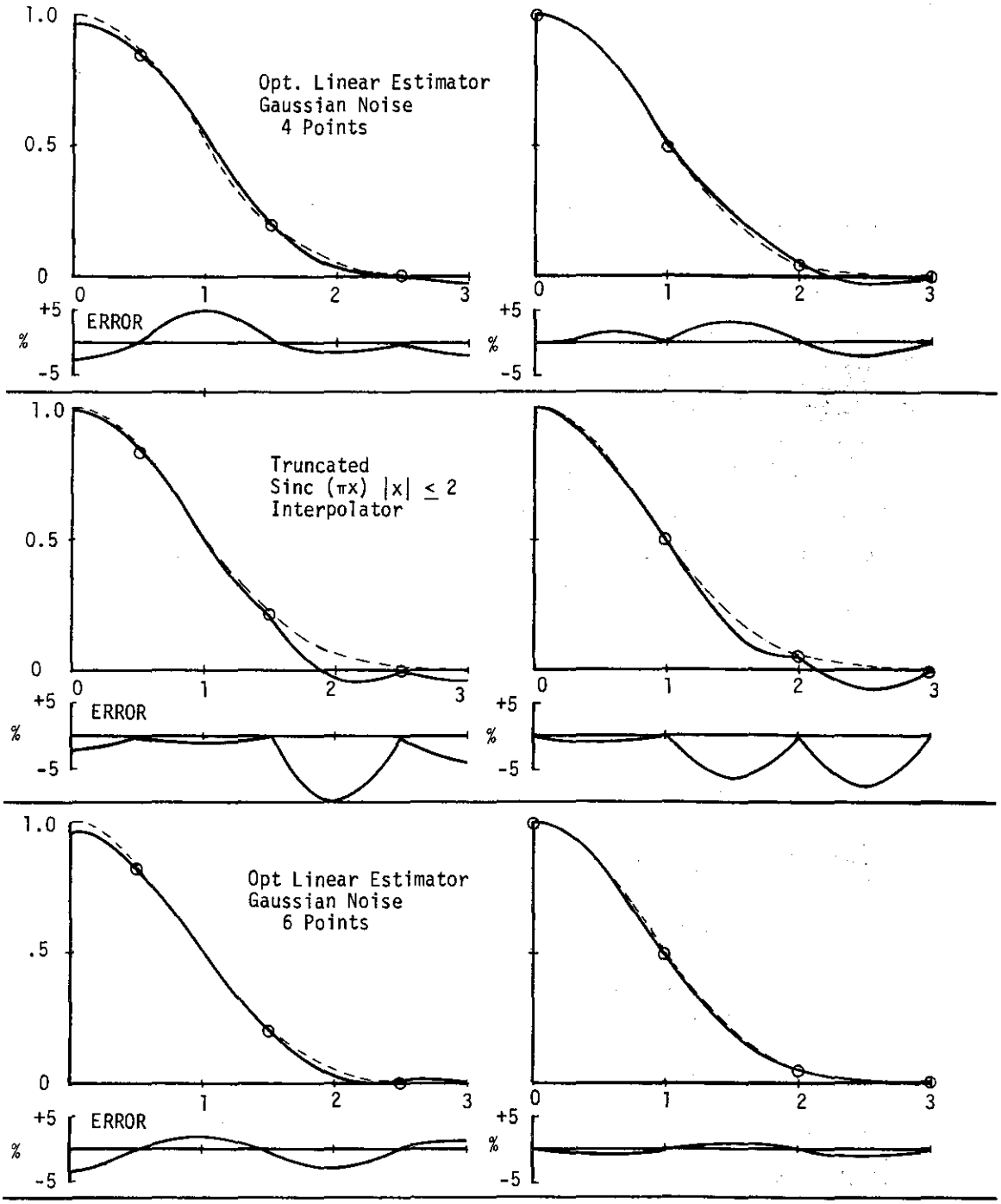

FIGURE 3 (CONTINUED) INTERPOLATION ERRORS

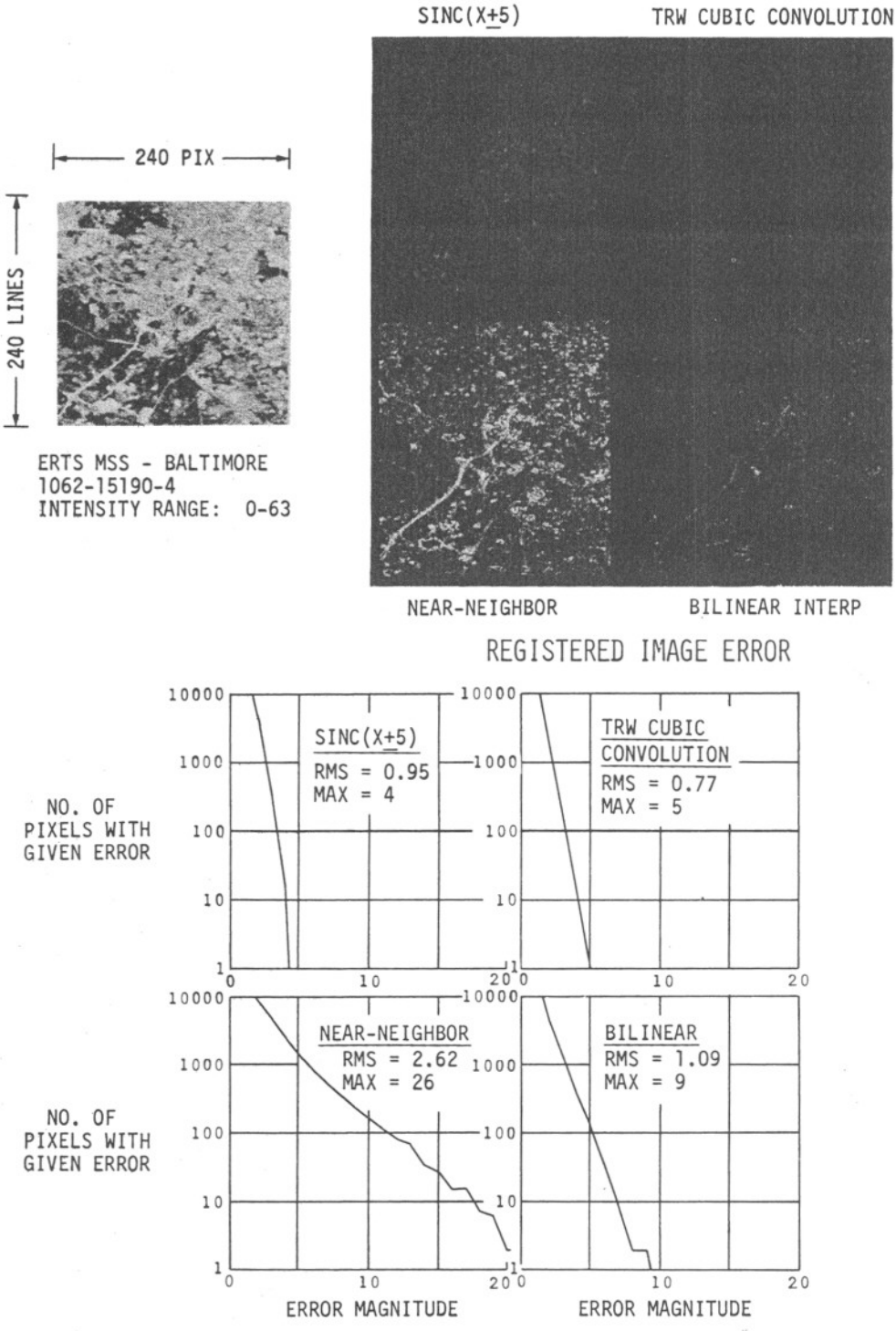

FIGURE 4 ERROR BETWEEN REGISTERED IMAGES DUE TO RESAMPLING

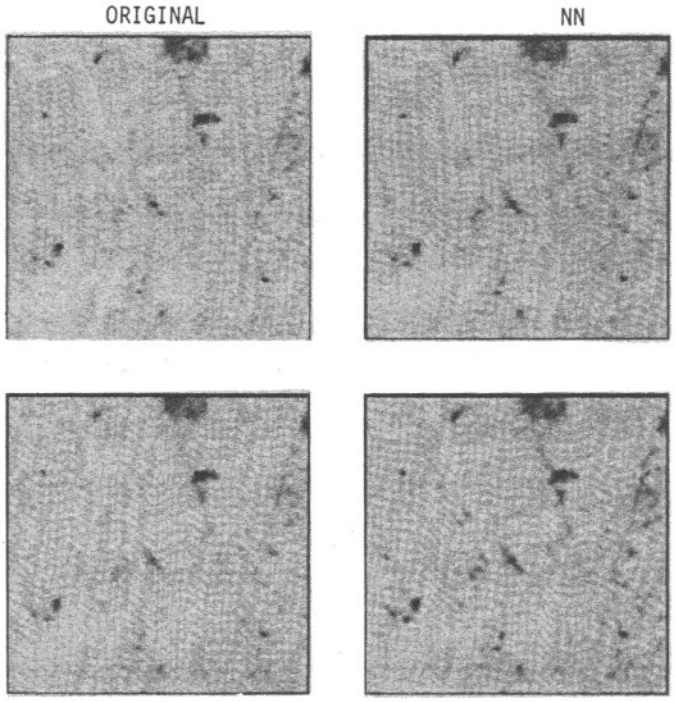

FIGURE 5A - 30M RESOLUTION IMAGE RECONSTRUCTION (54 x 78 PIXELS)

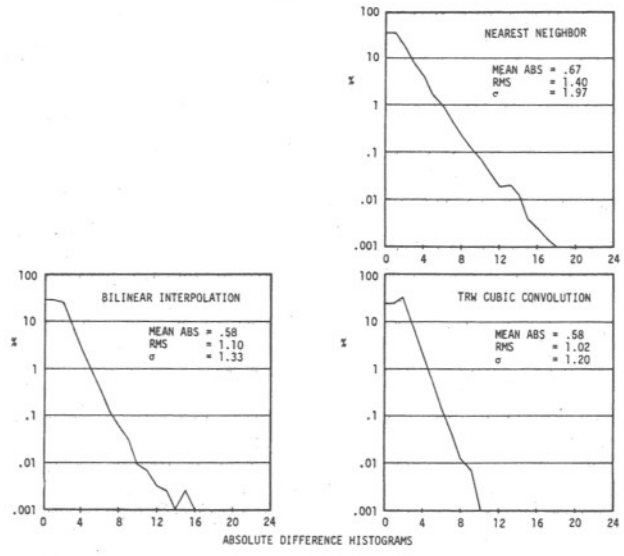

FIGURE 5B - RECONSTRUCTION DIFFERENCE IMAGES

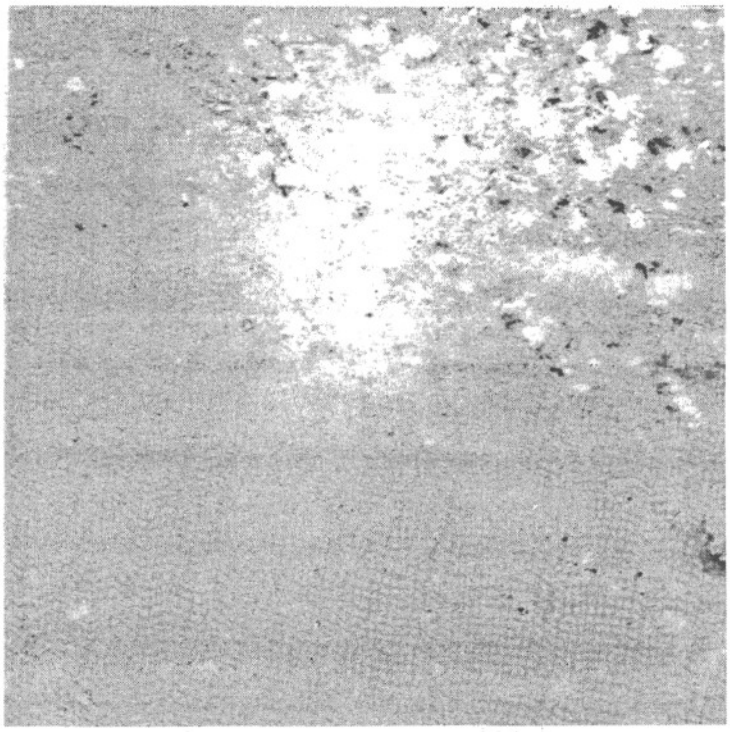

FIGURE 6C - DIFFERENCE IMAGE USING TRW CUBIC CONVOLUTION FOR WARPING
(GRAY = NO ERROR)

FIGURE 6D - DIFFERENCE IMAGE USING NEAREST NEIGHBOR RESAMPLING

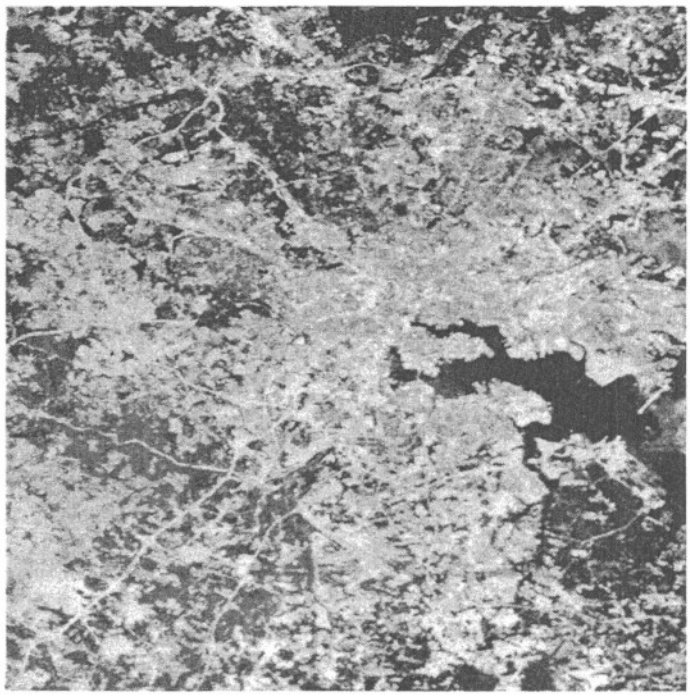

FIGURE 6A - ERTS BALTIMORE (SEPT '72) (512 x 512 PIXELS)

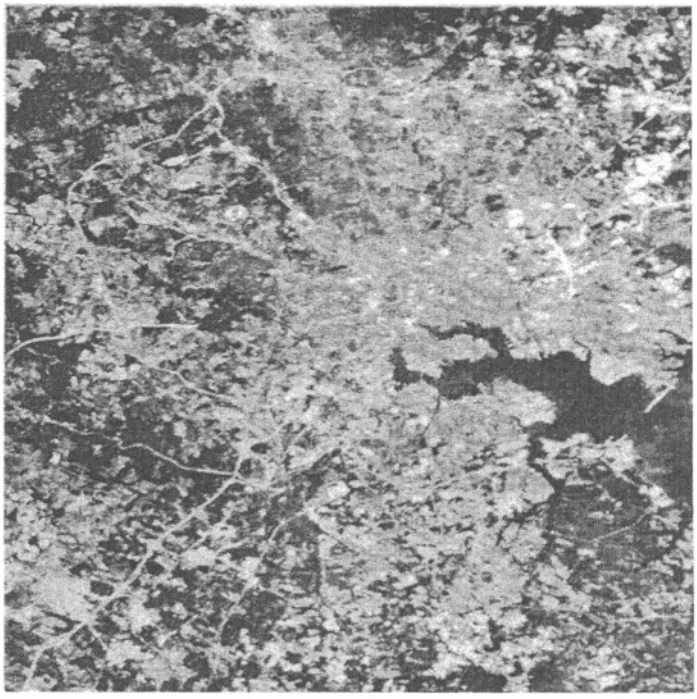

FIGURE 6B - ERTS BALTIMORE (OCT '72) REGISTERED TO PREVIOUS SCENE
(512 x 512 PIXELS)