Distribution Category:
General, Miscellaneous,
and Progress Reports
(UC-700)

ANL/ACTV-91/4 REVISED

ARGONNE NATIONAL LABORATORY
9700 South Cass Avenue
Argonne, Illinois 60439

UNDERSTANDING CORRELATION COEFFICIENTS IN TREATY VERIFICATION

by

A. DeVolpi

Engineering Physics Division

November 1991
(Revised February 1993)

Work sponsored by
U.S. Department of Energy
Office of Arms Control

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

UNDERSTANDING CORRELATION COEFFICIENTS IN TREATY VERIFICATION

by

A. DeVolpi

ABSTRACT

When a pair of images is compared on a point-by-point basis, the linear-correlation coefficient is usually used as a measure of similarity or dissimilarity. This report evaluates the theoretical underpinnings and limitations of the linear-correlation coefficient, as well as other related statistics, particularly for cases where inherent white noise is present. As a result of the limitations in linear-correlation, an additional step has been derived -- local-sum clustering -- in order to improve recognition of small dissimilarities in a pair of otherwise identical images. Results show an optimal three-stage procedure: first, establish congruence of the two images; second, use the linear-correlation coefficient as a test of true negatives; and, third, qualify a true positive by using the cluster (local-sum) method. These three algorithmic stages would be especially useful in application to arms control treaty verification, particularly for comparison of unique identifiers (tags or seals). This is illustrated by comparing scanning-electron microscope topographical images for an intrinsic-surface tag.

INTRODUCTION

Statistical methods have a potentially important role in validating collected arms control treaty verification data and in optimizing time and resources. A comparison of two images is a good example; such image pairs are often formed or reconstructed from tamper-resistant seals or tags used as unique identifiers of arms control treaty-limited equipment. An initial image is created when the seal or tag is placed on the item, and another image is collected when the item is verified, possibly years later. In order to quantify the image comparison, thereby removing subjective human judgment as an evaluation factor, a normalized correlation coefficient is usually created. This coefficient is expected to have a value close to unity when the two images are essentially the same and close to zero when they are entirely different images. The purpose of this paper is to focus and extend the theory and application of correlation coefficients so their uses and limitations can be better understood in a treaty verification context.

The result of this analysis is a three-stage process for a verification algorithm that provides more utility than the linear-correlation coefficient alone.

ANALYSIS OF CORRELATIONS

Let us start by reviewing the simplified mathematical treatment of two directly measured results A and B, whose functional relationship is anticipated by the theoretical expression

(1) $F = F[A(x,y,z),B(x,y,z)]$,

where A and B are functionally dependent on the causative parameters x, y, and z.[1] These parameters could be, for example, the measures (vector components) of a voxel in three-dimensional space. For simplicity in notation, the third parameter will henceforth be omitted without loss in generalization. Thus,

(2) $F = F[A(x,y),B(x,y)]$.

From statistical theory, the best estimates of the parameters x and y are their mean values for i=1,N where N is the size of the sample population:

(3) $\langle x \rangle = \Sigma x_i/N$ (and $\langle x^2 \rangle = (\sum x_i^2)/N$)

(4) $\langle y \rangle = \Sigma y_i/N$.

Also,

(5) $\langle F \rangle = \Sigma F_i/N$

is the minimum-variance unbiased estimate of F.

Next we evaluate the computed covariance of A and B, which would be

(6) $\Gamma_{AB} = \langle A*B \rangle - \langle A \rangle \langle B \rangle = \Sigma(A_i*B_i)/N - (\Sigma A_i*\Sigma B_i)/N^2$.

The correlation coefficient, which normalizes the covariances to the range ($-1 \leq \rho_{AB} \leq 1$), is

(7) $\rho_{AB} = \Gamma_{AB}/\sigma_A\sigma_B$,

where the standard deviations in A and B are σ_A and σ_B, which are measurable from the samples.

A_i and B_i are pixel values for images A and B at identical coordinates i, each pixel having a normalized value or magnitude (intensity or grey scale) represented by A_i and B_i. If the two images A and B are identical, but their grey scales are uniformly displaced by some linear correction factor, then we could write a relationship such that the values of B are linearly related to the values of A. The linear correlation coefficient ρ_{AB} is the usual statistical measure of that relationship. Procedurally, one can first align for shift, magnification, and rotation differences to ensure that two images are in registration (pixel congruence) by finding the highest value of the linear correlation coefficient of the images or of some fiducials in the images.

To estimate the standard deviation, let us assume that the images represented by the populations A_i and B_i each have their own random and completely independent (instrument) white noise values a_i and b_i respectively and that B is linearly related to A through their intrinsic intensities x_i and y_i, such that

(8) $A_i = x_i + a_i$ and

(9) $B_i = y_i + b_i = k + mx_i + b_i$,

where k and m are constants of the image system.

In this case,

(10) $\sigma_A^2 = \sigma_x^2 + \sigma_a^2$, and

(11) $\sigma_B^2 = m^2\sigma_x^2 + \sigma_b^2$,

which means that the slope of the linear relationship would in the absence of noise be

(12) $m = \sigma_B/\sigma_A$.

The autocorrelation in A can be computed, giving

(13) $\Gamma_{AA} = \langle A*A \rangle - \langle A \rangle * \langle A \rangle = \sigma_x^2 = \langle x^2 \rangle - \langle x \rangle^2$,

so that another way of computing m is

(14) $m^2 = \left(\sigma_B^2 - \sigma_A^2\right)/\Gamma_{AA}$

if the variance of the noise were equal for A and B.

The autocorrelation in B would be

(15) $\Gamma_{BB} = \langle B*B \rangle - \langle B \rangle * \langle B \rangle = m^2\sigma_x^2$,

and without any loss in generality,

(16) $m^2 = \Gamma_{BB}/\Gamma_{AA}$.

The covariance in angle-bracket notation is

(17) $\Gamma_{AB} = \langle A*B \rangle - \langle A \rangle * \langle B \rangle$.

Even if there is noise in the measurement, the (linear) covariance of Eq. 17 becomes from Eqs. 8 and 9

(18) $\Gamma_{AB} = m\left(\langle x_i^2 \rangle - \langle x_i \rangle^2\right) = m\sigma_x^2$.

With these assumptions the linear cross-correlation coefficient becomes

(19) $\rho_{AB} = m\sigma_x^2/\sigma_A\sigma_B = m/\left[\left(1+\sigma_a^2/\sigma_x^2\right)\left(m^2+\sigma_b^2/\sigma_x^2\right)\right]^{1/2}$.

Equation 19 corresponds to the coefficient for linearly cross-correlated functions A and B which each have random errors in their measurement, and σ_x is a measure of the inherent deviation between pixel (intensities) in images A and B.

We can also look at the linearly correlated difference image, that is the subtraction of A-B:

(20) $\Gamma_{A-B} = \langle (A-B)*(A-B) \rangle - \langle A-B \rangle * \langle A-B \rangle = (1-m)^2\sigma_x^2$, and

(21) $\rho_{A-B} = (1-m)^2\sigma_x^2/\sigma_A\sigma_B = (1-m)^2/\left[1+\left(\sigma_a^2/\sigma_x^2\right)\left(m^2+\sigma_b^2/\sigma_x^2\right)\right]^{1/2}$.

Observe that the correlation <u>difference</u> coefficient vanishes as the slope $m \rightarrow 1$ but could become very large if m and $\sigma_b \rightarrow 0$.

It is assumed that the two images are represented by N sets of data A_i and B_i, where each set of x_i and y_i are pixel locations on a two-dimensional plane, and the values of A_i and B_i are the components that contain the values to be compared at each point. The sets of data A and B might be values sensed by a scanning electron microscope (SEM) of electrical current that would be related to the depth of the point of electron scattering from a reference plane. Or they might be some other quantification of pixel values, such as the consolidated area of a reflecting contour point in the reflective-particle tag image. Optimization of the correlation coefficient will minimize systematic shifts, rotations, and distortions when attempting to make two images congruent. Having carried out that minimization through maximization of ρ, now the question is whether the two images are identical.

APPLICATIONS OF CORRELATION COEFFICIENTS

To assist in interpreting Eq. 19, first note that $\rho \rightarrow \pm 1$ as $\sigma_a \rightarrow 0$ and $\sigma_b \rightarrow 0$. In other words, a perfect correlation (or anti-correlation) is approached as two noiseless images of the same scent are compared by a system that introduces no noise, randomness, or bias in the measured values of the image amplitudes. Also note that if the linear displacement m between images is small, $\rho \rightarrow 0$, as expected.

Let us consider further some cases where $k \rightarrow 0$ and $m \rightarrow 1$, that is, A and B are congruent except for noise jitter. This ought to be the case for accurate digitization for two sets of data taken from the same original image. Then, $\rho_{AB} \rightarrow 1$, and any differential short of unity is a measure of the noise introduced in the image comparison. If we go one step further and choose one image A as a noise-free reference or baseline image, to be compared with the other image B, we could stipulate that all the net (comparative) systematic and random differences reside in the second image. Then

$$(22) \quad \rho_{AB} = \left(1+\sigma_b^2/\sigma_x^2\right)^{-1/2} \left(\text{for } \sigma_a=0 \text{ and } m=1\right) .$$

This expression has an interesting node at $\sigma_b = \sigma_x$, namely the value of 0.707. This function is plotted in Figure 1. Figure 1 can be useful in estimating the relative signal-to-noise-ratio from the measured correlated coefficient, or vice versa, if the signal-to-noise ratio corresponds to the variance ratio. Also assumed here is that any systematic differences can be represented by the constants k and m. In Fig. 1, the calculated and measured data show how random noise degrades the linear correlation coefficient. A very noisy image of an identical noiseless scene can create a false negative LCC.

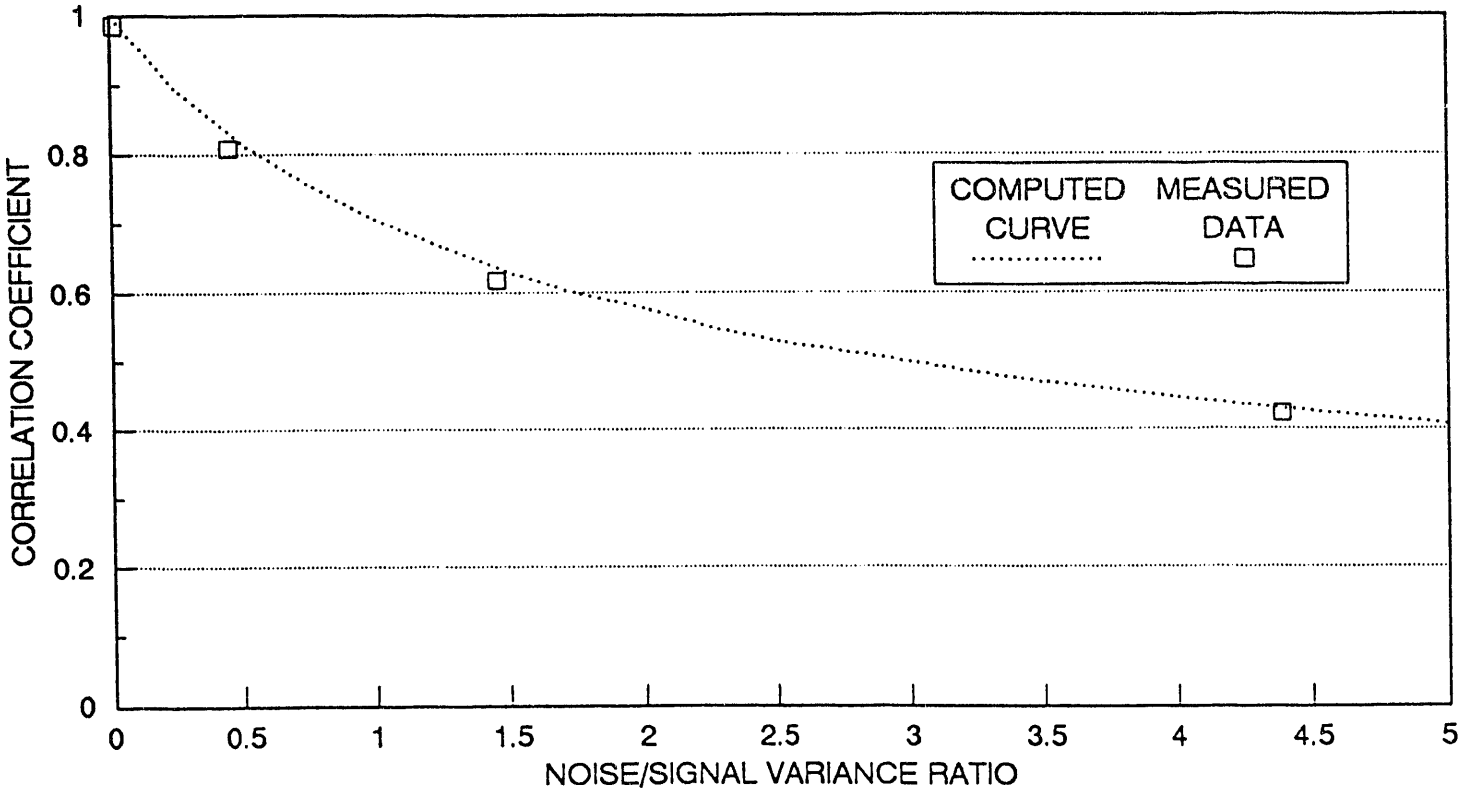

Figure 1. Dependence of Correlation Coefficient on Variance Ratio. The calculated and measured data show how random noise degrades the linear correlation coefficient. A very noisy image of an identical noiseless scene can create a false negative LCC.

Another interesting case is where both images are uniform (essentially featureless) throughout -- that is, there is no variation from pixel to pixel in either image. In that situation, any variances in the image are due to random noise, and the covariance and correlation coefficients would approach zero. A very small number of non-uniform features could result in a correlation coefficient slightly greater than zero, but the variance in that coefficient might be large enough to mask the true value of the correlation coefficient. In effect, σ_x would then be a measure of intrinsic image dispersion or non-uniformity of features.

For each sector of an image, similar conclusions regarding variances and covariances can expected. But as the number of pixels in each sector becomes smaller, the variance in the correlation coefficient will increase. If, for example, a bubble or defect in replica casting

occurs in a sector, the coefficient for that sector will become smaller, defining a "bad" sector. For two images to be considered identical, a standard will have to be set for the number of acceptable bad sectors.

When two entirely different images are compared with each other, the linear cross-correlation coefficient is expected to go to zero; however, the coefficient is a statistic that is subject to variance, and for a small number of pixels subject only to white noise the estimate of ρ might exceed the null value. To test the null hypothesis that $\rho = 0$, the parameter

(23) $t = \rho(N-2)^{1/2}\left(1-\rho^2\right)^{-1/2}$ can be used for the tabulated Student's distribution.

The number of samples N needed to achieve different probabilities that $\rho = 0$ is given in Table I as a function of the measured values of ρ. When only a half-dozen pixels are compared, there is a 10% chance of getting a false positive ($\rho>0.6$). Even a 50-pixel comparison could yield an ambiguous result ($\rho>0.3$) once in a hundred times.

Figure 2 schematically indicates two image-generalized pathways that occur, for example, in surface-image verification, resulting when plastic casting (fingerprints) are made of the intrinsic surface roughness. In the first case, there is a surface S which contains the intrinsic signature. The surface features could be measured directly with an SEM and used as a reference for comparing the plastic castings. In that case, possible non-linearities in the two pathways would have to be investigated. More routinely, two castings A and B made from the same surface S will be compared, with A being taken during the baseline and therefore becoming defined as the reference image. In fact, the sets of data for A and B are derived

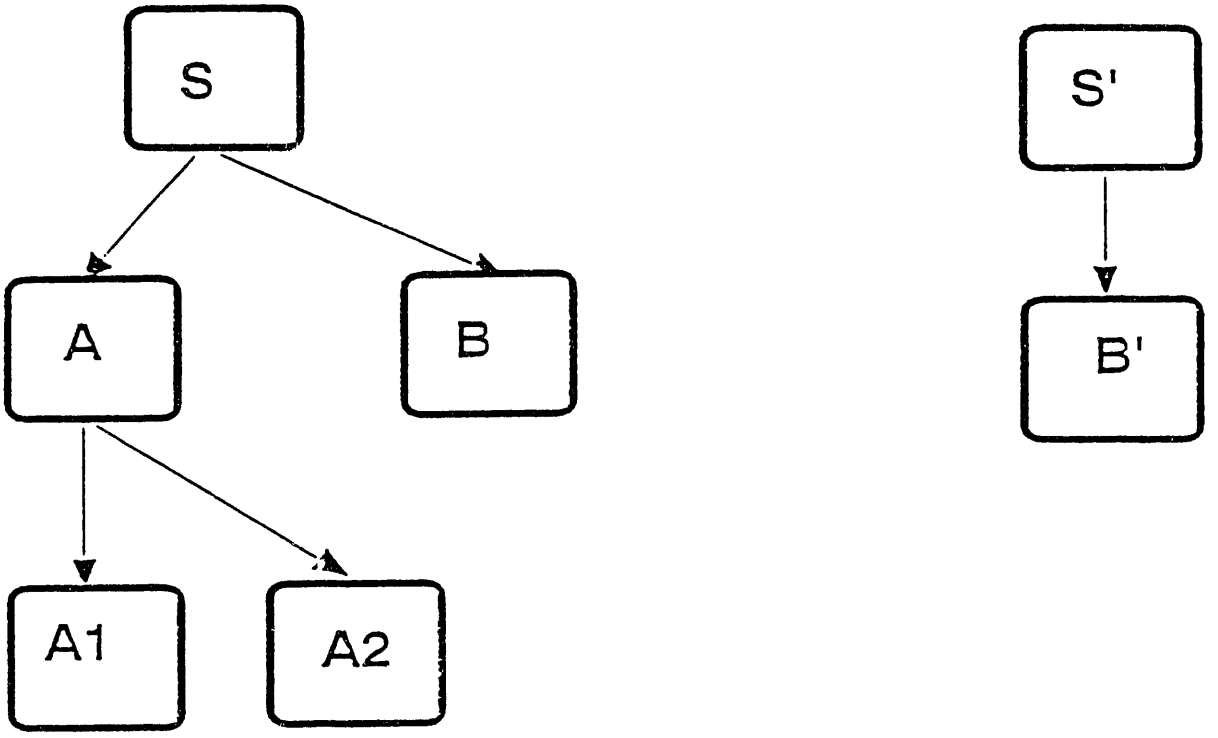

Figure 2. Pathways for Comparing Images. On the left-hand side are comparisons of two castings, A and B, of the same surface S. One of these castings might be digitized twice (A1 and A2) in order to evaluate digitization-instrumentation system noise. On the right-hand side is the pathway for a casting made from a dissimilar surface S′, from which we expect true negative to result from comparison of B and B′.

from a mensuration process using an SEM. In order to determine if this conversion process introduces noise into the comparison, the sample A could be digitized twice, resulting in data sets A_1 and A_2 which could be cross-correlated. This cross-correlation coefficient should give a measure of the instrument noise, while the correlation between A and B will give a combined measure of casting effects and instrument noise. In the pathway where two different surfaces S and S′ are to be compared, here we expect to test the null hypothesis for the correlation coefficient between A and B′. As additional measures of confidence, it would be useful to calculate the autocorrelations and difference correlations as a matter of routine practice.

TABLE I

APPROXIMATE NUMBER OF PIXELS REQUIRED TO CONFIRM A NULL HYPOTHESIS*

	PROBABILITY THAT TRUE VALUE OF ρ_{ab} IS ZERO						
Measured ρ_{ab}	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
0.9	4	5	8	11	13	16	17
0.8	4	8	12	16	20	25	28
0.7	5	10	16	21	28	36	43
0.6	6	16	24	32	43	53	63
0.5	8	20	35	50	65	80	95
0.4	12	30	55	80	105		
0.3	20	57	100				
0.2	45	135					
0.1	165						

*The true correlation coefficient is zero ($\rho_{ab}=0$) when the measured coefficient is ρ_{ab}.

TABLE II

COMPARISON OF COMPUTER-GENERATED NOISE-ADDED EFFECTS ON LINEAR CROSS-CORRELATION COEFFICIENT WITH CALCULATED COEFFICIENT BASED ON EQ. 22

Sample Number	Case	Measured Variance	Estimated Noise Variance	Estimated Noise/Signal Variance	Measured Correlation	Calculated Correlation
5321	reference	711	10.4	0.015	0.985	0.99
5322	reference	674	10.4	0.016		
5342	noise added	992	310.0	0.45	0.809	0.83
5352	noise added	1679	997.0	1.46	0.618	0.64
5362	noise added	3678	2996.0	4.39	0.423	0.43

As shown in Table II, the value of the correlation coefficient indeed departs from 1.0 as computer-generated (white) noise is added to images (of Fig. 3) without changing their inherent correlation. Although the "reference" cases in Table II are not noise-free, they are close enough to confirm the role of Eq. 22. Figure 3C indicates that care was taken to place the artificial noise within the digitization range of the system. The data tabulated from Fig. 3 also confirm the relationship of signal/noise and variance ratio expressed in Eq. 22 and depicted in Fig. 1. In each case of Fig. 3C, the lower and upper tails of the amplitude size distributions are within the digitization range (up to 256 bits).

As a rule of thumb, to avoid a false positive (or ambiguous) conclusion about the similarity of two images at a high (0.999999) level of confidence requires at least 100 pixels to be compared. On the other hand, if the measured value of $\rho_{ab} = 0.9$, then 16 pixels would give the same high level of confidence.

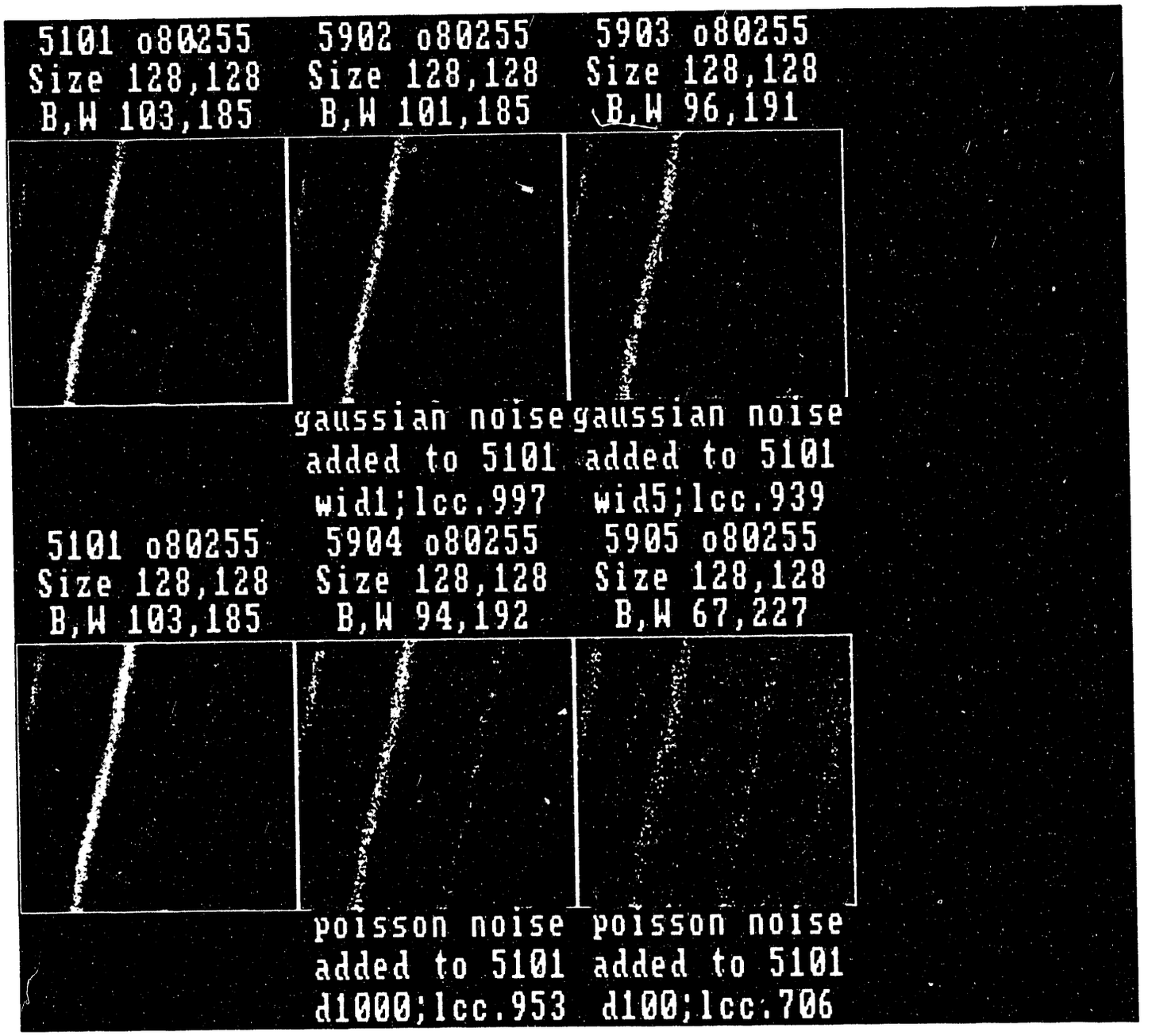

Figure 3A. **Output of SEMPER Noise Command, Showing Effects of Increasing Computer-Generated White Noise (Adding Gaussian and Poisson Noise)**

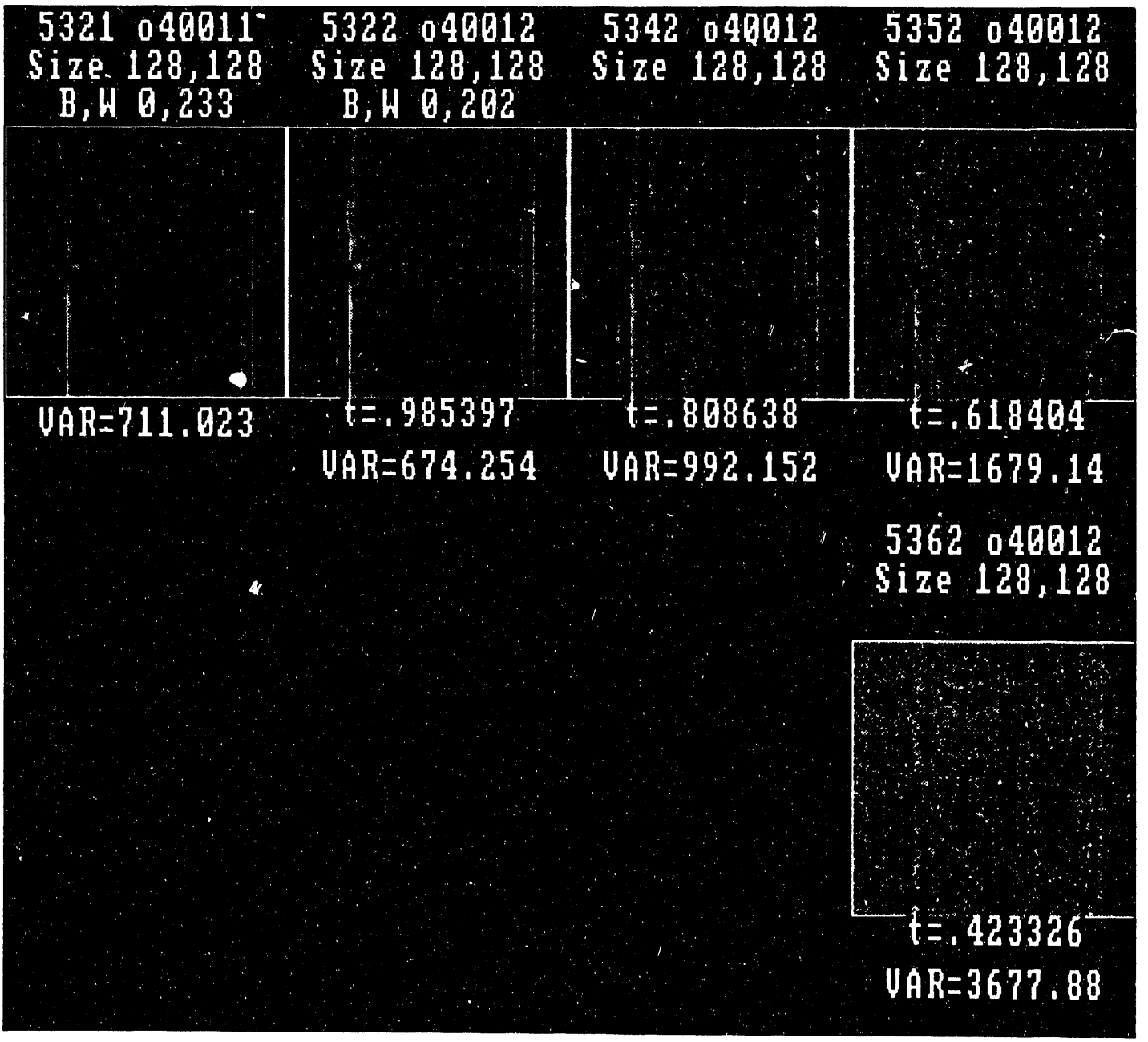

Figure 3B. Effect of Added Noise on Variance (VAR) and Linear-Correlation Coefficient (t)

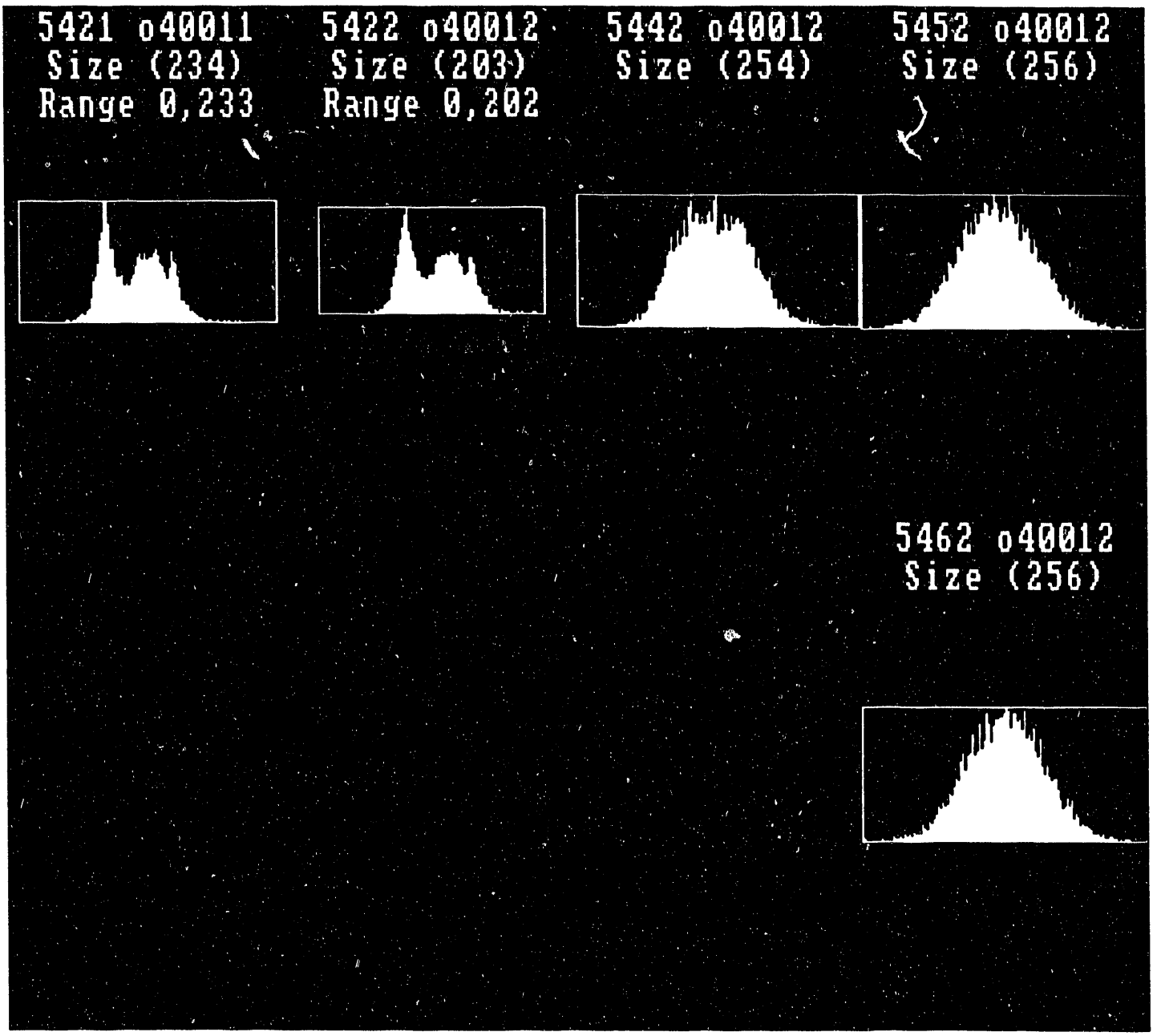

Figure 3C. Confirmation that Noise Level is Within Digitization Range of System. In each case the lower and upper tails of the amplitude size distributions are within the digitization range (up to 256 bits).

In Fig. 2, the left-hand side shows comparisons of two castings, A and B, of the same surface S. One of these castings might be digitized twice (A1 and A2) in order to evaluate digitization-instrument noise. On the right-hand side is the pathway for a casting made from a dissimilar surface S′, from which we expect a true negative to results from comparison of B and B′.

Table I applies when two images are known or expected to be entirely different (null hypothesis that the true correlation coefficient is zero, i.e., $\rho_{ab}=0$). The question answered by this table is, what actual values of ρ_{ab} would be obtained if only a small number of pixels are compared? Table I shows that unless many pixel values are compared, correlation coefficients significantly greater than zero can be computed; in fact, with just a few pixels intercompared, there would be a fairly high probability of getting a correlation coefficient close to 1.0.

MEAN-SQUARE-DEVIATION

A statistic sometimes used is called the "mean-square-distance" (MSD), which is related to the mean-square-deviation:

$$(24) \quad MSD = \frac{1}{2N} \sum_{ij} \left(A_{ij} - B_{ij}\right)^2 = \frac{1}{2N} \sum \left(A_i - B_i\right)^2.$$

This expression's relationship to the linear correlation coefficient can be determined by converting it to nomenclature used in this paper by setting the mean values <A> and <B> exactly to zero and the variances $\sigma_A = \sigma_B = 1$.
Then,

$$(25) \quad \Gamma'_{AB} = \langle A*B \rangle = \rho_{AB} = \sum A_i B_i / N,$$

$$(26) \quad \sigma_A^2 = \frac{\sum A_i^2}{N} - \left(\frac{\sum A_i}{N}\right)^2 = \frac{\sum A_i^2}{N} = 1,$$

and likewise for σ_B^2 because $\langle A \rangle = \frac{\sum A_i}{N}$ and $\langle B \rangle = \frac{\sum B_i}{N}$.

Hence,

$$(27) \quad MSD = \frac{\sum A_1^2}{2N} - \frac{2\sum A_i B_i}{2N} + \frac{\sum B_i^2}{2N} = 1 - \rho_{AB}.$$

Consequently the MSD is subject not only to the two given prescriptions for the mean and variance of each image-normalized distribution, but also to the limitations of the linear cross-correlation coefficient expressed through Eq. 19.

The MSD could be a useful tool for simply forcing the best congruence of two images: its calculation in a computer can usually be performed faster than the linear correlation coefficient because differences execute faster than multiplications. However, the MSD contributes nothing more to the understanding of image similarity beyond the linear cross-correlation coefficient as a normalized measure of correlation.

VERIFICATION ALGORITHM

For actual application to treaty-verification analysis of images, a two-step process appears optimal. The first step is to use Eq. 18 or 19 or the MSD (Eq. 24) to find as good a degree of congruence as possible. After this has been accomplished, for the second stage the difference correlation (Eq. 21) can be computed in order to highlight significant differences in images, e.g., identify replication attempts.[2]

In actual practice, data might be available in the form of a set of numbers for each pixel position i,j. The values in the matrix A might represent grey levels or other amplitude information, or they might be quantized into binary values.

After optimizing congruence of two images A and B, two data matrices A_{ij} and B_{ij} would exist. A normalized difference-correlation coefficient ρ_{A-B} can then be computed in any of three ways, depending on whether or when the values are quantized to a binary set. If threshold values A_{th} and B_{th} are first determined, then sets of binary values can be created:

(28) $I_{ij} = (A_{ij} - A_{th})/\langle A \rangle$ and

(29) $J_{ij} = (B_{ij} - B_{th})/\langle B \rangle$,

such that I_{ij} and J_{ij} equal 1 if the differences are positive and 0 if the differences are zero or negative.

Another way would be to compute the average $\langle A_{ij} - B_{ij} \rangle$ and quantize each difference value by a similar process into binary numbers:

(30) $K_{ij} = \left(A_{ij} - B_{ij}\right) / \langle A_{ij} - B_{ij} \rangle$.

All of these methods should be evaluated to achieve optimal computer processing.

CLUSTERING FOR LOCAL-SUMS

By clustering the image data into local-sums, the second stage of the verification algorithm is likely to yield results more sensitive to replication attempts. If i,j are the elements of a square matrix of N*N pixels, divided into a square matrix of M*M pixel clusters of size c*c, where M is the truncated integer resulting from dividing N/c, then the value of each normalized-difference local-sum coefficient D is

(31) $D_{k\ell} = \sum_{i,j} \left[\left(A_{ij} - B_{ij}\right)^2 / c^2\right] \qquad \left(0 \le D_{k\ell} \le 1\right),$

where i ranges from (k-1)c to kc, j ranges from (ℓ-1)c+1 to ℓc+1, and both k and ℓ range from 1 to M.

The normalized sum of the normalized-difference local-sum coefficient is therefore:

(32) $D_{A-B} = D_M = \left(\sum_{k,\ell} D_{k\ell}\right) / M^2 \qquad \left(0 \le D_M \le 1\right).$

Computationally, rather than use the binary values of differences between grey-scale values, it is more convenient to first quantize the grey-scale values and then subtract the binary values in the cluster. In that case, we derive a local-sum for the cluster $K_{k\ell}$:

(33) $K_{k\ell} = \Sigma_{ij} \left(I_{ij} - J_{ij}\right)^2,$

where I and J are defined as in Eqs. 28 and 29, and the normalized- total binary local-sum C becomes

(34) $C_{A-B} = C_M = \Sigma_{k\ell} (K_{k\ell}/c^2)/M^2.$

For subtraction and summing of binary values, the sum $K_{k\ell}$ in Eq. 33 has the same result as the absolute value of the unsquared differences, which computationally might be easier to program and execute.

Figure 4 shows some sample results from 3 by 3 clusters. Each score is derived from the absolute value after subtraction of binary values (0,1) assigned for cases A and B. The expected features of replicas would have relatively high scores (which could be normalized to 1 by if divided by 9). With a pair of images in good correlation, most of the difference clusters will score 0. Random noise can create a single point difference in A or B, resulting in scores of 1, and occasionally a cluster score of 2 if single-point noise is in both clusters. In effect, this will establish a threshold of low average scores that would be considered below the threshold of true-negative significance.

A true negative would normally have cluster scores of 3 or more. For example, any three points in B but not A would suggest a narrow crack-like feature. A narrow feature that is broadened might have a score of as much as 6, and a microbubble might cause a cluster score of up to 9.[2]

If higher weights are be given to line broadening and microbubbles that are likely to occur in replication attempts using casting techniques, the differences could be deliberately exaggerated, especially if the images are not too noisy. For example, the cluster difference correlation coefficient could be squared:

$$(33) \quad \rho_M = \left(\sum_{k,\ell} \rho_{k\ell}^2\right)/M^2 \qquad \left(0 \le \rho_M \le 1\right)$$

This would result in a greater numerical separation of values between noisy pixels and systematic deficiencies.

Appendix A provides supporting data for the clustering/local-sum method.

CALIBRATION ERRORS

The expressions derived above implicitly take into account calibration errors, whether random or systematic.

If the error is purely random, it can be considered explicitly by, for example, adding a component c_i to Eq. 8, and the equivalent to Eq. 9, such that σ_a^2 is really the sum of the variances of instrument and calibration noise terms.

On the other hand, if the calibration error might be systematic, as a result of constant bias in the measurement x_i or y_i or a relative error between the two, then it can be treated in the same way as Eq. 9, where a linear relationship is assumed.

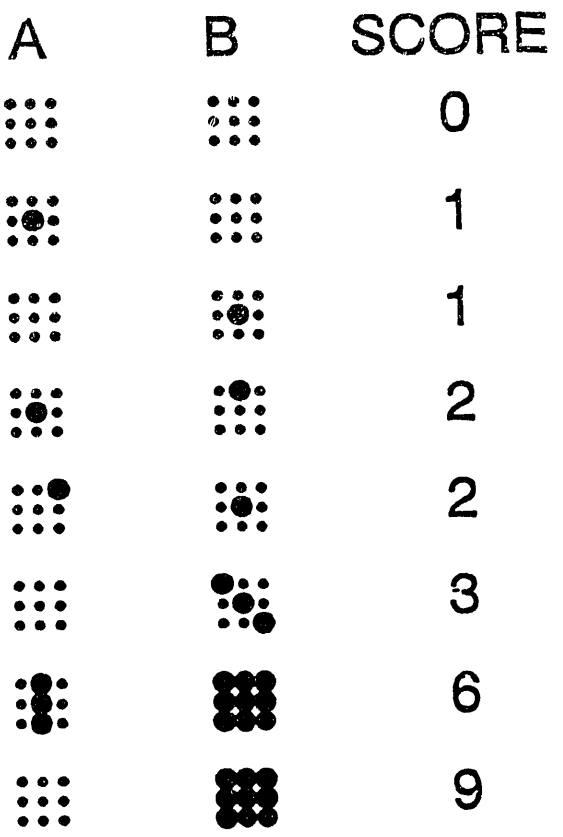

Figure 4. Cluster Scores for Difference Correlations. Each score is derived from the absolute value after subtraction of binary values (0,1) assigned for cases A and B.

It is this linear relationship which underlies the linear correlation coefficient, and the explicit inclusion of a calibration bias will not change the results. On the other hand, if the calibration bias (or the fundamental linear assumption of the primary data) is invalid, so is the linear correlation result, which should be derived for some other known or unknown relationship between the two blocks of data. Because of the unpredictability of such non-linear relationships, it is not particularly productive to go beyond the linear correlation coefficient, but potential non-linearities are a reminder of the limits of this commonly used measure.

SUMMARY

There are a number of statistical measures to help assess the validity of data. Autocorrelations, cross-correlations (covariances), and the normalized covariance, usually called the correlation coefficient, are minimum-variance unbiased statistics. Each of these can play a role in finding the optimum congruence (that is, in helping to correct for linear shifts, rotations, magnifications, and distortions) and in guiding conclusions after the optimum value is attained. Each has its own level of uncertainty that depends largely on the number of sample points (or pixels for images). In every case, the random "noise" or uncorrelated component must be taken into account as well as the systematic effects. The linear cross-correlation and difference-correlation coefficients can be computed with explicit attention to inherent noise.

The MSD is a computationally useful measure of congruence of two image matrices, mathematically being a mirror-image of the linear-correlation coefficient. A linear correlation coefficient that does not take into account sample noise might be useful for comparing results for data of consistent origin, but the linear coefficient should be used with caution in comparing data of different origin. Moreover, the linear cross-correlation coefficient does not take into account higher-order effects (such as image warpage and other non-linearities). A constructive operational approach would be to plot results from false positives and negatives along with true positives and negatives. There should be a comparison of at least 100 pixels when two images are to be accepted as true positive based solely on the linear-correlation coefficient.

One strategy designed to differentiate between similar (true positive) and dissimilar (true negative) images is a three-stage procedure consisting of separate registration and verification algorithms: first, a linear cross-correlation coefficient or a mean-square-deviation is used to optimize the congruence between two pixel image arrays; second, the highest value obtained for the linear cross-correlation coefficient is compared against an array value (≤ 0.3) that defines a true negative result; third, if the correlation coefficient averaged over the array exceeds the true negative threshold, the cluster (or local-sum) method is invoked as a further test to qualify a true positive.

REFERENCES

1. A. DeVolpi, "Estimate of Variance and Merit Ratios from Measured Quantities of Fluctuating Origin and Inherent Correlation," Int. J. Applied Radiation Isotopes 22, 103-110 (1971).

2. R. Palm and A. DeVolpi, "Intrinsic-Surface-Tag Image Authentication," Argonne National Laboratory Report ANL/ACTV-91/5 (December 1991).

APPENDIX A

DATA IN SUPPORT OF LOCAL-SUM CORRELATION

R.G. Palm

As described in the main body of the report, clustering data after carrying out some earlier image-data manipulations has the potential of yielding an objective measure of differentiation between two images, one of which differs in small but systematic ways from each other. The linear correlation coefficient (LCC) fails to provide objective discrimination scores in such cases. Although the method is illustrated in this Appendix with images produced by a scanning-electron microscope (SEM), the algorithm is applicable to any generic situation where two images need to be compared.

Both gray scale and binary images can be compared to determine tag authenticity. These tests for agreement are described in Section 1 of this appendix. Proposed acceptance criteria for sub-regions of images obtained for an SEM-authenticated tag are also presented in Section 1. The verification algorithm has been implemented using SEMPER 6.2 image-processing software. Congruent registration of the images is necessary before the tag images can be numerically compared; this process is described in Section 2.

1. Image Registration and Correlation of Gray-Scale Images

Before two digital images can be compared, a certain amount of processing is necessary to place them in congruent registration. Details of this registration process as implemented by the SEMPER software are described in a Section 2. In the case of two digital images A(x,y) and B(x,y), registration is achieved when the x,y addresses of each image correspond to the same point on the scene being compared. One measure of image agreement known as the linear correlation coefficient can be calculated from the registered images. An LCC of 1.0 indicates perfect agreement while an LCC near 0.0 would be expected for totally uncorrelated images.

Figures A1a and A1b compare two sets of gray-scale images. Figure A1a is a comparison of two casting images from the same original, and these two images have an LCC of 0.919. Figure A1b is a comparison of an original with an attempted replica, and these two images have an LCC of 0.875. Note that on Fig. A1b, the LCC falsely indicates a true positive, while the local sum correctly indicates a true negative. Visual inspection shows that the images in Fig. A1a agree better

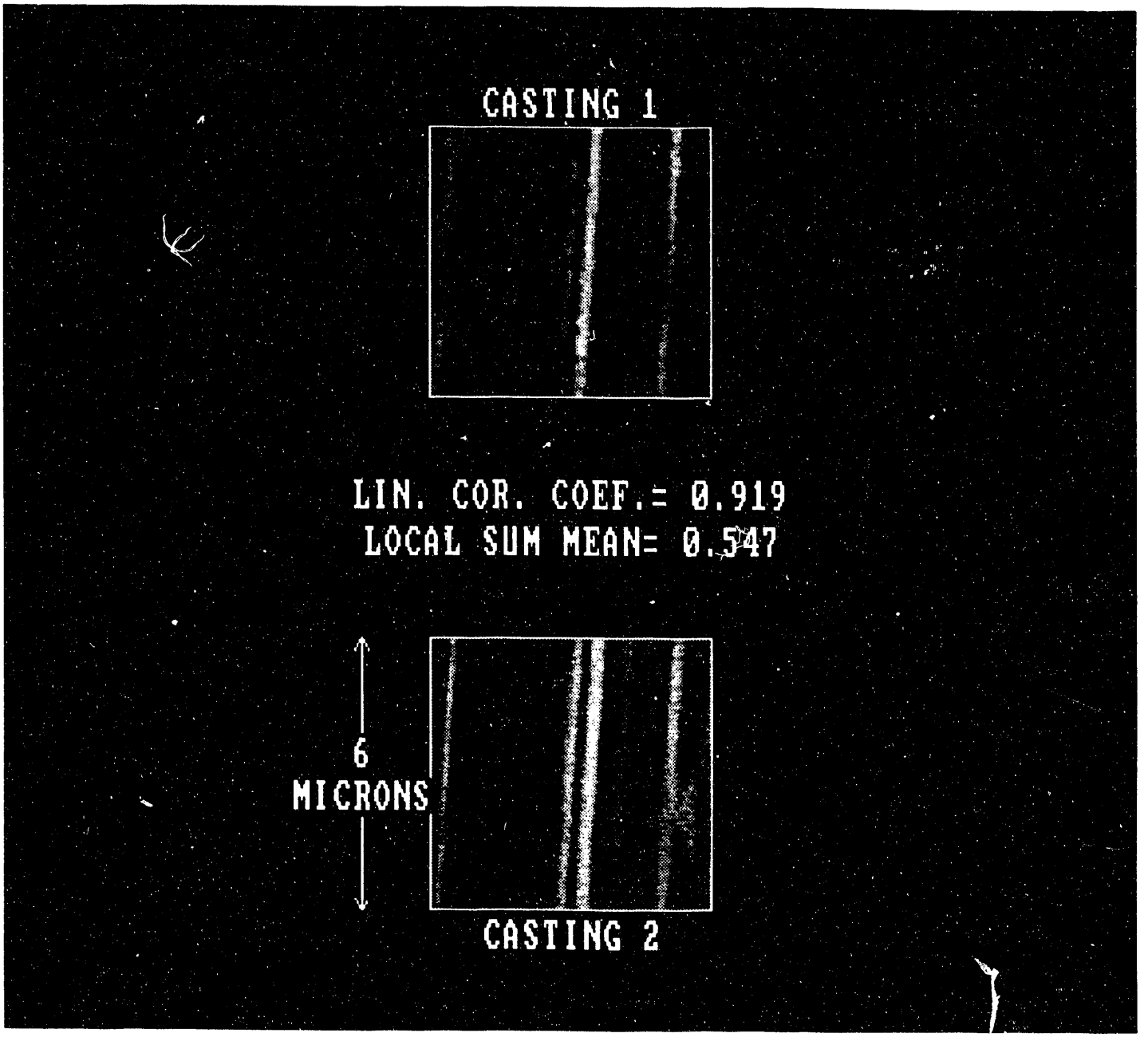

Figure A1a. Comparison of the Local Sum and Linear Correlation Coefficient for Two Castings from the Same Original Surface.

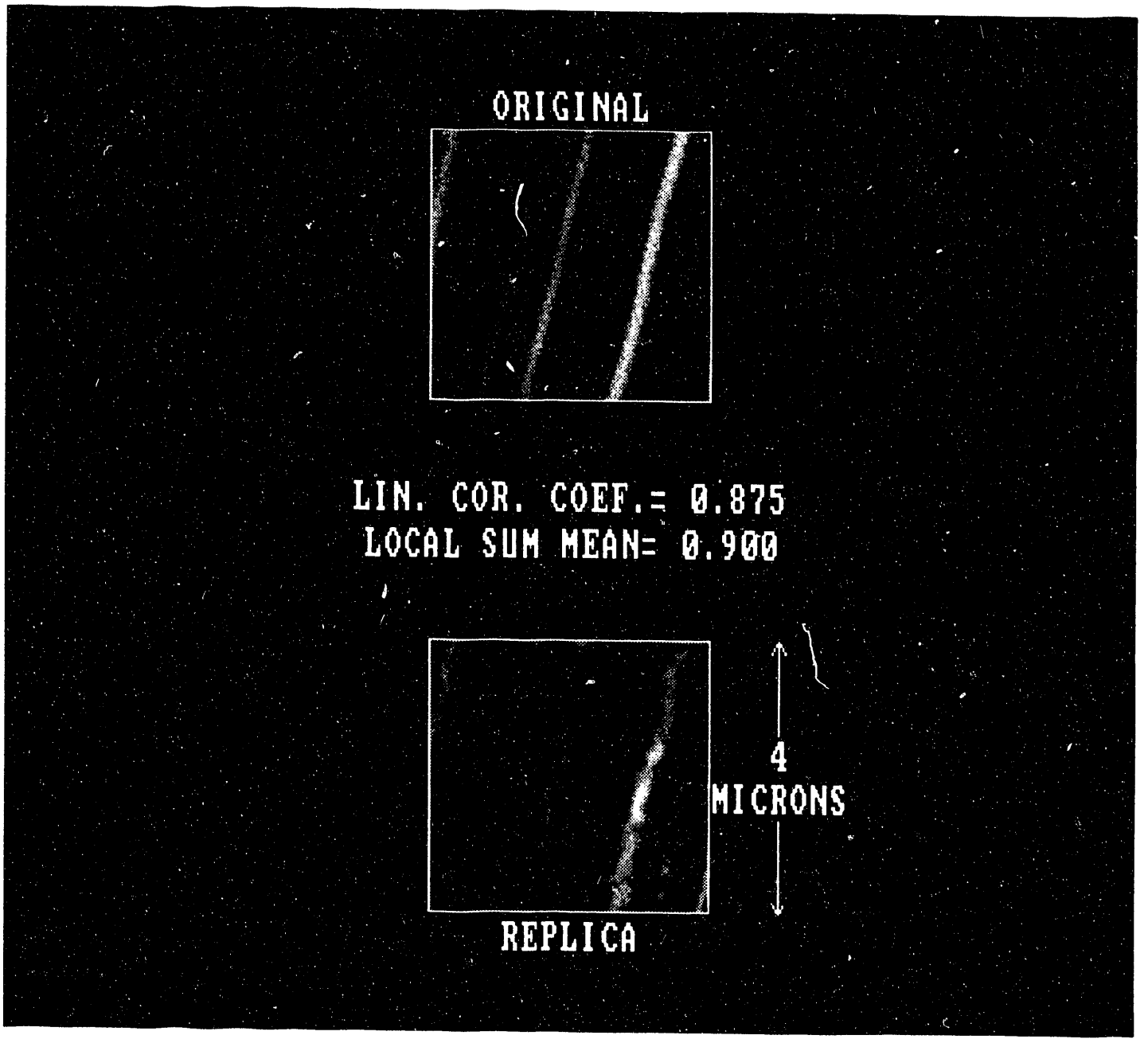

Figure A1b. Comparison of the Local Sum and Linear Correlation Coefficient for an Original and a Replica. Note that the LCC falsely indicates a true positive, while the local sum correctly indicates a true negative.

in the brighter areas that represent high areas in the surface. However, there is only a 5 difference in the LCC discrimination ability so the LCC provides a crude score with limited discrimination. Note that Figs. A1a and A1b also present another image-comparison statistic called the local sum. This statistic is described in a subsequent section.

Figure A2 presents two pairs of the same original and replica images with LCC = 0.832, highlighting disagreement among the brightest pixels. It is possible to extract sub-areas around the peaks (brightest pixels) to determine the correlations between areas. These sub-areas are shown in the lower part of this figure, taken from the areas marked in the top. Figure A2 shows that the sub-area isolated in the lower left of Figure A2 agrees to 0.828, still only 5% less than the LCC for the whole image comparison above. The boxed areas appear and are, in fact, noticeably different. The LCCs show no significant difference until the boxes are narrowed to a small area. In the lower right is a narrower sub-area comparison with a low or LCC of 0.414, coinciding better with visual observation of the disagreement. The LCC of the original and replica images are markedly smaller if one carefully chooses small sub-regions for correlation. However, choosing such small regions is difficult to implement.

From examining Figures A1 and A2 it is apparent that the LCC reported on the whole data set is a rather insensitive measure of image congruence. Nevertheless, the data suggests that a sensitive and computationally simple means to score an image based upon only its brightest pixels could be devised.

2. Binary-Image Comparison

It is possible to form binary images from the gray-level images thresholded according to some characteristic of the image, such as its brightness. In this case the binary images I(x,y) and J(x,y) derived, respectively, from the gray-scale images A(x,y) and B(x,y) are compared to determine a numerical score for tag comparison. Figure A3 shows binary images derived from an original and an attempted replication. The brightest 15% of the original and replica pixels are isolated by setting a threshold on the image intensity. Next, the gray-scale values are converted to binary values by assigning a null score to all values below the threshold and a unit value to those above. Then, the absolute differences are taken of the binary values, and a gray-scale image is reconstructed of the local-sum highlights. The binary images are set to pixel values equal to one if they are part of the brightest 15% set of pixels; otherwise they are zero. Inspection of Figure A3 shows that the ridge features of the replica are wavy and discontinuous compared to the original. The binary brightness-threshold process captures the differences between the original and a replica.

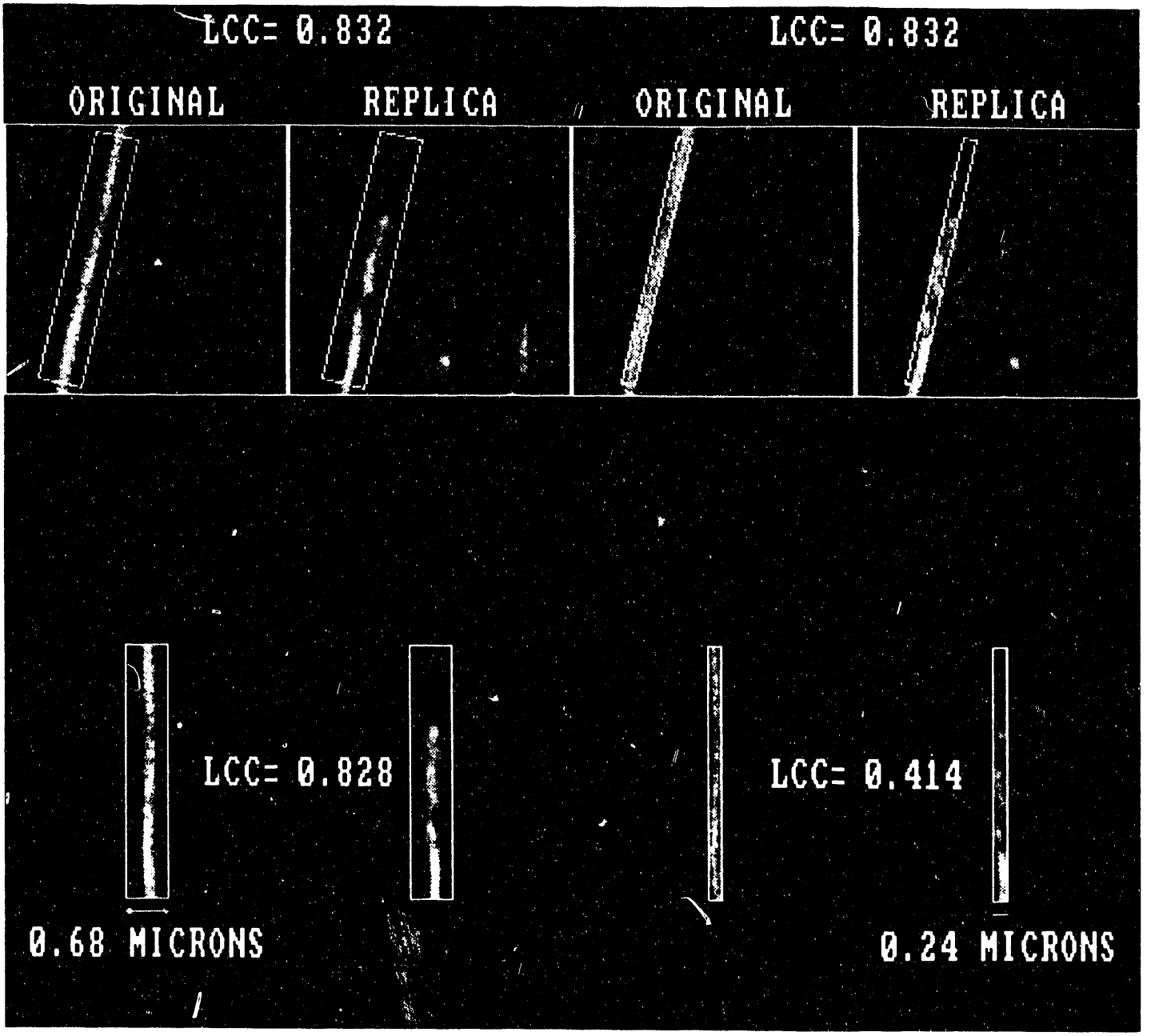

Figure A2. Isolation of High-intensity Peaks in and Original and Replica. The boxed areas appear and are, in fact, noticeably different. The LCCs show no significant difference until the boxes are narrowed to the small area.

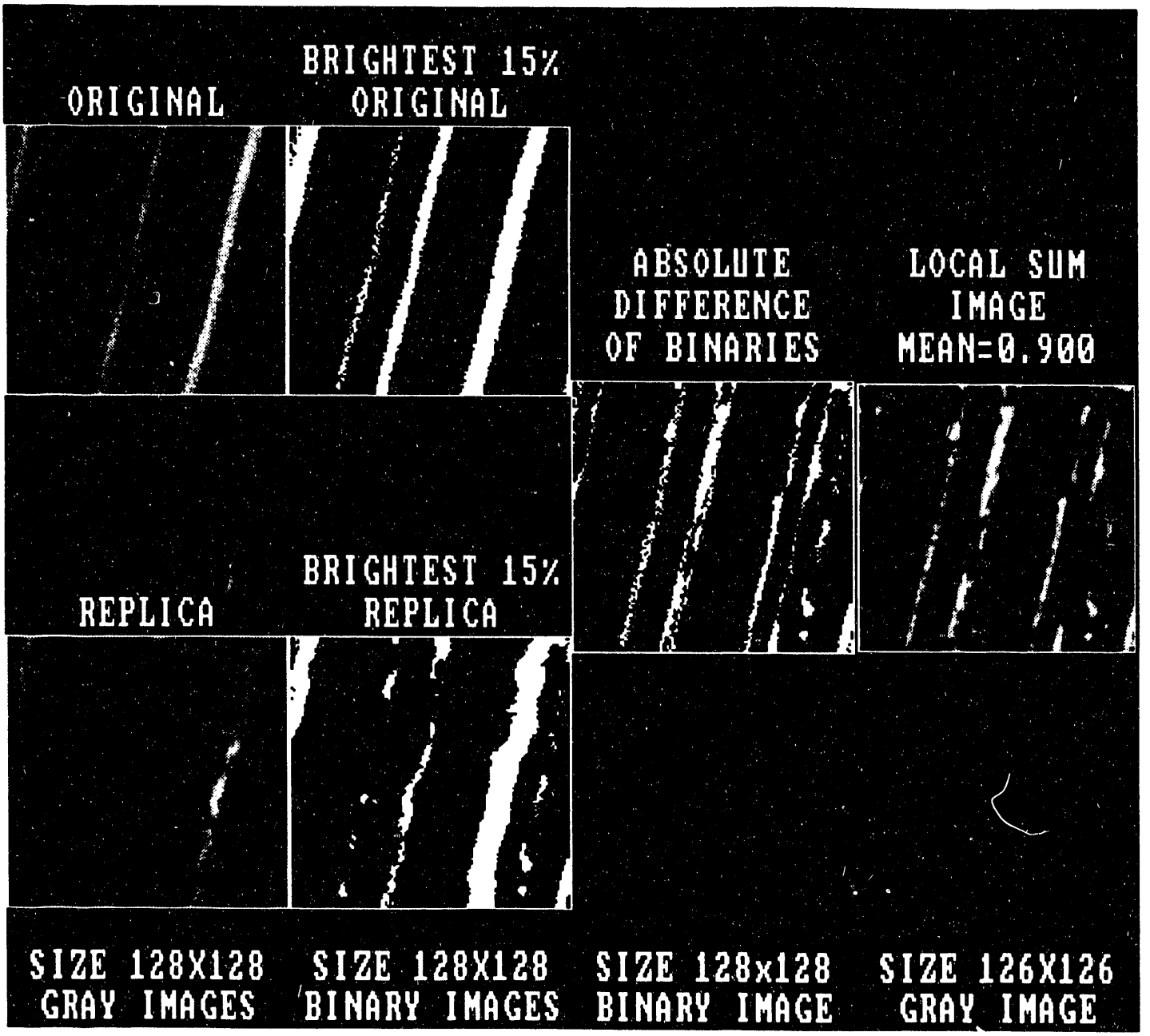

Figure A3. Steps Illustrating Calculation of Local-sum Image. The brightest 15% of the original and replica pixels are isolated by setting a threshold on the image intensity. Next, the gray-scale values are converted to binary values by assigning a null score to all values below the threshold and a unit value to those above. Then, the absolute differences are taken of the binary values, and a gray-scale image is reconstructed of the local-sum highlights.

2a. Absolute-Difference Image

The simplest way to compare the binary images is to form the absolute-difference image C from images I and J according to the following formula:

$$C = | I(x,y)-J(x,y) |$$

Image C gives a binary indication of the brightest pixels in I and J that disagree. The authentication score derived from C is simply its mean averaged over all x,y. Low mean values (i.e., small differences) indicate good agreement and vice versa. The mean of C can span from 0.0 (perfect agreement) to 1.0.

2b. Local-Sum Image

Another way to define the image-comparison score is to process image C further to define a local-sum image. The local-sum image gives a gray-scale rendition of the disagreement between images A and B. This local-sum image, referred to as image D, renders in terms of a gray scale the degree of disagreement of local p x p pixel clusters in C.

In order to implement a local-sum algorithm in SEMPER, the theoretical steps described in the main body of this report were modified. In particular, a moving-average cluster was created, in which the origin of the first cluster is displaced to an adjacent pixel which thereupon forms the centroid of the next cluster used to compute a local sum. The theoretical formulation would move the centroid from the first cluster to an entirely different adjacent cluster, the centroid of which might be several pixels away. The moving-average process smoothes out the appearance of the image, so that it is easier to visualize the local-sum differences. More testing needs to be done to see which method gives the most contrast between true positives and negatives.

The local-sum image D(x,y) is produced by summing each pixel in C(x,y) with its p x p nearest neighbors and placing the sum in position x,y of D. Because the pixels within a distance p-2 of the border of C(x,y) can't be summed, the local-sum image has dimensions M-p+1 by N-p+1 if the image C has dimensions M x N. The pixel values of the local-sum image can span the range from 0 (complete agreement) to p^2 (complete disagreement).

Figure A3 shows the local-sum image formed from comparison of an original and a replica. Disagreement amongst the brightest pixels is highlighted by a local-sum gray-scale image. For the 3 x 3 local-sum image shown in Figure A3, the local-sum pixel values range from 0 to 9. The score derived from local-sum image D is simply its mean value averaged overall (x,y). Low mean values indicate good agreement (small differences). The mean of D is about nine times the mean of C for 3 x 3 local-sum images.

3. Tentative Acceptance Criteria for Subareas

The adopted tentative criteria are to accept the images in a subareas if the LCC > 0.7 and the local-sum mean is < 0.6. When the images are overlaid, the LCC is determined before the local-sum mean. If LCC < 0.7, the subarea would fail to pass, and the local-sum need not be calculated. Figures A1 and A2 show that despite visual differences it is relatively easy to satisfy the LCC part of the acceptance criteria. However, meeting the local-sum criterion is much more demanding; only image pairs with their bright pixels in registration can satisfy it.

Two general kinds of images were compared to formulate these empirical acceptance criteria. For two castings taken from the same original surface, acceptance criteria were formulated that all of the casting image comparisons had passing scores. For originals and positive replicas of the original, the acceptance criteria were formulated so that all but one of these image comparisons at high magnifications had failing scores. At low magnifications the originals and replica attempts had passing scores. This is reasonable, as differences in the finer features that can distinguish the originals from the replicas were not resolved at the low magnifications.

Table A shows the results for several image comparisons reporting both LCC and mean values D of the two images based upon binary thresholds. All images in Table A were 128 x 128 pixel images. Note that Table A also gives the magnification of the images being compared. Six of seven original-versus-replica comparisons of images acquired at a magnification of 5500 fail to meet the LCC acceptance criteria. The left side of Figure A4 shows the one replica subarea that passed the acceptance criteria. However, note that the original and replica set on the right-hand side of Figure A4 did not have a passing score. The LCC in both cases incorrectly indicated true positive; the local-sum incorrectly indicated true positive in subarea 1, but correctly recognized the true negative in subarea 2. Both replica images in this figure were from the same replica, and the two subareas were located only 11 microns from each other. It is also important to note that each of the subareas in Figure A4 represent only 10^{-7} of a 1 cm^2 authentication surface. It is expected that the great majority of the subareas in each replica will fail to meet the local-sum acceptance criteria.

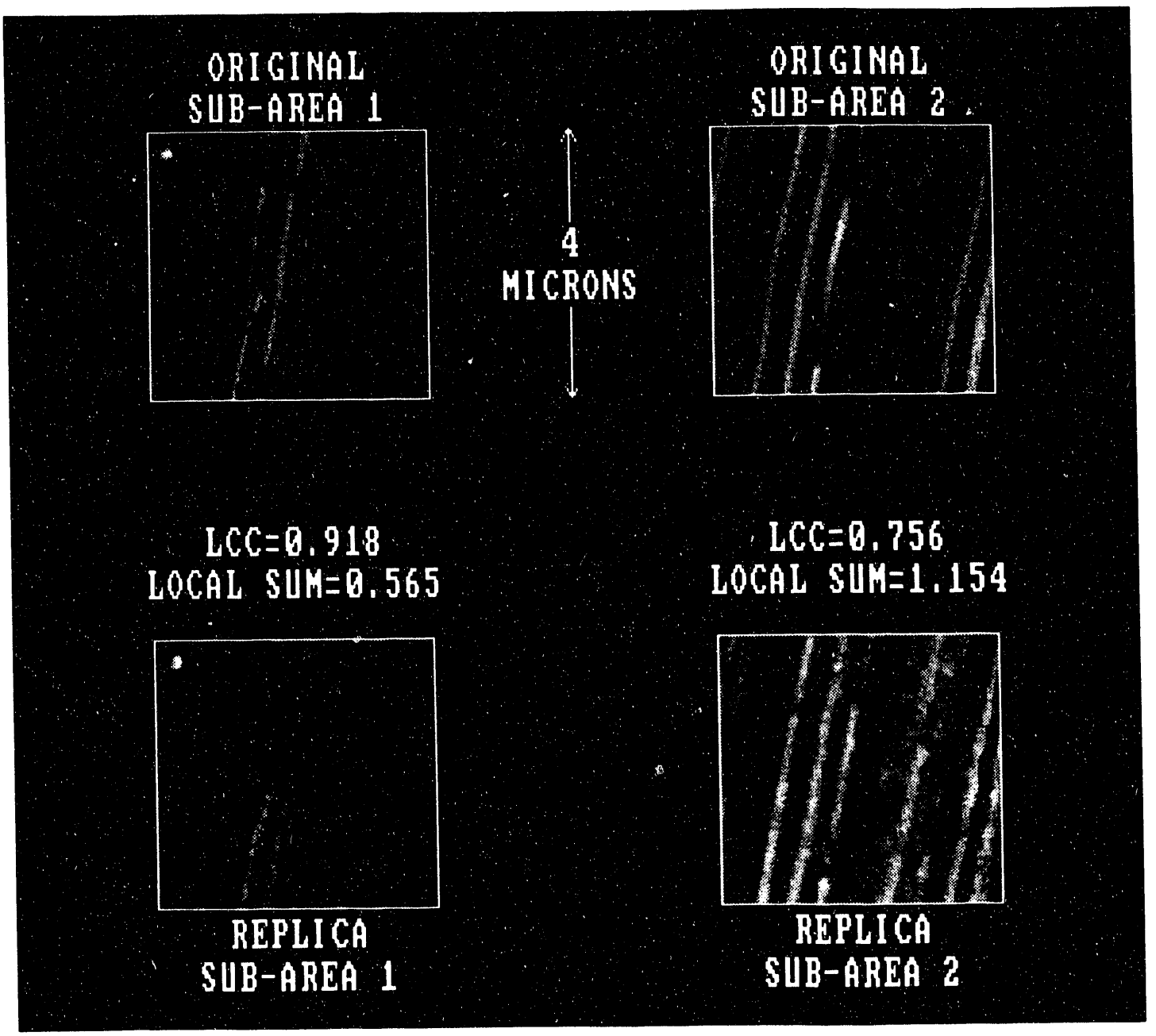

Figure A4. Comparison of Two Nearby Subareas from an Original and a Replica. The LCC in both cases incorrectly indicated true positive; the local-sum incorrectly indicated true positive in sub-area 1, but correctly recognized the true negative in sub-area 2.

One further acceptance criterion must be developed as more data is examined. This criterion would average the subarea pass/fails to provide a overall score. An example of an overall authentication criterion would be to accept the item as genuine if 90% of the subareas passed.

Figure A5 shows that three attempts to replicate the same surface failed the acceptance criteria. One of the three replication attempts passed the LCC criteria. While all three attempts to replicate the original were correctly identified by the local-sum test as true negatives, one of the LCCs gave a false-positive result, and the other two are borderline.

4. Processing Software

The gray-scale and binary-image processing steps have been implemented in the SEMPER 6.2 image-processing software. This software is a product of Synoptic's Ltd. Most of the computer effort is spent in making the image congruent, which is described in this section. As previously mentioned, the congruent LCC value is used as the first acceptance criterion. The binary-image processing is straightforward and its SEMPER implementation will not be discussed further.

All image matching must undergo a certain amount of processing before the two images can be compared mathematically. In general, two images of the same or similar scenes would, of course, be slightly dissimilar if the imaging device or the scene is changed between acquiring the images. These dissimilarities can be described either as translation, rotation, or magnification differences. In the case of digital images acquired by a scanning electron microscope, all the differences must be corrected before the images are congruent.

The initial goal of the image correlation is to adjust one or both of the images, so the images overlay or are congruent. In the case of two digital images A(x,y) and B(x,y), congruence is achieved when the x,y addresses of each image correspond to the same point on the scene being compared.

A program has been written that uses several SEMPER commands or routines to accomplish congruence using standard Fourier techniques, which are used because they are much more computationally efficient for the large images being compared. The most important correlation routine translates two images over each other and reports the x,y shift that provides the best overlay, which is determined when the linear correlation coefficient is maximized. Another SEMPER correlation routine determines the rotational correlation between two images; this routine

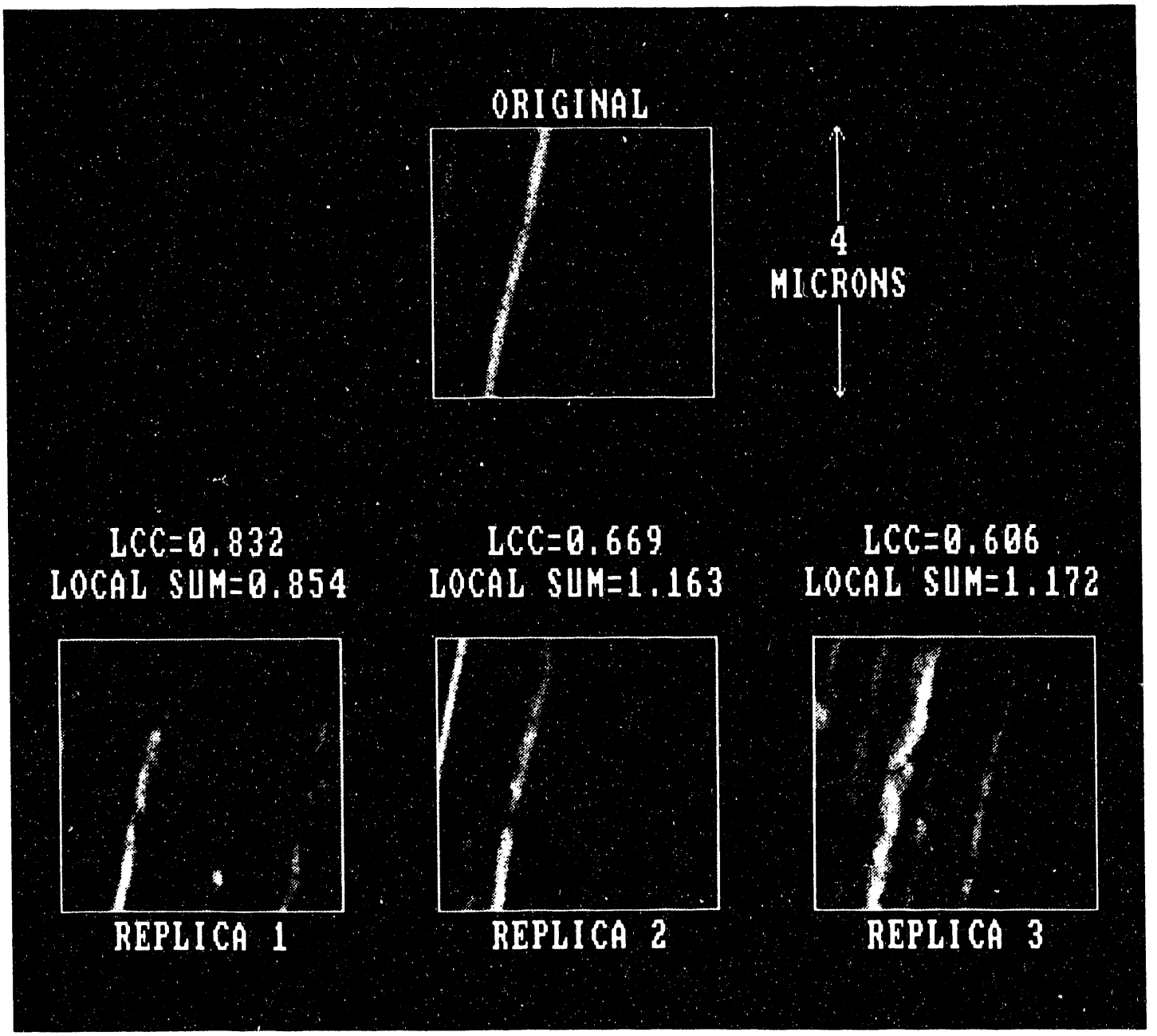

Figure A5. Comparison of an Original and Three Replicas. While all three attempts to replicate the original were correctly identified by the local-sum test as true negatives, one of the LCCs gave a false-positive result, and the other two are borderline.

as implemented in this program operates on the real images rather than their Fourier power spectra. The last important SEMPER routine extracts shifted, rotated, and translated sub-images from the as-acquired images.

As implemented in the program, three overlays with successively higher correlations are computed. The first overlay corrects for shift differences and is used as input into the second overlay. The second overlay corrects for magnification differences and is used as input into the third overlay. Finally, the third overlay corrects for rotational differences completing all adjustments.

The LCC's reported in Table A are for the third correlation. This correlation is computed to within 0.5 pixel and 0.1 degree of the theoretically best congruence; it is precise to at least the third decimal point. The casting images could be made congruent by a less resource-intensive process; however, false maximum correlations might then occur.

TABLE A

NUMERICAL IMAGE COMPARISON RESULTS

Image #1 File Number	Image #2 File Number	Magnification	Linear Correlation Coefficient	LCC Result	Mean Local-sum Image	LS Result
5242ca	5241ca	4000	0.907	TP	0.571	TP
5244ca	5243ca	4000	0.919	TP	0.547	TP
5225ca	5226ca	2000	0.869	TP	0.566	TP
5221ca	5222ca	2000	0.895	TP	0.451	TP
5223ca	5224ca	2000	0.757	TP	0.795	TN*
5227ca	5228ca	2000	0.915	TP	0.598	TP
5201ca	5202ca	700	0.922	TP	0.407	TP
5111or	5112rp	5500	0.875	FP	0.900	TN
5121or	5122rp	5500	0.881	FP	0.947	TN
5165or	5168rp	5500	0.756	FP	1.154	TN
5155or	5168rp	5500	0.918	FP	0.565	FP**
5101or	5102rp	5500	0.832	FP	0.854	TN
5101or	5103rp	5500	0.606	TN	1.172	TN
5101or	5104rp	5500	0.669	TN	1.163	TN
5193or	5194rp	700	0.898	FP	0.410	FP***
5193or	5195rp	700	0.898	FP	0.408	FP***

LS = local sum; LCC = linear-correlation coefficient ; ca= casting; or = original; rp = replica; T = true; F = false; P = positive (LCC > 0.7, LS < 0.6); N = negative (LCC < 0.7, LS > 0.6)

* Although this is expected to be TP, its value corresponds to FN; however, visual examination indicates that the second casting is flawed and so the result is actually TN.

** Although this sub-area should be TN, FP will occur in some sub-areas (see Ref. 2, pp. 26 - 29).

*** These sub-areas of the replicas visually at low magnification appear to be areas of good agreement, so that the low value of the local-sum is correct.

DISTRIBUTION FOR ANL/ACTV-91/4

Internal:

TIS Files
S. Bhattacharyya
F. Cafasso
R. Boyar
T. Braid
A. DeVolpi (28)
C. Dickerman
B. Dickerson (3 - CVR)
L. Gaines
C. Herzenberg
C. Johnson
D. Kupperman
R. Palm
S. Pratt
A. Raptis
E. Rhodes
N. Sather
S. Sheen
G. Stanford
R. Stajdohar
M. Steindler
A. Travelli
T. Wolsko
RE Files: 54920
RE Tagging File

External:

DOE/OSTI (2)
Manager, Chicago Operations, DOE
ANL-E, Libraries (2)
ANL-W, Library
C. Makris, DOE, Washington, DC (10)
P. Olinger, PSR
W. Nicholson, PNL

END

DATE FILMED

11/5/93