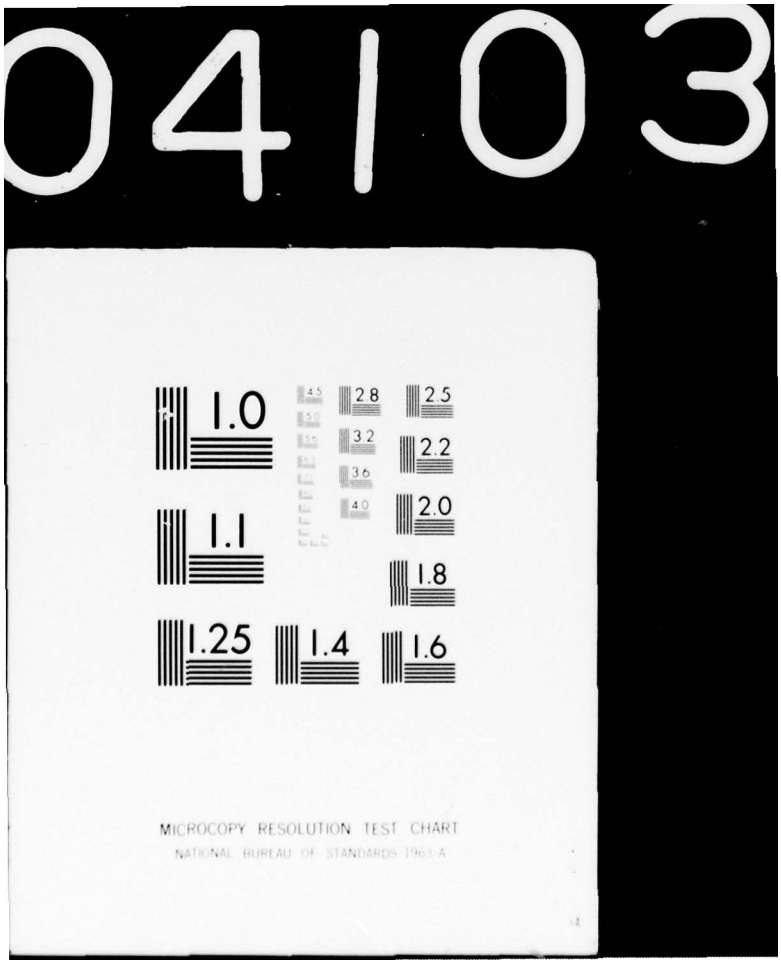
AD-A041 034 HONEYWELL INC MINNEAPOLIS MINN SYSTEMS AND RESEARCH --ETC F/G 17/5
AUTOMATED IMAGE ENHANCEMENT TECHNIQUES FOR SECOND GENERATION FL--ETC(U)
MAY 77 P M NARENDRA, L G WILLIAMS DAAG53-76-C-0195
UNCLASSIFIED 77SRC49 NL

1 OF 2
AD A041034


04103
1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
7.1
8.0
9.0
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

AUTOMATED IMAGE ENHANCEMENT TECHNIQUES FOR SECOND GENERATION FLIR

P. M. NARENDRA
L. G. WILLIAMS
W. G. CHAPLIN

SECOND INTERIM REPORT

CONTRACT NO. DAAG53-76-C-0195

DECEMBER 1, 1976 TO APRIL 15, 1977

Prepared for

U.S. ARMY ELECTRONICS COMMAND
AMSEL-NV-VI
NIGHT VISION LABORTORY
Fort Belvoir, Virginia 22060

Honeywell

SYSTEMS & RESEARCH CENTER

2600 RIDGWAY PARKWAY MINNEAPOLIS, MINNESOTA 55413

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (WHEN DATA ENTERED)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOV'T ACCESSION NUMBER	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (AND SUBTITLE) AUTOMATED IMAGE ENHANCEMENT TECHNIQUES FOR SECOND GENERATION FLIR		5. TYPE OF REPORT/PERIOD COVERED Second Interim Report. no. 2, December 1, 1976 to April 15, 1977 1 Dec 76 - 15 Apr 77
		6. PERFORMING ORG. REPORT NUMBER 77SRC49
7. AUTHOR(S) P. M. Narendra, L. G. Williams W. G. Chaplin		8. CONTRACT OR GRANT NUMBER(S) DAAG53-76-C-0195
9. PERFORMING ORGANIZATIONS NAME/ADDRESS Honeywell Systems and Research Center 2600 Ridgway Parkway Minneapolis, Minnesota 55413		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 15 May 77
11. CONTROLLING OFFICE NAME/ADDRESS U.S. Army Electronics Command, AMSEL-NV-VI Night Vision Laboratory Fort Belvoir, Virginia 22060		12. REPORT DATE May 15, 1977 13. NUMBER OF PAGES 129
14. MONITORING AGENCY NAME/ADDRESS (IF DIFFERENT FROM CONT. OFF.) 129 p.		15. SECURITY CLASSIFICATION (OF THIS REPORT) Unclassified 15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (OF THIS REPORT)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (OF THE ABSTRACT ENTERED IN BLOCK 20, IF DIFFERENT FROM REPORT)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (CONTINUE ON REVERSE SIDE IF NECESSARY AND IDENTIFY BY BLOCK NUMBER)

20. ABSTRACT (CONTINUE ON REVERSE SIDE IF NECESSARY AND IDENTIFY BY BLOCK NUMBER)

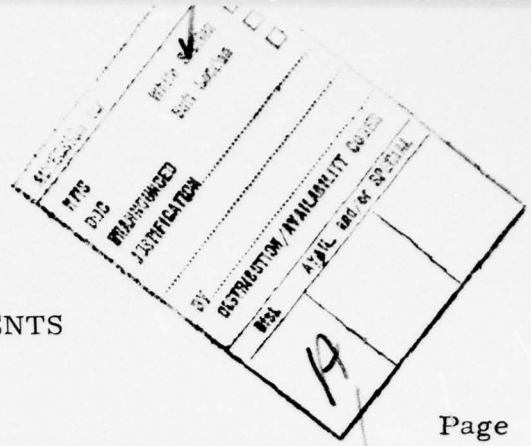
This is the second interim progress report for contract number DAAG53-76-C-0195, Automated Image Enhancement Techniques for Second Generation FLIR. It describes results of Phase I studies of image processing applications to FLIR imaging systems from December 1, 1976 to April 15, 1977.

HD-168 REV 11/74

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 55 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (WHEN DATA ENTERED)



CONTENTS

CONTENTS (continued)

CONTENTS (concluded)

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (continued)

LIST OF ILLUSTRATIONS (continued)

LIST OF ILLUSTRATIONS (continued)

LIST OF ILLUSTRATIONS (concluded)

LIST OF TABLES

SECTION I

INTRODUCTION

This is the second interim progress report for Contract DAAG53-76-C-0195, Automatic Image Enhancement Techniques for Second Generation FLIR. It describes the results of Phase I studies of image processing applications to FLIR imaging systems from December 1, 1976 to April 15, 1977, performed by Honeywell Systems and Research Center.

This report is organized along the following major headings corresponding to Phase I tasks:

- Contrast Enhancement
- MRT Enhancement
- Resolution Restoration
- Statistical Analysis
- Integrated Approach to Image Enhancement (T. S. Huang)
- Search Effectiveness Model Development

Contrast enhancement deals with the problems of transforming thermal images of wide dynamic range into the limited dynamic range of the display medium without losing essential information content. MRT enhancement involves spatial or temporal smoothing for noise reduction. Resolution restoration treats detail recovery from imagery which has been blurred by the optics and electronics of the sensor. Statistical analysis of FLIR

imagery is performed to provide quantifiable characterizations through shape, brightness, and textural features statistics. An integrated approach to contrast, MRT, and resolution enhancement resulted from work performed by Professor T. S. Huang of Purdue University under contract to Honeywell Systems and Research Center.

The last section reports the progress on the Search Effectiveness Model add-on to the present contract to develop a model and procedure for predicting search performance using electro-optical sensors in tactical situations.

Accompanying this report is a companion volume which describes work performed by Honeywell Radiation Center on Autofocus, detector responsivity equalization, and other relevant aspects of CCD technology.

SECTION II

CONTRAST ENHANCEMENT

INTRODUCTION

This section describes the progress made on the Phase I contrast enhancement tasks in this reporting period. All algorithm development, coding, and initial parameter selection were completed. Production of the enhanced images for evaluation is well on the way. The images will be evaluated by using the imagery statistics specifically developed to characterize FLIR imagery, and by the human factors enhancement evaluation add-on part of this contract. In this section we briefly review the contrast enhancement objectives, outline the approaches taken, and discuss their implementability and effectiveness with the aid of representative contrast enhanced FLIR imagery.

The need for contrast enhancement arises because, when the high dynamic range scene temperature information (available at the focal plane) is squeezed into the limited luminance range of a typical display, low contrast local details (targets, etc.) tend to fall below the contrast sensitivity threshhold of the human eye and are not perceived. It is of course possible to adjust the gain and brightness controls on a FLIR imager to selectively expand a given temperature range of interest, but this involves extensive interactive manipulation of the controls and is therefore not desirable. Contrast enhancement algorithms are being investigated to enhance the

local detail and compress the overall scene dynamic range of the FLIR. The following techniques were investigated:

- High frequency emphasis linear and homomorphic filtering
- Local area gain/brightness control (LAGBC)
- Histogram modification techniques

The linear high frequency emphasis filter and the LAGBC schemes were found to be very effective for contrast enhancement. Moreover, simple recursive filter formulations were developed for both high frequency emphasis filtering and LAGBC. The recursive filter approach is not only effective but could result in a CCD analog implementation an order of magnitude simpler than an equivalent nonrecursive approach. On the other hand, histogram modification techniques (both full frame and local area) were found (see the first quarterly report) to be not as consistently useful for enhancing FLIR imagery.

HIGH FREQUENCY EMPHASIS FILTERING

This is based on the premise that FLIR scenes can be modeled as a large slowly varying temperature envelope corresponding to the background and superimposed higher spatial frequency information corresponding to targets, background, detail, etc. Attenuating the low frequency components compresses the low frequency envelope and therefore compresses the dynamic range. At the same time, high frequency emphasis enhances the local detail above the contrast sensitivity threshold. The filter frequency response required to achieve this is shown in Figure 1a. Figure 1b shows how this can be done using a low pass filter as a basic building block. The

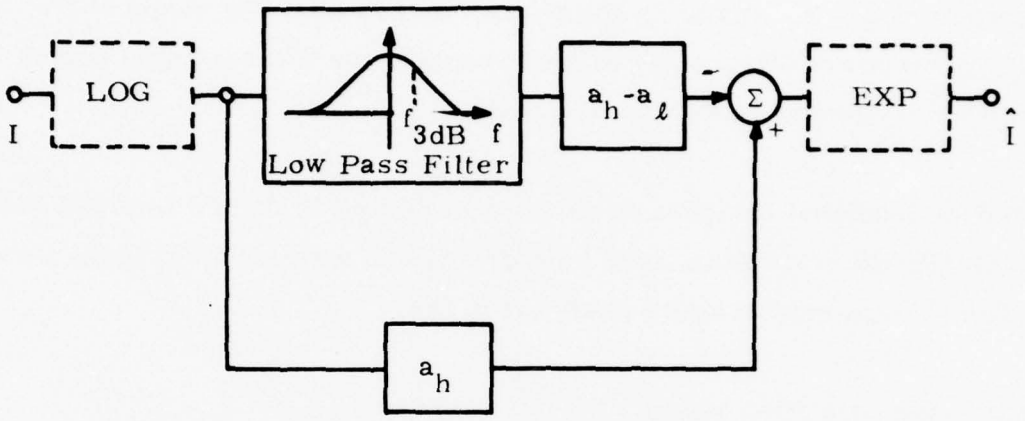


Figure 1a. Desired High Frequency Emphasis Response

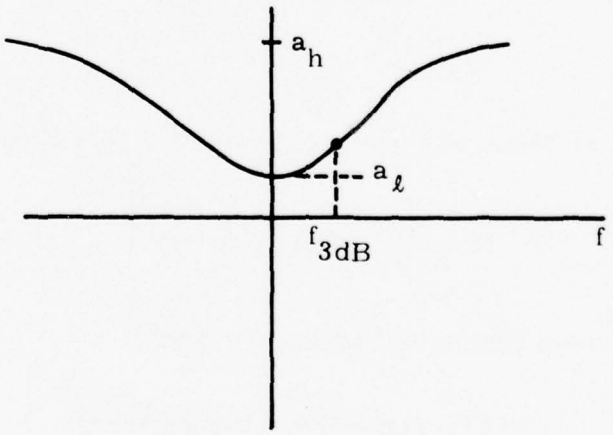


Figure 1b. High Emphasis Filter Using a Basic Low Pass Filter

reasons for realizing a low pass filter first are that (recursive) low pass filters are more stable and easier to design. Moreover, by simply changing the gains a_L and a_h in Figure 1b, we can alter the shape of the frequency response (Figure 1a) without changing the filter weights (which are harder to change) on the low pass filter.

The two-dimensional low pass filter used in Figure 1b can be realized with two basically different structures--recursive and nonrecursive realizations. Each has its own merits as we shall see below.

Nonrecursive Low Pass Filter

In a nonrecursive filter, the filtered output is a convolution of a finite sized impulse response of size M with the input image (N x N) as follows:

$$y(m,n) = \sum_{k=0}^{M-1} \sum_{\ell=0}^{M-1} x\,(m-k, n-\ell)h(k,\ell),$$

$$m, n = 0, \ldots, N-1$$

The low pass filter frequency response selected was Gaussian,

$$H(f_x, f_y) = \text{Exp}\left[-\frac{f_x^2 + f_y^2}{2\sigma^2}\right]$$

which in turn gives a Gaussian impulse response

$$h(x,y) = \frac{1}{\sqrt{2\pi}\sigma'} \text{Exp}\left[-\frac{x^2 + y^2}{2\sigma'^2}\right]$$

where

$$\sigma' = \frac{1}{\sqrt{2\pi}\sigma}$$

The Gaussian filter is separable, i.e.,

$$h(x,y) = h_x(x) \cdot h_y(y)$$

Therefore the two-dimensional convolution can be cast as two successive one-dimensional convolutions:

$$y(m,n) = \sum_k h_x(k) \sum_\ell h_y(\ell)\, x\,(m-k,\ n-\ell)$$

i.e., one-dimensional filtering of the rows with $h_y(\ell)$ followed by one-dimensional filtering of the columns with $h_x(k)$. Note that straight two-dimensional convolution requires N^2M^2 multiplications and additions (MADs) whereas the above separability property reduces this to $2MN^2$ MADs, a saving by a factor of M/2. Since M required for contrast enhancement is typically 11 to 15, this was found to be a considerable saving in the computer simulation of this filter.

The filter size M was related to the 3 dB cutoff frequency by specifying that the truncated filter coefficient at the tail of the Gaussian (edge of the window) should be 1 percent of the center value. With this,

$$f_c = \frac{0.2843}{M} \cdot f_s$$

where f_s is the sampling frequency. This helps us to choose equivalent parameters for other filter shapes (not necessarily Gaussian) and for sampling rates which change in different FLIR configurations.

Recursive Low Pass Filter

Two-dimensional recursive low pass filters are investigated in parallel with the nonrecursive realizations because the recursive filters possess considerably simpler implementation. The two-dimensional recursive low pass filters were designed as a separable product of two one-dimensional Butterworth filters.

For an n^{th} order Butterworth filter, the frequency response is given by

$$|H(f)| = [1 + (f/f_c)^{2n}]^{-\frac{1}{2}}$$

Figure 2 shows the frequency responses of the first- and second-order Butterworth filters along with a Gaussian low pass filter with the same 3 dB cutoff frequency. We see that the first-order filter is similar in shape to the Gaussian. Further, the cutoff transition is gradual. This is in fact desirable because we do not want to create artifacts by selectively emphasizing some spatial frequencies much more than their neighboring frequencies. Further, the impulse response of the first-order filter is an exponential; this implies that there is no "ringing," a phenomenon associated with higher order filters possessing oscillating impulse responses.

The corresponding two-dimensional separable first-order recursive filter has a frequency response given by the product of the two-dimensional filter responses:

$$|H(f_x, f_y)| = [1 + (f_x/f_c)^2]^{-\frac{1}{2}} [1 + (f_y/f_c)^2]^{-\frac{1}{2}}$$

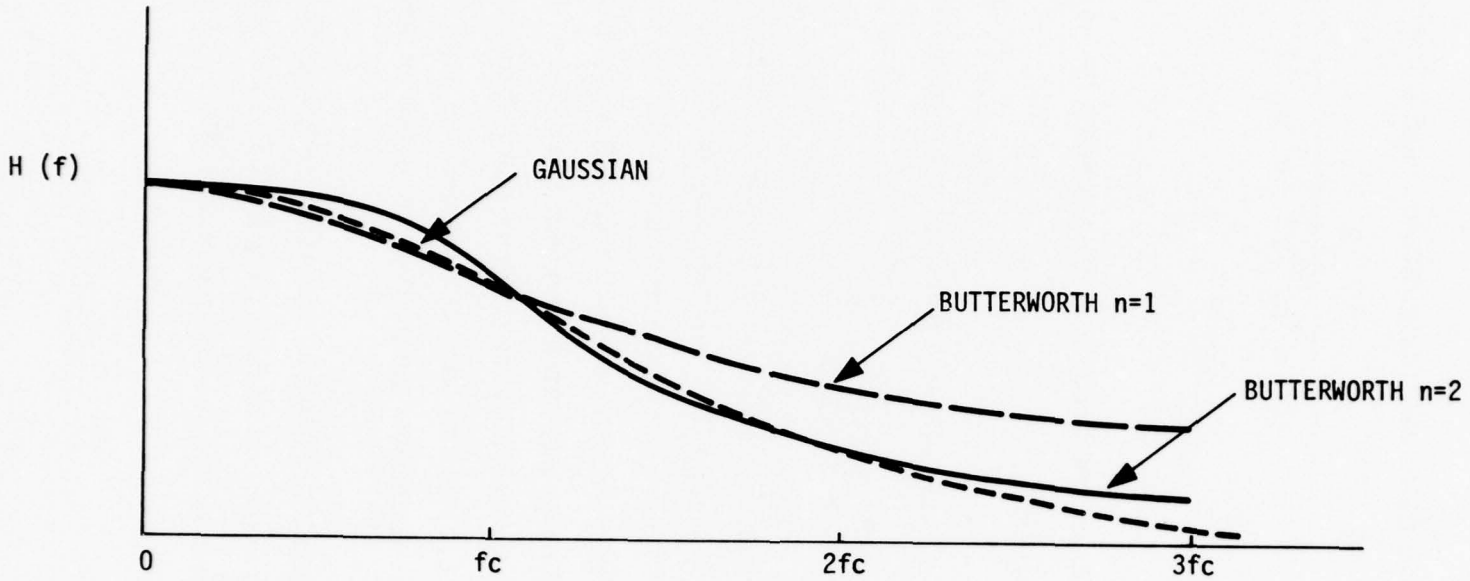


Figure 2. Frequency Responses of Gaussian, First- and Second-Order Butterworth Filters

The equivalent sampled data filter has a two-dimensional Z transform,

$$H(z_1, z_2) = \frac{\gamma^2}{(1 - e^{-\gamma} z_1^{-1})(1 - e^{-\gamma} z_2^{-1})}$$

where

$$\gamma = 2\pi f_c / f_s$$

is a function of the 3 dB cutoff frequency f_c. This gives a two-dimensional recursive relation for the filter

$$y(m,n) = \gamma^2 x(m,n) + e^{-\gamma} y(m-1,n) + e^{-\gamma} y(m,n-1)$$
$$- e^{2\gamma} y(m-1,n-1)$$

Figure 3 shows the filter structure for the first-order recursive filter. We see that the current filtered output is a function only of the previous filtered pel, the previous filtered line, and the current input pel. So we need just one line delay and four filter weights to realize this filter. By changing the filter parameter γ, practically any cutoff frequency can be realized. Incidentally, this two-dimensional recursive filter is stable because of the separability property, i.e., the two one-dimensional low pass filters are stable for all positive γ. Therefore the product is stable as well.

Several NVL supplied thermal (FLIR and thermoscope) images have been enhanced by the two high emphasis filters. Figure 4a is the original thermal image, and Figures 4b and 4c are the high emphasis filtered versions using the nonrecursive and recursive filters, respectively.

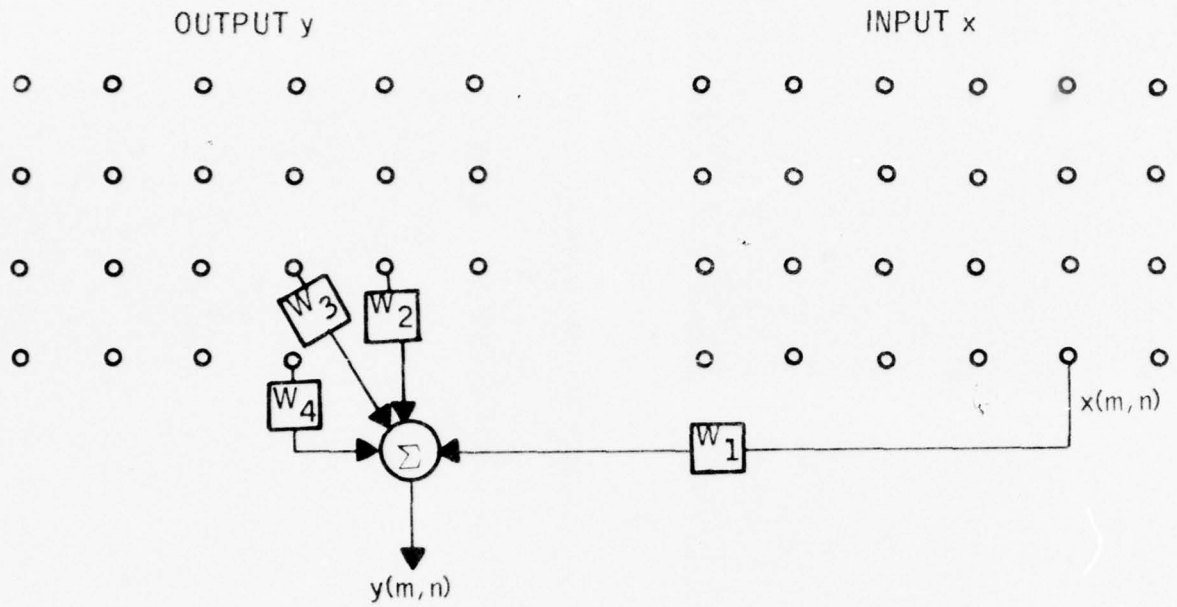


Figure 3. First-Order Recursive Filter Structure

Figure 4a. Original Thermal Image

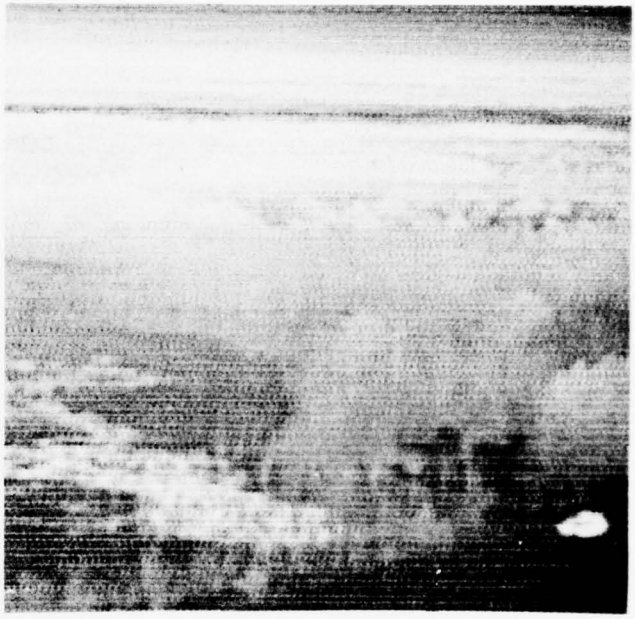


Figure 4b. High Frequency Emphasis Nonrecursive Filter

Figure 4c. High Frequency Emphasis Recursive Filter

For the nonrecursive filter, the size was M = 11; for the recursive filter, the cutoff frequency was $f_c = 1/20\ f_s$. For both filters, the low frequency gain was $a_\ell = 0.5$ and the high frequency gain was $a_h = 2.5$. Another set of images--original and similarly enhanced by high emphasis filtering-appear in Figures 5a, 5b, and 5c. We note that the overall dynamic range has been reduced in both high frequency emphasized versions, while the local contrast (especially the target detail) has been enhanced. The recursive filter did as well as the nonrecursive filter. It is instructive to examine histograms of the original and the high frequency emphasized images in Figures 4a and 4b (in Figures 6a and 6b). The scene dynamic range has been compressed, while significant target detail has been enhanced in Figure 4b. It should be pointed out that the parameters used in Figures 4 and 5 may not be optimal. The optimal parameter selection is being done as part of the production runs to produce 150 contrast-enhanced thermal imagery for evaluation. The parameters for high frequency emphasis filtering are M, the size of the nonrecursive filter(related to the cutoff frequency f_c), the cutoff frequency of the recursive filter f_c, and the low and high frequency gains a_ℓ and a_h.

There is no objective criterion for selection of these parameters, but some heuristic guidelines exist. For example, f_c should correspond to target sizes of interest (10 to 20 pixels in diameter, for example) and the ratio of gains a_h/a_ℓ should be chosen so that all the low frequency information is not lost. A systematic parameter selection process is underway to enhance representative images with a sequence of parameter values and to select the optimal parameter set. This is outlined at the end of this section.

Figure 5a. Original Thermal Image

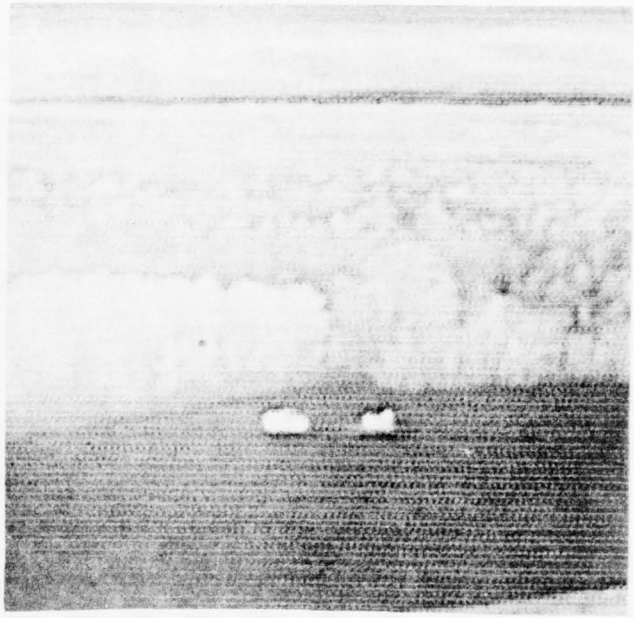


Figure 5b. High Frequency Emphasis Filter

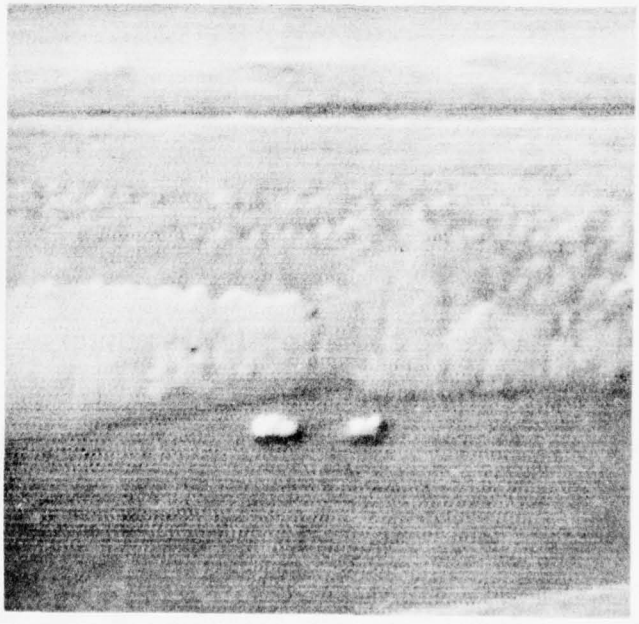


Figure 5c. High Frequency Emphasis Recursive Filter

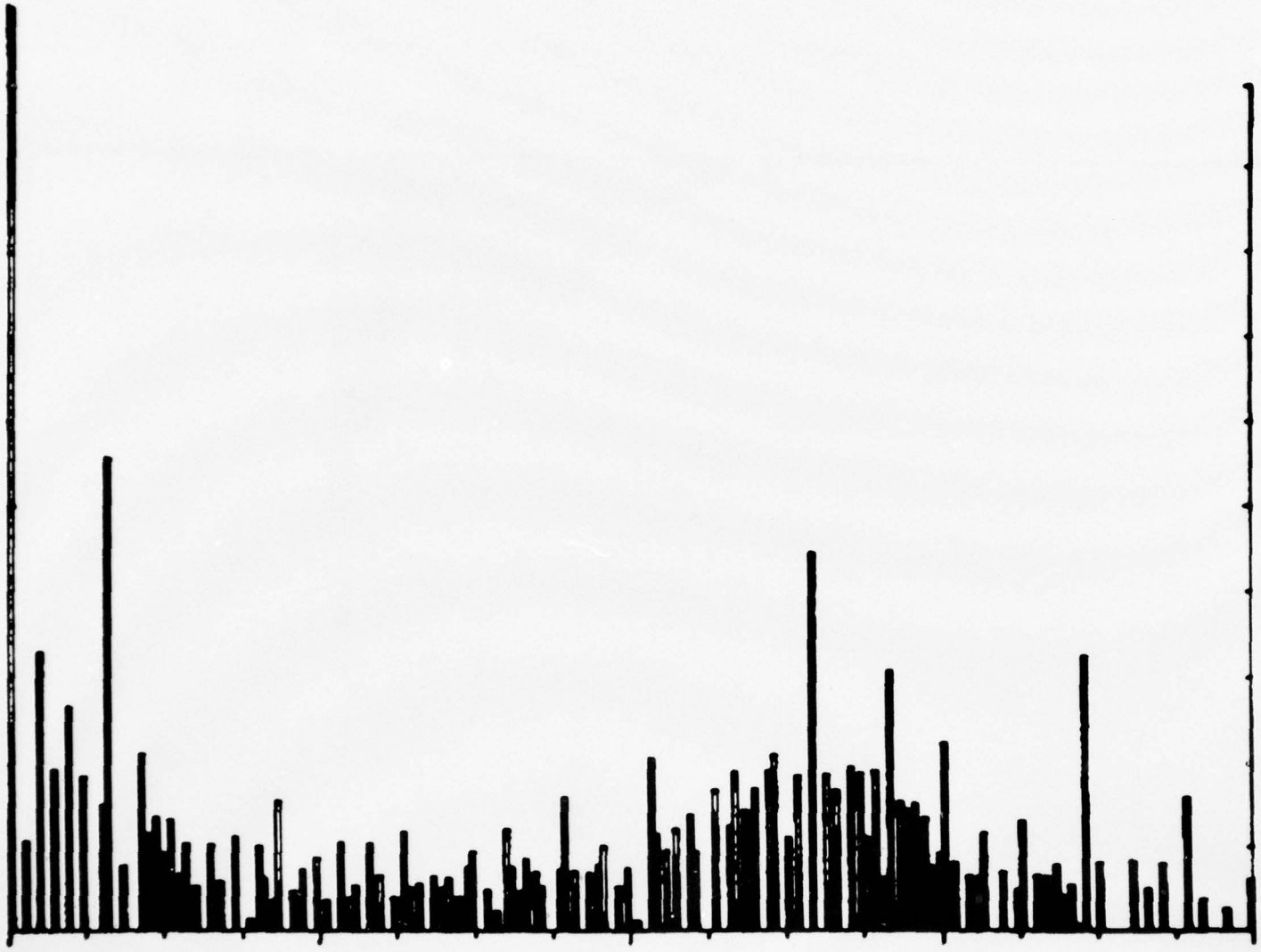


Figure 6a. Histogram of Original Thermal Image in Figure 4a

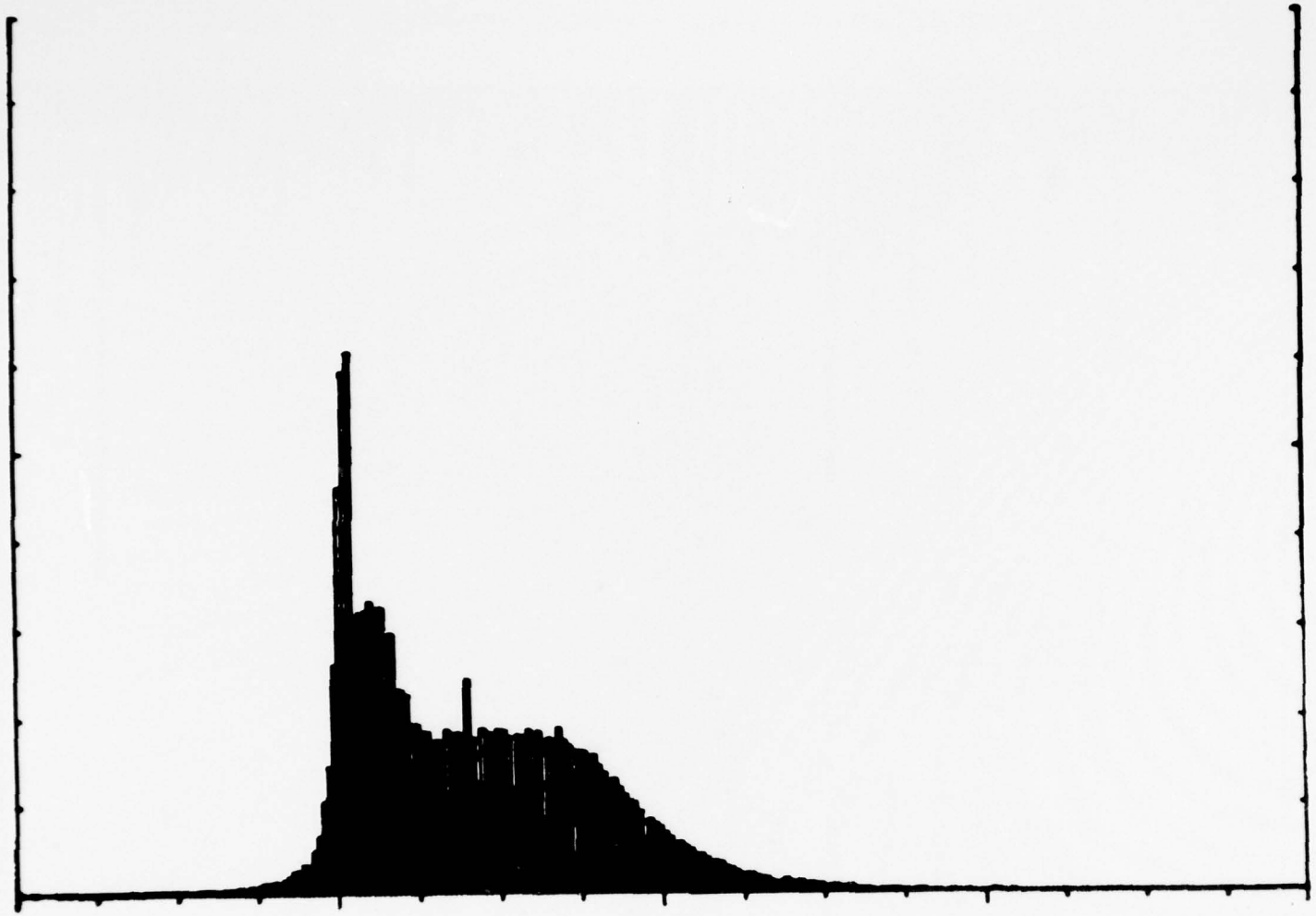


Figure 6b. Histogram of High Frequency Emphasis Image in Figure 4c

Comparison of Recursive and Nonrecursive Approaches

For purposes of comparison, we assume that M = 11 for the nonrecursive Gaussian filter and that the recursive filter is first order. This nonrecursive filter size (M = 11) was found to give the best results with the images enhanced at this time. In Figures 4 and 5 we saw that the recursive filter and the nonrecursive filter yielded approximately equivalent results. But there are some important differences in the two structures that will affect their real-time implementation potential:

- The nonrecursive filter needs access to M (11) lines at the same time, which implies M-1 line delays for a real-time implementation on a serial data stream. The first-order recursive filter needs just the previous filtered line, i.e., one (1H) line delay.
- The nonrecursive filter requires M^2 (121) filter weights in the nonseparable case and 2M (22) filter weights in the separable case, whereas the recursive filter requires only four weights.
- To change the cutoff frequency of the nonrecursive filter, we need to change the filter size, i.e., the entire filter structure. Any cutoff frequency can be realized in the recursive filter by simply changing the four filter weights without altering the filter structure.
- The nonrecursive Gaussian filter is isotropic (possesses circular symmetry and has zero phase). The same is not true of the recursive filter. But the nonlinear phase effect does not appear to be very evident in the enhanced images.

A detailed implementation study for real-time application of these approaches will be made during Phase II of the current contract. At this time, the recursive filter with its simple structure seems to offer great promise for analog CCD implementation.

LOCAL AREA GAIN/BRIGHTNESS CONTROL

Local area gain/brightness control schemes have been proposed to locally control the gain and bias of the video so that the full luminance range of the display is used to display local information. Ideally such a scheme should

- Vary the local average brightness (bias) so that overall dynamic range of scene is compressed,
- Enhance local variations above the contrast sensitivity threshhold of the human eye, and
- Automatically fit the intensity extremes in the enhanced video scene to the display limits.

The image intensity at each point is transformed based on local area statistics (the local mean M_{ij} and the local standard deviation σ_{ij} computed on a local area surrounding the point). The transformed intensity is then

$$\hat{I}_{ij} = G_{ij}[I_{ij} - M_{ij}] + M_{ij}$$

where the local gain

$$G_{ij} = \alpha \frac{\sigma_{ij}}{M_{ij}} \quad , \; 0 < \alpha < 1$$

The locally transformed intensities $\hat{I}_{ij}$ are further sealed to the display by

$$\hat{\hat{I}}_{ij} = A_G \cdot \hat{I}_{ij} + B_G$$

where A_G and B_G are global gain and bias computed from the minimum and maximum values of $\hat{I}_{ij}$ from the previous frame.

The LAGBC equation merely amplifies the local intensity variation around the local mean M_{ij}. The local gain G_{ij} is itself locally adaptive, being proportional to M_{ij}, to satisfy the psychovisual considerations and is inversely proportional to σ_{ij} so that areas with small local variance receive larger gain. It is instructive to analyze the effect of G_{ij} on the new local standard deviation:

$$\sigma'_{ij} = G_{ij}\sigma_{ij} = \alpha M_{ij}$$

This requires that the new local variance be directly proportional to the local mean everywhere in the image. This satisfies the contrast sensitivity criterion. To prevent the gain from being inordinately large in areas with large mean and small standard deviation, the local gain is actually controlled as in Figure 7. LAGBC was implemented using two essentially different approaches to the computation of the local area statistics: mean and variance.

Interpolation of Local Area Statistics

The primary computation involved in the simulation of LAGBC is in evaluation of the local means and standard deviations over a sliding window (of size M) centered at each point. This requires M^2N^2 MADs. To reduce

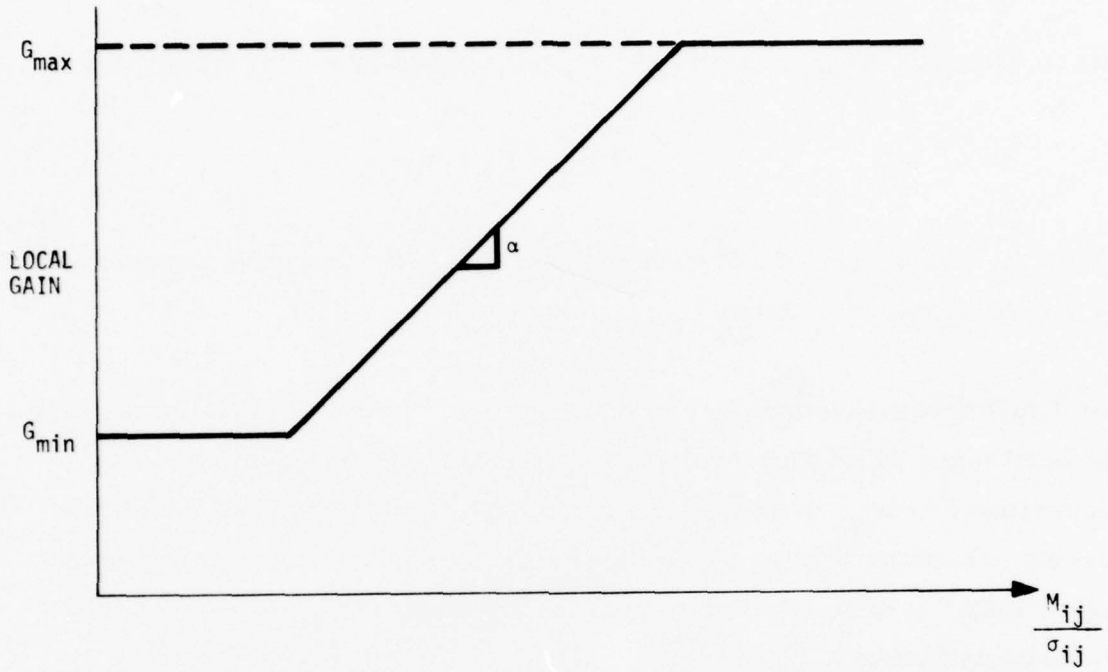


Figure 7. Gain Curve Used for Local Area Gain/Brightness Control

this effort significantly, a novel approach was taken as follows. The local mean and standard deviation were computed on nonoverlapping local areas of size M and stored. As Figure 8 shows, the local mean and standard deviation at any point in the image were estimated by a simple bilinear interpolation. This approach proved very effective because it cut down the mean and standard deviation computation to N^2 MADs versus M^2N^2 (a saving by a factor of over 100) without any of the grid-like artifacts that would be inherent in other nonoverlapping window approaches.

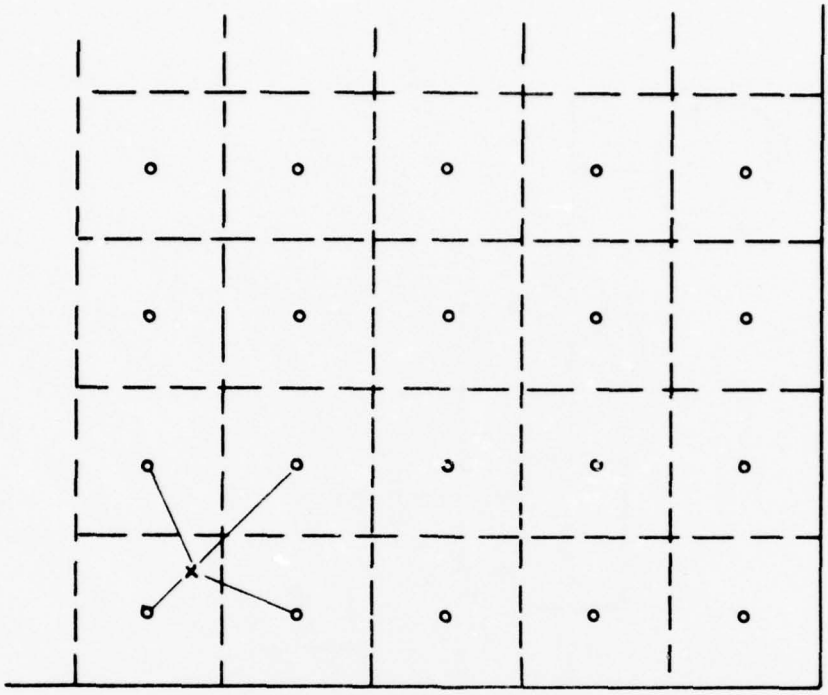


Figure 8. *Nonoverlapping Areas for Local Area Statistics Computation* (Shows the bilinear interpolation scheme used (LAGBC))

Recursive Filter Approach to LAGBC

Computing the local area means and standard deviation is similar to non-recursive spatial filtering because the local mean is really the convolution of a Rect function with the image. Convolution with the Rect function is really low pass filtering since the Sinc function is a low pass filter. Hence we can replace the local average function with a low pass filter (unity gain). Figure 9 is a realization of the LAGBC using linear low pass filters to estimate the local mean and standard deviation. Since the local mean is the low passed version of the input, the local standard deviation is approximated by the low passed version of the absolute difference between the

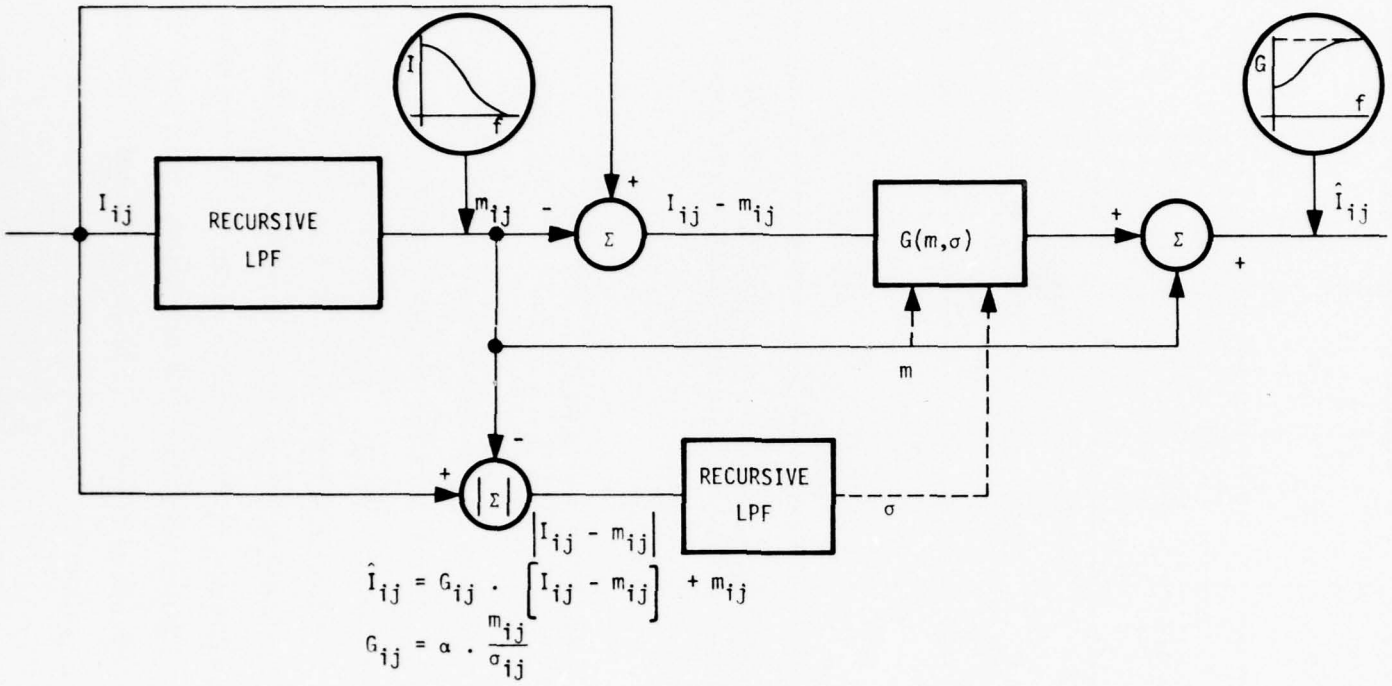


Figure 9. A Recursive Filter Approach to Local Area Gain/Brightness Control

image intensity I_{ij} and the local mean estimate M_{ij}. The following are observations concerning Figure 9:

- The frequency response at the output is that of a high emphasis filter with low frequency gain unity, and an adaptive high frequency gain G which is a function of local image statistics. Therefore, LAGBC is merely an extension of high emphasis filtering techniques.
- First-order recursive low pass filters can be used in the filter boxes. This makes real-time implementation very simple because all the advantages (one line delay, simple structure) of the recursive filter carry over.

The recursive approach was also programmed for computer simulation on the FLIR imagery. Figures 10a and 10b are nonrecursive and recursive LAGBC transformed versions, respectively, of the image in Figure 4a. The corresponding LAGBC transformed versions of Figure 5a can be found in Figures 11a and 11b. In all these images, the minimum high frequency gain Gmin = 2, Gmax = 10, and $\alpha = 0.3$. The size of the window is M = 10 for the nonrecursive case, and the cutoff frequency is fc/fs = 1/15 for the recursive filters used in the recursive formulations. Figures 12a and 12b show a vertical cross section over the target in the original and LAGBC transformed images in Figures 4a and 10a, respectively. Note that each local area brightness in Figure 12b is now closer to the global mean, and the local standard deviation is everywhere proportional to the local mean. The local contrast is emphasized while the global dynamic range has been reduced. The detail at the hot target evident in Figure 12b is quite legitimate as we see in Figure 10b.

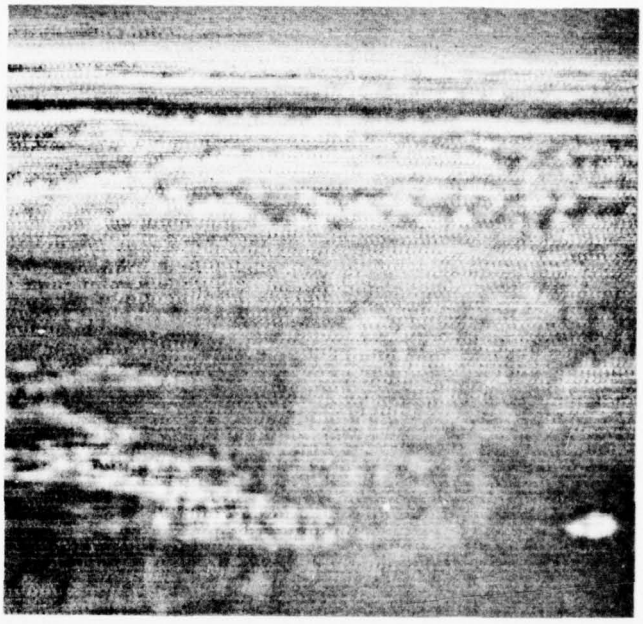


Figure 10a. Nonrecursive Local Area Gain/Brightness Control Transformed Version of Figure 4a

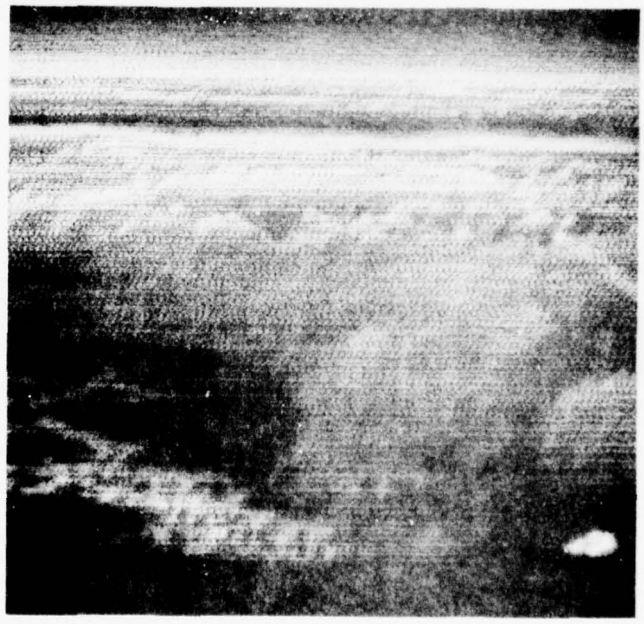


Figure 10b. Recursive Local Area Gain/Brightness Control Transformed Version of Figure 4a

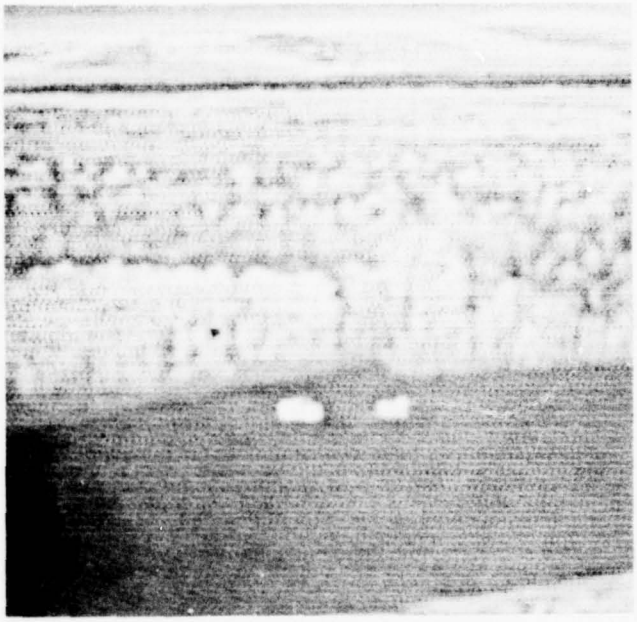


Figure 11a. Nonrecursive Local Area Gain/Brightness Control Transformed Version of Figure 5a

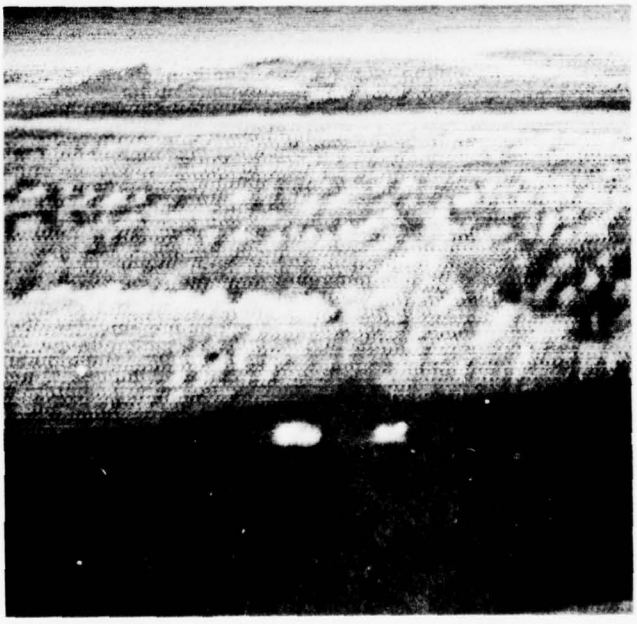


Figure 11b. Recursive Local Area Gain/Brightness Control Transformed Version of Figure 5a

Figure 12a. Intensity Profile of a Vertical Cross Section over the Target in Figure 4a (Original)

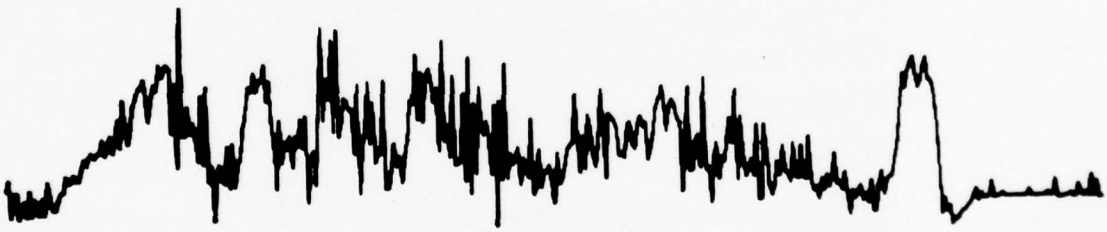


Figure 12b. Corresponding Intensity Profile of Figure 10b (Local Area Gain/Brightness Equalized)

CONTRAST ENHANCEMENT PRODUCTION RUNS

At this point, there are five contrast enhancement algorithms:

- High Frequency Emphasis Recursive Filter
- High Frequency Emphasis Nonrecursive Filter
- LAGBC Nonrecursive
- LAGBC Recursive Realization
- Local Area Histogram Modification Scheme

Each of these schemes is governed by several basic parameters which should be tuned for best performance (for example, the cutoff frequency used in the high frequency emphasis filter, the high and low frequency gains, etc).

To judge the effect of parameter changes within a scheme, we should by necessity confine our attention to a small set of images. On the other hand, for the enhancement evaluation (by human factors) we need a large set of images that have all been enhanced by the same set of algorithms so that we can choose among the algorithms. Therefore, the enhancement production runs were designed to first select the parameters for each algorithm and then to choose the best three out of the five algorithms to enhance the large set of FLIR images. This was done in three stages. A systematic set of parameter combinations for the five algorithms on two images in Stage I results in a total of 45 images. Using the selected parameters, the five algorithms are applied to nine representative thermal images. Based on the resultant 50 (10 enhanced by each algorithm) images, three algorithms will be chosen.

These three will be applied to 20 FLIR images to get 60 enhanced FLIR images. This will make up a subset of the 120 (40 x 3) contrast-enhanced FLIR images to be input to the FLIR image enhancement evaluation (add on).

SECTION III

MRT ENHANCEMENT

INTRODUCTION

Minimum resolvable temperature (MRT) is a measure of the system sensitivity of a FLIR imager as a function of the spatial frequency. A typical MRT curve is shown (solid line) in Figure 13a. The MRT approaches infinity at the cutoff frequency of the modulus transfer function (MTF) of the system. The MRT at a given frequency can be enhanced if the noise power can be reduced at that frequency. However, we do not want to degrade the signal by lowering the effective MTF of the system. There are two approaches to MRT enhancement: intraframe and interframe averaging.

Intraframe averaging involves smoothing the image spatially. A spatially invariant low pass filter reduces the noise power at higher frequencies, but it also attenuates the higher frequency information corresponding to edges, resulting in blurred detail. This effect is shown by the dotted line in the MRT curve of Figure 13a, where the MRT of the enhanced image approaches infinity at a lower frequency due to the degradation of the system MTF by the low pass filter. We are investigating algorithms for intraframe MRT enhancement that yield higher overall S/N ratios while preserving the edges in FLIR imagery.

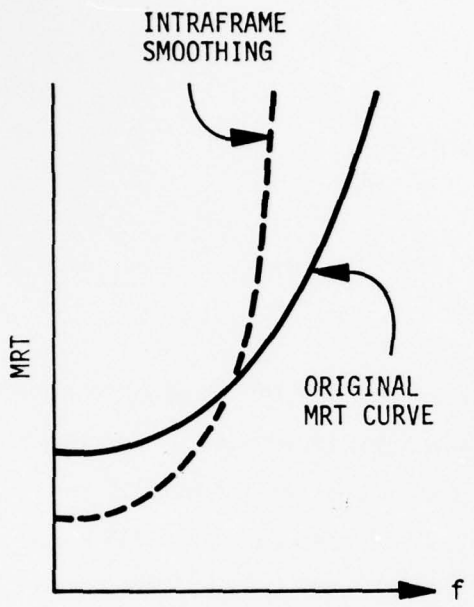


Figure 13a. Original MRT Curve and Intraframe Smoothed Curve

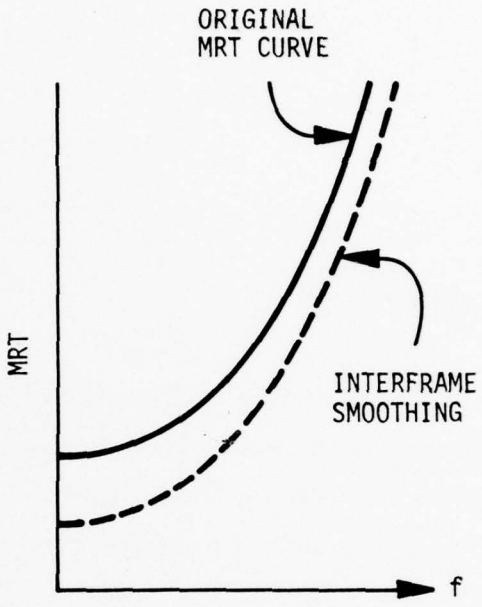


Figure 13b. Original and Interframe Smoothed MRT Curves

Interframe MRT enhancement is temporal averaging of several registered and stacked FLIR frames. Since the noise is uncorrelated between successive frames, the noise variance is reduced by a factor n, the number of stacked frames. The limitations are that resolution is lost if registration is not maintained, and, also, the data rate is reduced. Figure 13b (dotted line) shows the MRT improvement that can be expected from interframe averaging.

In this reporting period, all the MRT enhancement algorithm coding and parameter selection for optimum performance was completed. In the following subsections, we briefly describe the schemes employed, present

simulation results on representative images, and discuss the relative merits of the schemes in light of their effectiveness and potential for real-time implementation.

Measurement of S/N Improvement by MRT Enhancement

To quantify the S/N gains accruing from the MRT enhancement algorithms, relatively noise-free FLIR (and Thermoscope) images were degraded by adding uncorrelated noise using a BLIP detector model. The noise was generated from a zero mean Gaussian pseudo-random number generator with standard deviation σ_{ij}, a function of the image intensity at the point (i, j). In the following examples, the noise standard deviation refers to the average σ used for the whole image, i.e., the σ computed for the mean intensity of the image. The S/N ratio refers to the following ratio: (average intensity)/ RMS noise σ.

To measure the S/N gain in the enhanced images, histograms of the intensity differences at every pixel between the original (o)-noisy (n) image and original (o)-enhanced (e) image were formed. The standard deviations of these histograms, σ_{o-e}, also reflect the degradation due to excessive smoothing introduced, for example, by the MRT enhancement schemes. The S/N gain was then defined to be

$$\text{S/N Gain} = 10 \text{ Log } \frac{\sigma_{o-n}}{\sigma_{o-e}}$$

The S/N gain was measured not only on the entire image, but also on a small box surrounding the target in the FLIR images. The reason for this is that the S/N gain measured on the whole image does not reflect the

blurring of the sharp target edges because the pixels corresponding to the target edges are but a minuscule portion of the entire image. On the other hand, the measurement on the small box surrounding the target shows the blurring of the target as the mean square error between the original and the enhanced boxes. This measure was used to aid the parameter selection in the variable width filters that follow and to supplement subjective evaluation of the blurring induced.

INTRAFRAME SMOOTHING FILTERS

We are investigating three classes of intraframe smoothing filters to smooth the noise in the images without blurring the high frequency detail:

- Median Filter
- Hysteresis Filter
- Scene Adaptive Variable Width Filters

Median Filter

Figure 14a shows the basic structure of a 3 x 3 median filter; the median intensity in a small window replaces the pixel intensity at the center of the window. We have investigated 3 x 3 and 5 x 5 window median filters. The advantage of the median filter is that it does not blur a step edge because the median intensity of the window is dominated by the intensities in that side of the edge the window is centered on. Corners, however, are slightly affected, but this is deemed insignificant for FLIR imagery. This filter is most useful with spiky noise (as a prefilter operation for edge operators, for example). Figure 15a is an original NVL supplied thermal image with

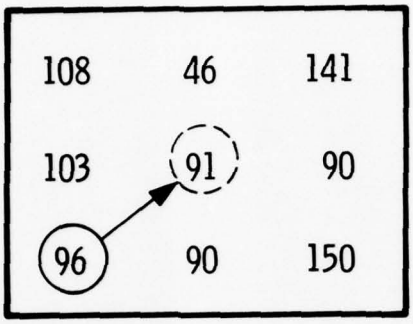


Figure 14a. 3 x 3 Median Filter Concept

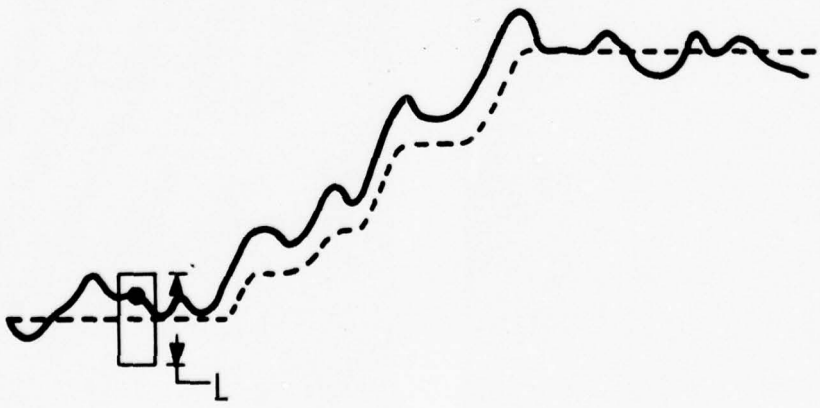


Figure 14b. Hysteresis Filter Concept

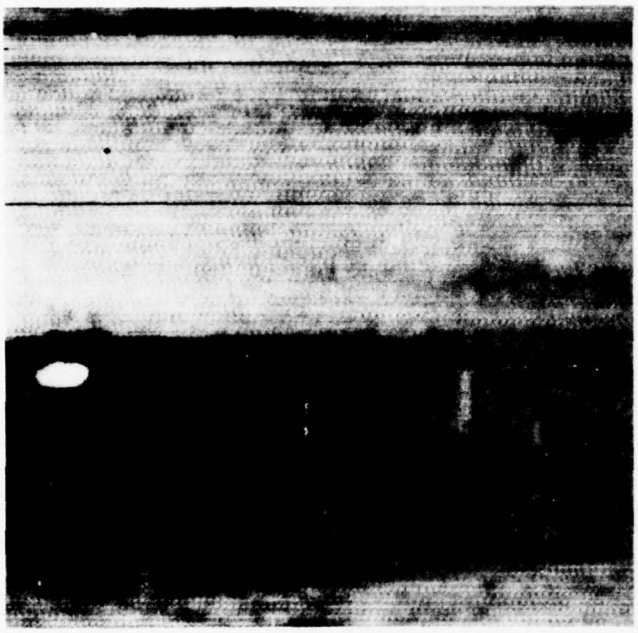


Figure 15a. Original Thermal Image

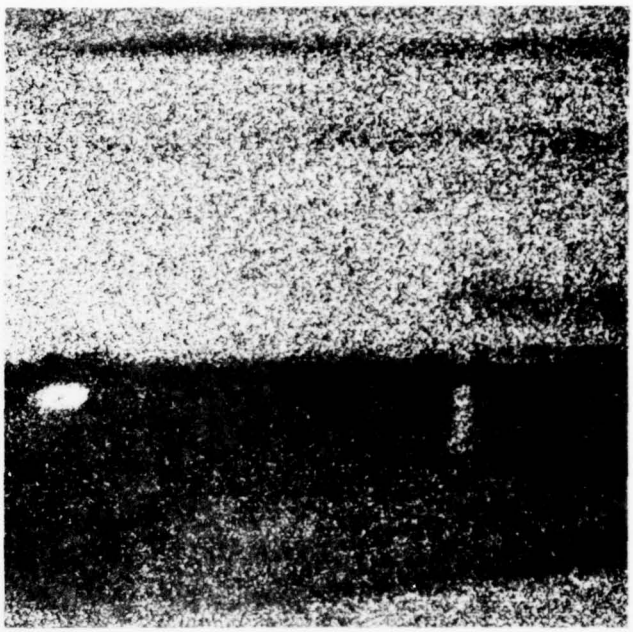


Figure 15b. Thermal Image with Noise Added SNR = 3

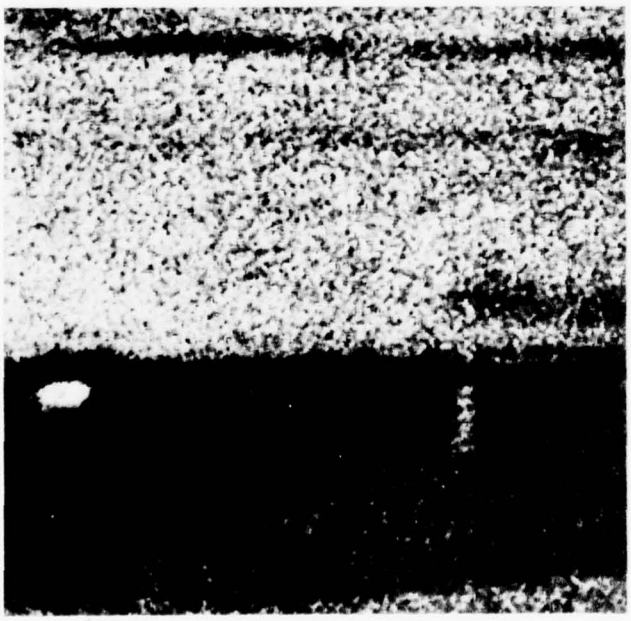


Figure 15c. Median Filtered
3 x 3 S/N Gain = 1.0 dB

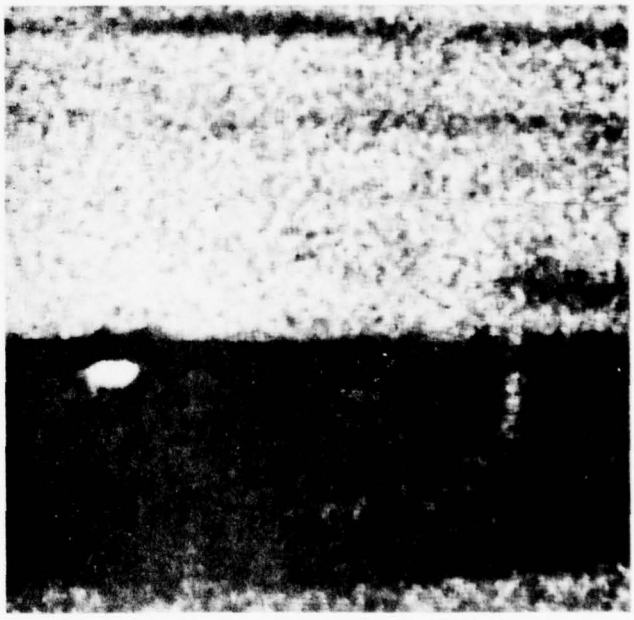


Figure 15d. Median Filtered
5 x 5 S/N Gain = 4.3 dB

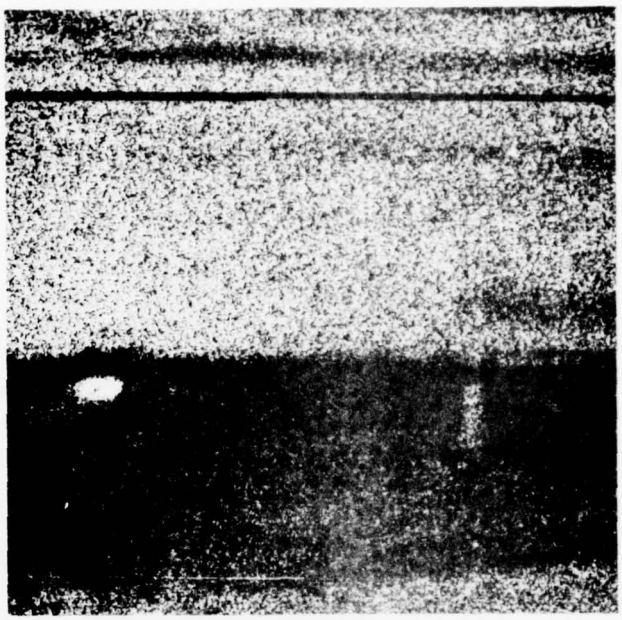


Figure 15e. Two-Dimensional Hysteresis Filtered S/N Gain = 0.9 dB

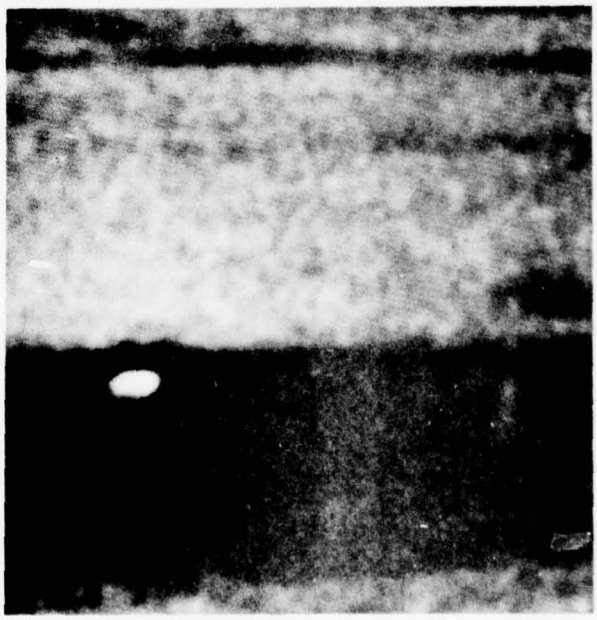


Figure 15f. Nonadaptive Smoothing with a Linear Low Pass Filter (W = 7)

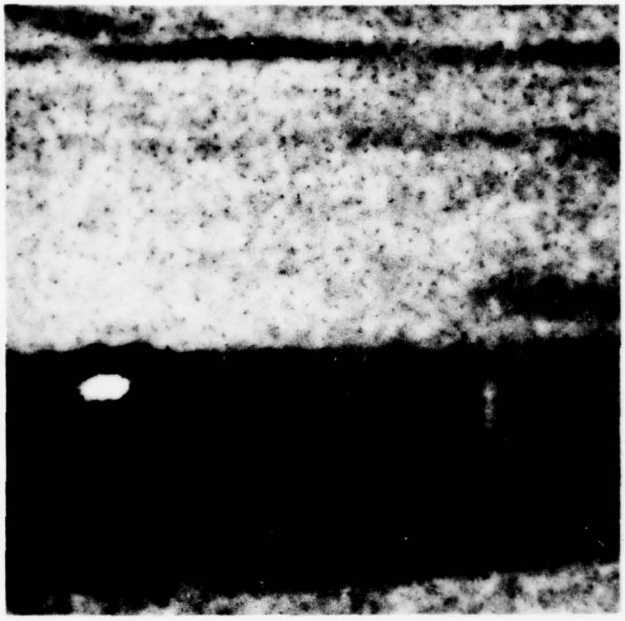


Figure 15g. Curvature Directed Gaussian (*Nonrecursive Filter*)

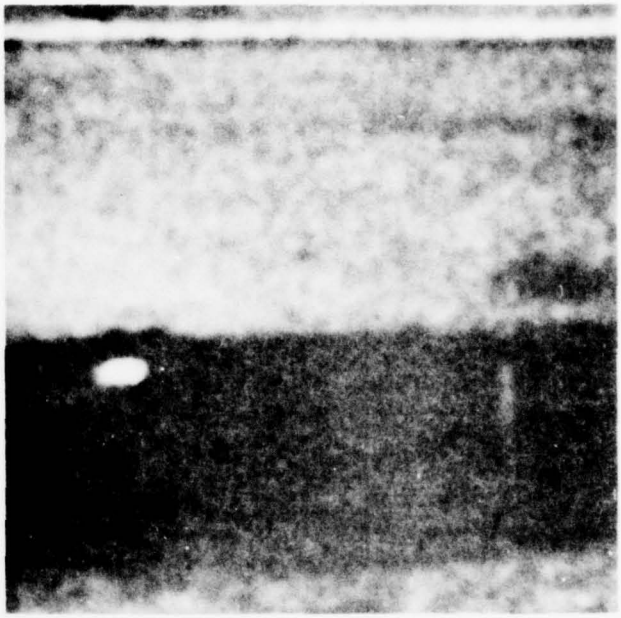


Figure 15h. Gradient Directed Gaussian Filter

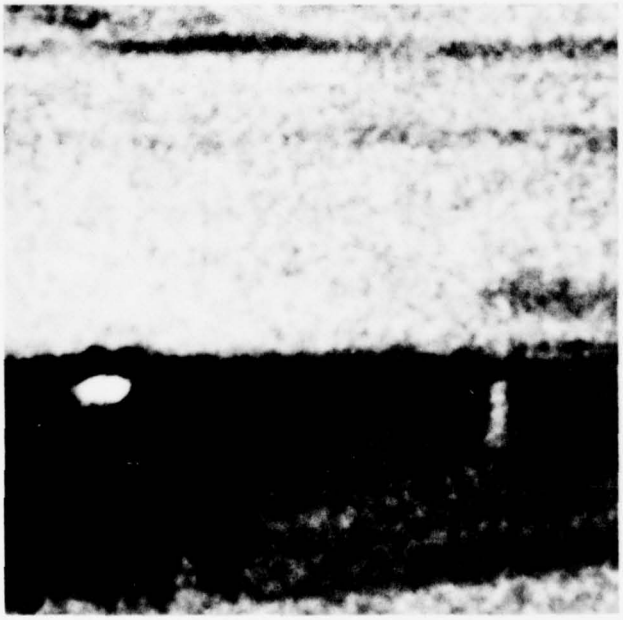


Figure 15i. Curvature Directed Recursive Filter

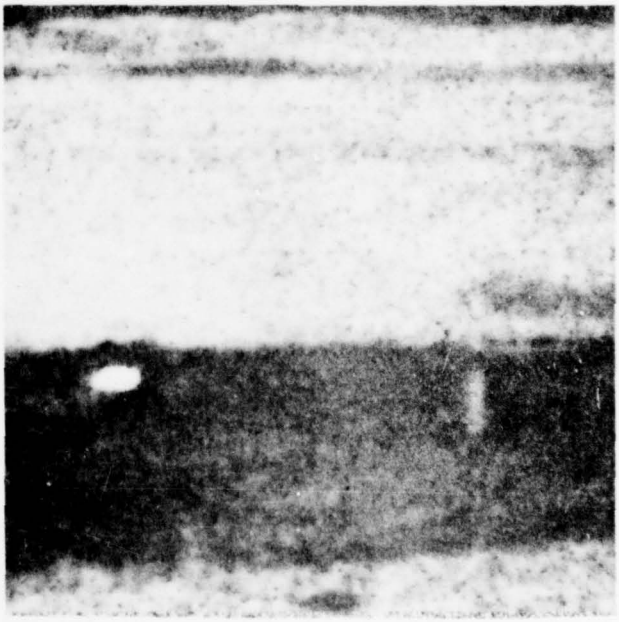


Figure 15j. Gradient Directed Recursive Filter

an average signal to rms noise ratio of SNR = 3.0 (5 dB). Figures 15c and 15d are 3 x 3 and 5 x 5 median filtered versions of this image, respectively. We see that the 5 x 5 median filter is far superior to the 3 x 3 filter. Indeed, this subjective impression was confirmed by the S/N gains measured on the full frames (as described earlier): 1.0 dB for the 3 x 3 median filter versus 4 dB for the 5 x 5 median filter.

Hysteresis Filter

Hysteresis filtering is a method which preserves significant extrema in an image while smoothing out local fluctuations. Two-dimensional hysteresis filtering is achieved simply by applying one-dimensional filters sequentially in orthogonal directions (separability assumption). For one dimension, the algorithm is as follows:

$$\text{If } (k = 1),\ y_k = x_k$$

$$\text{If } (y_k < x_{k+1} - L/2),\ y_{k+1} = x_{k+1} - L/2$$

$$\text{If } (y_k > x_{k+1} + L/2),\ y_{k+1} = x_{k+1} + L/2$$

$$\text{Otherwise, } y_{k+1} = y_k$$

Here x_k is the original pixel intensity, y_k is the filtered intensity, and L is the hysteresis window width. Figure 14b illustrates the one-dimensional hysteresis filter concept. Figure 15e is the hysteresis filtered version of the noisy Figure 15b with a hysteresis "window" L = 8 grey-levels (out of 256). The S/N gain is minimal (less than 1 dB) which we can conclude subjectively by looking at Figure 15e as well.

Adaptive Variable Width Filters

The problem with smoothing with a nonadaptive linear low pass filter is that, while the overall S/N ratio improves by a factor K (K is the size of an equivalent rectangular averaging window), the target edges get blurred which is surely an undesirable effect. Figure 15f shows that the image was filtered with a Gaussian low pass filter with an equivalent rectangular width of 7 to the side. The S/N gain measured by the technique outlined above was 6 dB, but the target has lost the edge detail. What is required is an adaptive filter that detects target and other significant edges and directs the degree of smoothing in each local area, adaptively. We are investigating adaptive variable width filters that are directed by (1) the local curvature, and (2) the local gradient of the image.

Local Gradient-Directed Adaptive Filters--The approach to a gradient-directed, variable-width spatial filter we chose to investigate was to compute the local contrast in each of four principal directions (0 deg, 90 deg, $\pm$ 45 deg) and choose one of five possible Gaussian low pass filters with equivalent rectangular width L = 0, 2, 3, 5, and seven pixels based on the magnitude of the local contrast. L = 0 denotes the condition where no spatial filtering is performed.

For each direction, $\Delta\vec{x}$, the local contrast at $\vec{x}$,

$$C(x) = \frac{\left| \sum_{m=1}^{3} f(x + m\Delta\vec{x}) - \sum_{m=1}^{3} f(x - m\Delta\vec{x}) \right|}{\sum_{m=-3}^{3} f(x + \Delta\vec{x})}$$

is compared with a set of thresholds,

$$\theta_k = \frac{1}{\epsilon 2^{k-1}}; \quad k = 1, 2, 3, 4$$

The maximum L for each direction is determined by the following logic:

$L = 7$ if $C \leq \theta_4$

$L = 5$ if $\theta_4 \leq C \leq \theta_3$

$L = 3$ if $\theta_3 \leq C \leq \theta_2$

$L = 2$ if $\theta_2 \leq C \leq \theta_1$

$L = 0$ if $\theta_1 \leq C$

Finally, the min-max L for the four directions is chosen for the size of the spatial filter at point x in the image. The local contrast as measured by C(x) is a smoothed measure of the local image activity. If there is a genuine edge in the picture at x, C(x) will have a large value and, correspondingly, a smaller filter size L will be selected. Note that isolated noise points will not give a large value of C(x) because this is essentially a smoothed measure of the local contrast.

Local Curvature-Directed Adaptive Filters--Use of the gradient as a criterion for a selection of a variable filter width to avoid smoothing across edges leaves something to be desired. The reason is that no error in smoothing an edge results if the edge gradient is constant within the width of the smoothing window.

To investigate this further, assume that an edge has the form of a plane ridge, at least over the smoothing window area. If the x-coordinate is in the direction of the gradient, the edge can be described by

$$f(x,y) = a + bx + cx^2 + dx^3$$

The smoothed intensity (for a box filter) is then

$$F(x,y) = \frac{1}{L^2} \int_{x-L/2}^{x+L/2} du \int_{y-L/2}^{y+L/2} dv \; f(u,v)$$

$$= (a + \frac{cL^2}{12}) + (b + \frac{dL^2}{4})\, x + cx^2 + dx^3$$

For specific error in the smoothed value at the center of the window,

$$\left| \frac{F(0,0) - f(0,0)}{f(0,0)} \right| \leq \epsilon \text{ if } \frac{L^2}{12} \left| \frac{c}{a} \right| \leq \epsilon$$

Thus, the smoothing error does not depend upon the local gradient, b, but rather on the local curvature, c. Furthermore, we now have a criterion for selecting the smoothing window size to achieve a specific error.

This analysis suggests the following approach. At each pixel, the least-square error estimate of the coefficient a and c in (x, y) is computed under the assumption that x is in one of the four principal directions (0 deg, 90 deg, $\pm$ 45 deg). For a seven-pixel, one-dimensional window these are

$$a = (7g_a - g_c)/21 \qquad c = (g_c - 4g_a)/84$$

where

$$g_a = \sum_{i=-3}^{3} f_i$$

$$g_c = \sum_{i=1}^{3} i^2 (f_i + f_{-i})$$

$$f_i = f(x + i\Delta x, y)$$

Use of a seven-pixel window provides the advantage of some noise smoothing in the filter width selection criterion.

Having computed a and c for each of the four principal directions, the maximum L for which $|c| L^2 \leq 12\epsilon |a|$ is found in each direction. Then the minimum of the max L for the filter width is chosen.

Filter Shape and Implementation for the Smoothing Filters

The five smoothing filters required in the above approaches were realized using both recursive and nonrecursive structures because recursive filters are easier to implement in real-time hardware than nonrecursive filters.

Nonrecursive Filter Structure--A Gaussian frequency response was shown in Figure 2. The corresponding impulse response is also Gaussian. The four Gaussian filters of equivalent rectangular width (L = 2, 3, 5 and 7) correspond to Gaussian impulse responses of size W given by W = L + (L+1)/3 where W gives the size of the Gaussian FIR filter in pixels equivalent to the rectangular averaging window of side L. This ensures that the truncated filter has a smooth transition with no Gibbs' phenomenon

occurring in the frequency domain. The advantage of these nonrecursive filters is that they have zero phase and therefore preserve the linear phase of the total transfer function. But from an implementation viewpoint, the largest filter requires a 10-pixel diameter convolution and is time consuming in a simulation as well as expensive in a hardware implementation. Also, the gradient and curvature estimates computed at each image point in four directions are nonrecursive in nature and require access to seven lines of the image at the same time. To alleviate this, recursive alternatives that would facilitate eventual implementation were sought.

Recursive Realization of the Smoothing Filters--Second-order Butterworth recursive filters were programmed for the four smoothing filters of different widths required. The frequency response of a second-order filter was shown in Figure 2 and closely approximates the Gaussian shape. The recursive equations for the filters are given by

$$y_k(i,j) = d_k\, g(i-1,j-1) + \sum_{m=0}^{2} \sum_{n=0}^{2} c_k(m,n)\, y_k(i-m,j-n)$$

$$c_k(0,0) = 0;\quad c_k(1,1) = -a_k^2;\quad c_k(2,2) = -b_k^2$$

$$c_k(1,0) = c_k(1,0) = a_k;\quad c_k(2,0) = c_k(0,2) = -b_k;\quad c_k(2,1) = c_k(1,2) = a_k b_k$$

$$d_k = (1 - a_k + b_k)^2 \qquad a_k = 2\exp(-\gamma_k)\cos(\gamma_k)$$

$$b_k = \exp(-2\gamma_k) \qquad \gamma_k = 2/w_k;\quad w_k = 0, 2, 4, 5, 16 \quad (k = 0, 1, 2, 3, 4)$$

where $y(i,j)$ is the smoothed image and $g(i,j)$ is the input image.

The advantage of this recursive filter structure is that only three line delays (two previous filtered lines and the previous input line) are needed to filter the input for any equivalent rectangular filter size.

Of course, it is not enough to have a recursive implementation for the smoothing filters; we also need recursive estimators for the local gradient (contrast) and the local curvature in the four principal directions (used to direct the smoothing).

For the gradient-directed filters, the normalized contrast at a point $\vec{X} = (x, y)$, $C(x) = (b/a)$ in any given direction $\Delta\vec{X}$ is given by the recursive equation

$$b(\vec{X}) = c_1 b(\vec{X} - \Delta\vec{X}) + c_2 |g(\vec{X}) - g(X - 2\Delta\vec{X})|$$

and

$$a(\vec{X}) = c_1 a(\vec{X} - \Delta\vec{X}) + c_2 g(\vec{X})$$

where

$$c_1 = \text{Exp}\ (-\gamma)$$

$$c_2 = 1 - c_1$$

$$\gamma = 2/W$$

Similarly, the local curvature at a point $\vec{X}$ normalized by the local mean is given by d/a, where

$$d(\vec{X}) = c_1\ d(\vec{X} - \Delta\vec{X}) + c_2[g(\vec{X}) - 2g(\vec{X} - \Delta\vec{X}) + g(\vec{X} - 2\Delta\vec{X})]$$

Figures 15g through 15j are the filtered versions of the noisy image in Figure 15b, using the curvature-directed and gradient-directed nonrecursive and recursive filters. Note that the adaptive filters succeed in smoothing out the noise in relatively large uniform areas of the image while preserving the edges of the target and detail in the image. In particular, a comparison of Figure 15f, which was nonadaptively filtered with a Gaussian filter with an equivalent width of seven pixels, shows that the adaptive filters are superior in this respect. Also note that the recursive filters are as effective as the Gaussian nonrecursive filters, which is a welcome discovery because they are so much easier to implement.

Parameter Selection for the Adaptive Filters

In the adaptive filters, the permissible error parameter ϵ should be selected so that maximal overall noise smoothing is achieved while the target edges are preserved. In addition, each of the four algorithms (recursive and nonrecursive implementations of the curvature-directed and gradient-directed smoothing filters) has to be individually optimized.

In order to do this, we selected a small box surrounding a sharp target and another box containing a relatively uniform background in the noisy image. The algorithms were then applied to the two boxes and the S/N gains against the noisy image were measured both over the target box and the background box. As ϵ is allowed to increase, successively larger filters are selected at a given point in the image. For $\epsilon = 0$, for example, no filtering would be done at all, while for a sufficiently large ϵ, the largest filter ($W = 7$) would be chosen at all points over the image. The measured S/N gain on the target and background boxes peaks out at a certain ϵ value due to the

nature of this S/N gain defined earlier. For very small ϵ, S/N gain is small because little filtering is done, and for very large ϵ, the noise points are smoothed. However, the MSE of the smoothed image from the original image actually increases because of the blurring of high frequency detail by excessive smoothing. Hence, the optimum ϵ should be close to the value at which the S/N gain over the target/background box is maximized.

To fine tune this parameter further, entire noisy images were processed and the filter selection at every point in the image was histogrammed. Obviously the only true criterion for judging target edge sharpness and smoothness of background is subjective and visual. But this criterion was strengthened by examining the filter selection histograms and choosing ϵ to ensure that the large filters would be chosen frequently over the background, but the smaller (0 to 2) filters are also represented over the target regions.

INTERFRAME TEMPORAL AVERAGING

The key issue in interframe temporal averaging for MRT enhancement (aside from frame storage) is temporal responsiveness with and without frame registration in a dynamic environment.

One can visualize implementing a hierarchy of interframe averaging techniques that are selectable depending upon prevailing conditions. For example, in heavy haze or rain, one may choose interframe stacking without frame registration because registration would be ineffective anyway.

In this case, both background and targets would suffer motion blur; perhaps moving targets could be enhanced to the point of detection, recognition, and identification, whereas moving targets would only be detectable.

At the final level, a target may be detectable but not recognizable or identifiable because of noise. Registration on the target in this case would enhance its definition (especially if moving so that the background blurred in the frame stacking process) and would perhaps permit recognition and identification.

To temporally average successive frames, they first have to be registered with one another with respect to some feature of interest in the scene because motion of the sensor platform causes the scene to translate and rotate substantially (several pixels) even from one frame to the next. A simple registration scheme would involve pure translational correction, i.e., no rotation or scale change. Such a scheme was simulated with successive digitized and noise-added sequences of FLIR frames of tactical targets containing a stationary target (a truck) and another containing a target (an APC) moving at right angles to the platform trajectory. The registration was done using an automatic maximum correlation tracker algorithm against a corner of the target, in each case, to estimate its translation from frame to frame. The frames were then averaged after being aligned to account for the target motion.

Frame Registration

Figures 16a and 16b show the correlation and reference windows used in the registration algorithms. The 5 x 5 reference window is chosen on the first frame in the sequence to encompass a high contrast target corner. Note that the 5 x 5 window can be at grid spacings of 1, 2, or 3 which means that the corresponding windows are actually 5 x 5 to 15 x 15 pixels wide in the image. The correlation window is a bigger window (11 x 11) or the next frame, chosen around the same center picture coordinates as the reference window. Both these windows are smoothed heavily with a 15 x 15 Gaussian filter of equivalent rectangular width of 7 x 7 prior to correlation. The correlation window is correlated with the reference window on frame 2 and the translation (Δx, Δy) that yields the highest value of the correlation is found. This translation (Δx, Δy) is taken into account when shifting the second frame by this amount and adding to the first. The process continues with the point of maximum correlation on the second frame being the center of the reference window on that frame, and the third frame is correlated against this reference window, and so on. Two correlation criteria were evaluated:

a. Squared Canonical Product Correlator (SCPC)

$$\rho_k(m,n) = \frac{\left(\sum_{|p|,\,|q| < 2} r(p,q)\; c_k(m+p, n+q) - 0.04\bar{r}\;\bar{c}_k(m,n) \right)^2}{\sigma_r^2\;\sigma_{mn}^2}$$

b. Minimum Absolute Difference Correlator (MADC)

$$\rho_k(m,n) = \left| \sum_{|p|,\,|q| < 2} [c_k(m+p,\, n+q) - r(p,q)] - \bar{c}_k(m-n) + \bar{r} \right|$$

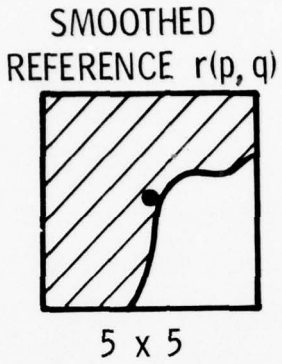


Figure 16a. Smoothed Reference Window

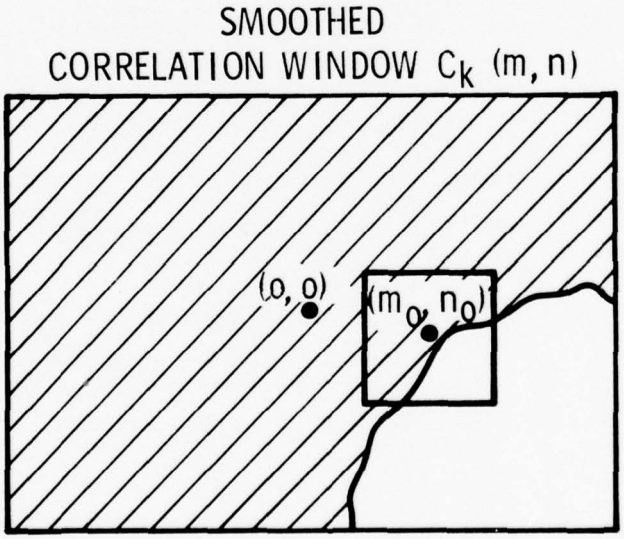


Figure 16b. Smoothed Correlation Window Used in the Correlation Tracker Algorithm for Translational Registration

where

$$\bar{r} = \sum_{|p|,|q|\le 2} \sum r(p,q)$$

$$\overline{r^2} = \sum_{|p|,|q|\le 2} \sum r^2(p,q)$$

$$\sigma_r^2 = \overline{r^2} - 0.04\,\bar{r}^2$$

$$\bar{c}_k(m,n) = \sum_{|p|,|q|\le 2} \sum c_k(m+p,n+q)$$

$$\bar{c}_k^2(m,n) = \sum_{|p|,|q|\le 2} \sum c_k^2(m+p,n+q)$$

$$\sigma_{mn}^2 = \overline{c_k^2}(m,n) - 0.04[\overline{c_k}(m,n)]$$

Experiment 1--From each sequence, a set of nine noise-added ($\overline{SNR}$ = 3) FLIR frames 1/15 of a second apart were stacked and the S/N gain against the noise-free frame was measured at intervals of two, four, six and nine frames. The total integration time here is 9 x 1/15 = 0.60 second. Figures 17a through 17f show the results of the interframe stacking for the stationary target case. Note that after stacking nine frames (Figure 17f) considerable noise reduction has been achieved. Ideally, of course, when we stack N frames, we should realize gains in the S/N ratio against the noisy image of 10 $\text{Log}_{10}\sqrt{N}$ dB because the noise is uncorrelated from frame to frame while the scene is assumed to be the same. In Figure 18, the solid curve shows the actual S/N gain measured (as previously described) over the whole frame versus the number of frames stacked, and the broken curve shows the theoretical expectation in the S/N gain. The reason we do not monotonically continue to realize increasing gains in S/N ratio (as predicted theoretically) by stacking more frames is that the translational

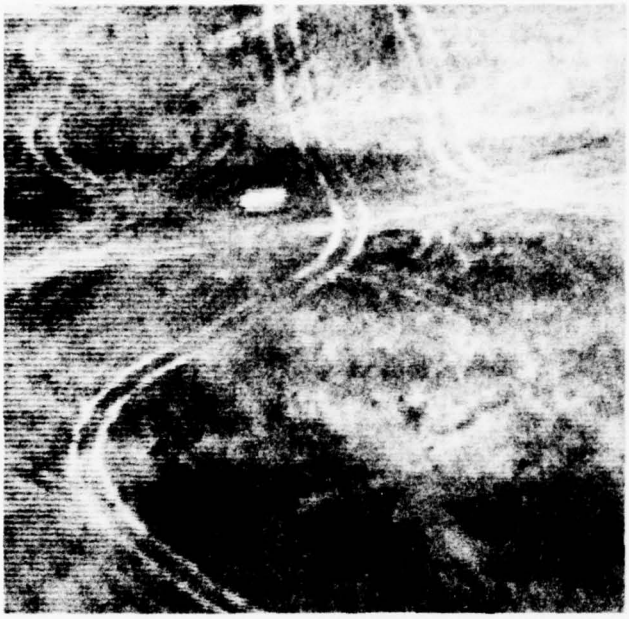


Figure 17a. First Noise-Free Frame (Stationary Target)

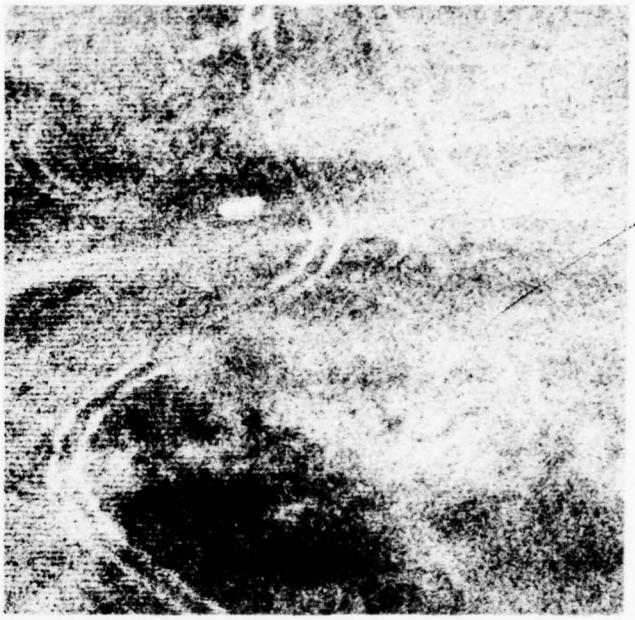


Figure 17b. First Noisy Frame SNR = 3

Figure 17c. After Averaging Two Noisy Frames

Figure 17d. After Averaging Four Noisy Frames

Figure 17e. After Averaging Six Noisy Frames

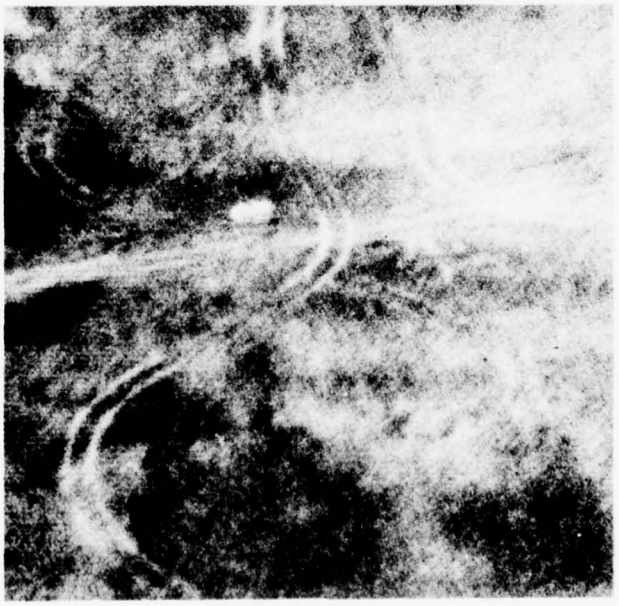


Figure 17f. After Averaging Nine Noisy Frames

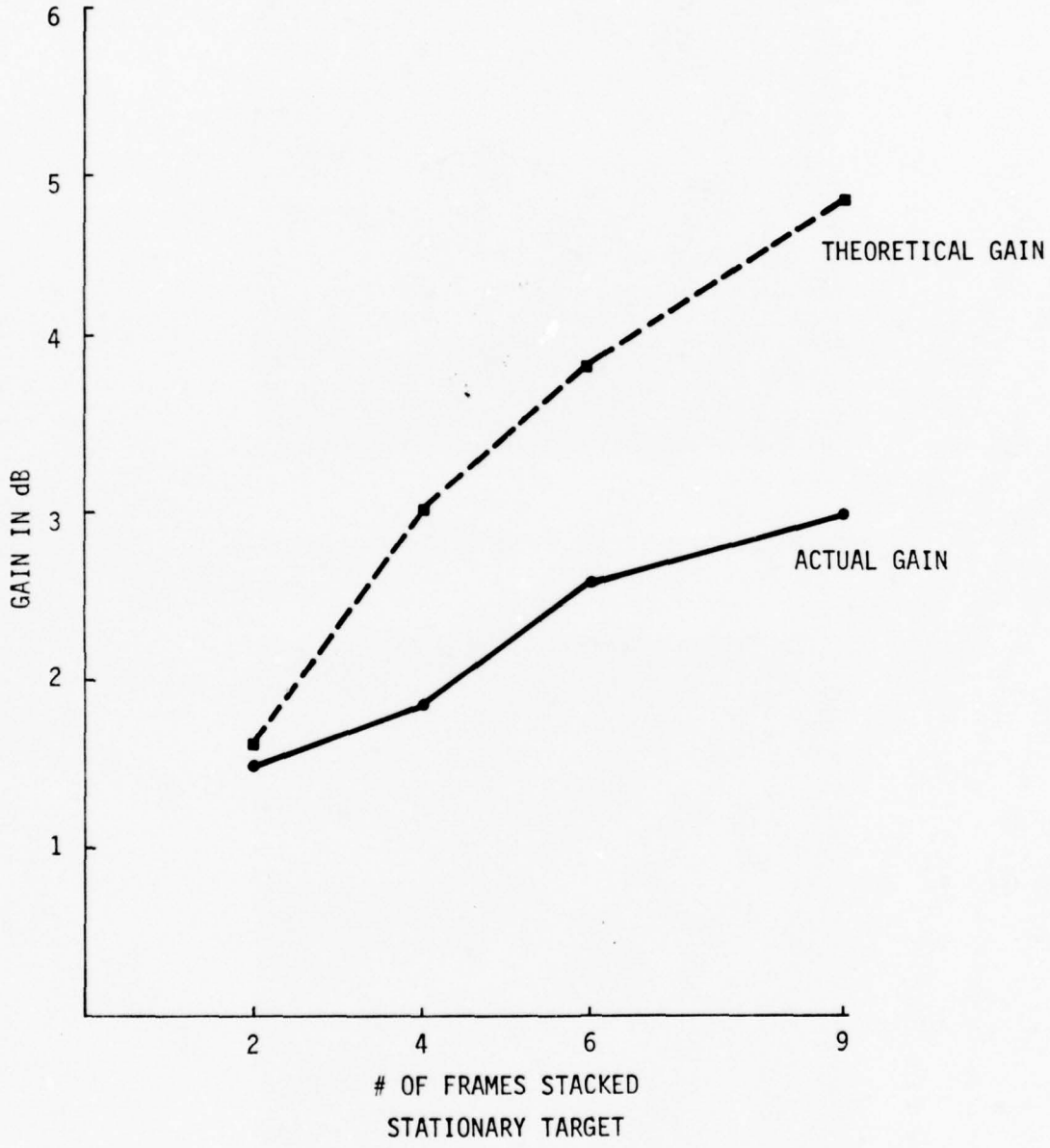


Figure 18. Signal-to-Noise Ratio Gains Accrued by Frame Averaging vs. the Number of Frames Stacked for the Stationary Target Case

registration is not fully correcting the sensor platform motion (because of skew and rotation) and the attendant scene blurring causes the S/N gain measured to fall short of the theoretical. A similar set of sequences with a moving target (at right angles to the sensor motion) is shown in Figures 19a through 19f, and we see that while the target definition improves (because the registration is done against the target) the background starts getting blurred near the end of the sequence. The corresponding S/N gains measured against the original image also show this trend (Figure 20).

It is interesting to note that, with the stationary target sequence in Figure 17, the S/N gain measured continues to be monotonic because registration is on the stationary target. On the other hand, when the registration was done on the moving target in Figure 19, the background became so blurred in the registered sequence that the S/N gain measured actually deteriorated from the six to nine frame stacking.

Experiment 2--The human eye temporally integrates over approximately 0.20 second. Therefore, to achieve perceptibly superior performance in practice, we must prove feasibility of stacking frames that span longer durations. Accordingly, a second experiment was performed identically to the first but with the nine stacked frames being precisely 0.20 second apart (i.e., every sixth frame). Also much larger noise was added to these frames (S/N = 1.0). The larger time lapse between successive frames taxes the registration algorithms somewhat and, moreover, blurs the background very quickly.

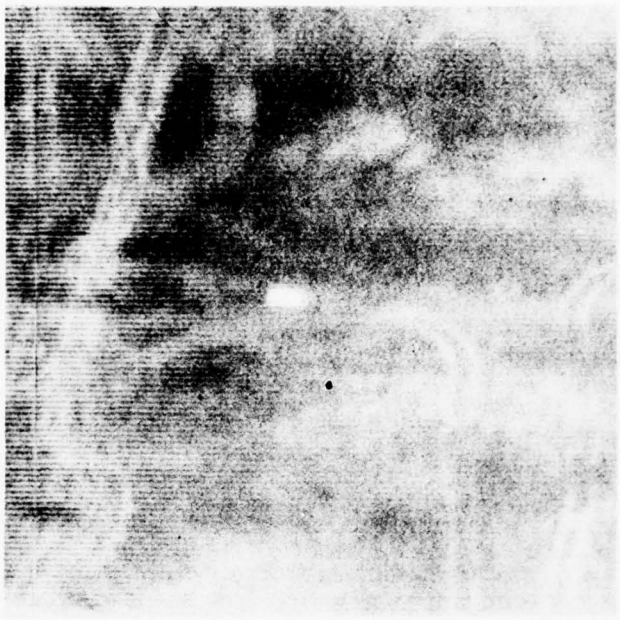


Figure 19a. First Noise-Free Frame (Moving Target)

Figure 19b. First Noisy Frame SNR = 3

Figure 19c. After Averaging Two Noisy Frames

Figure 19d. After Averaging Four Noisy Frames

Figure 19e. After Averaging Six Noisy Frames

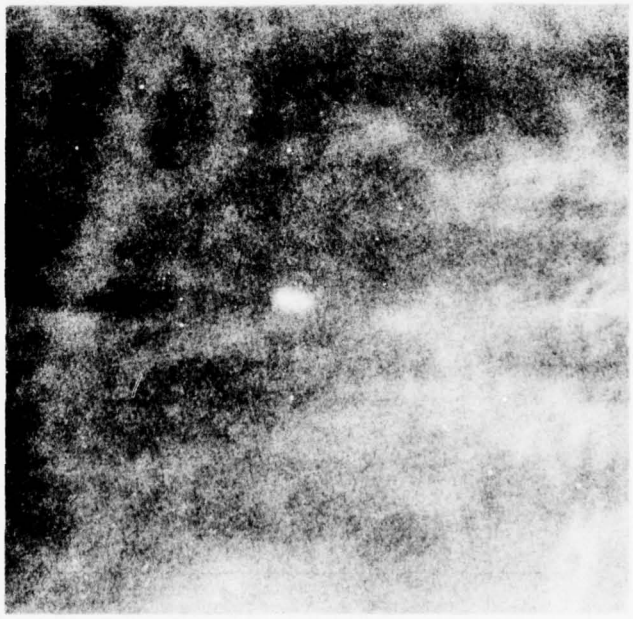


Figure 19f. After Averaging Nine Noisy Frames

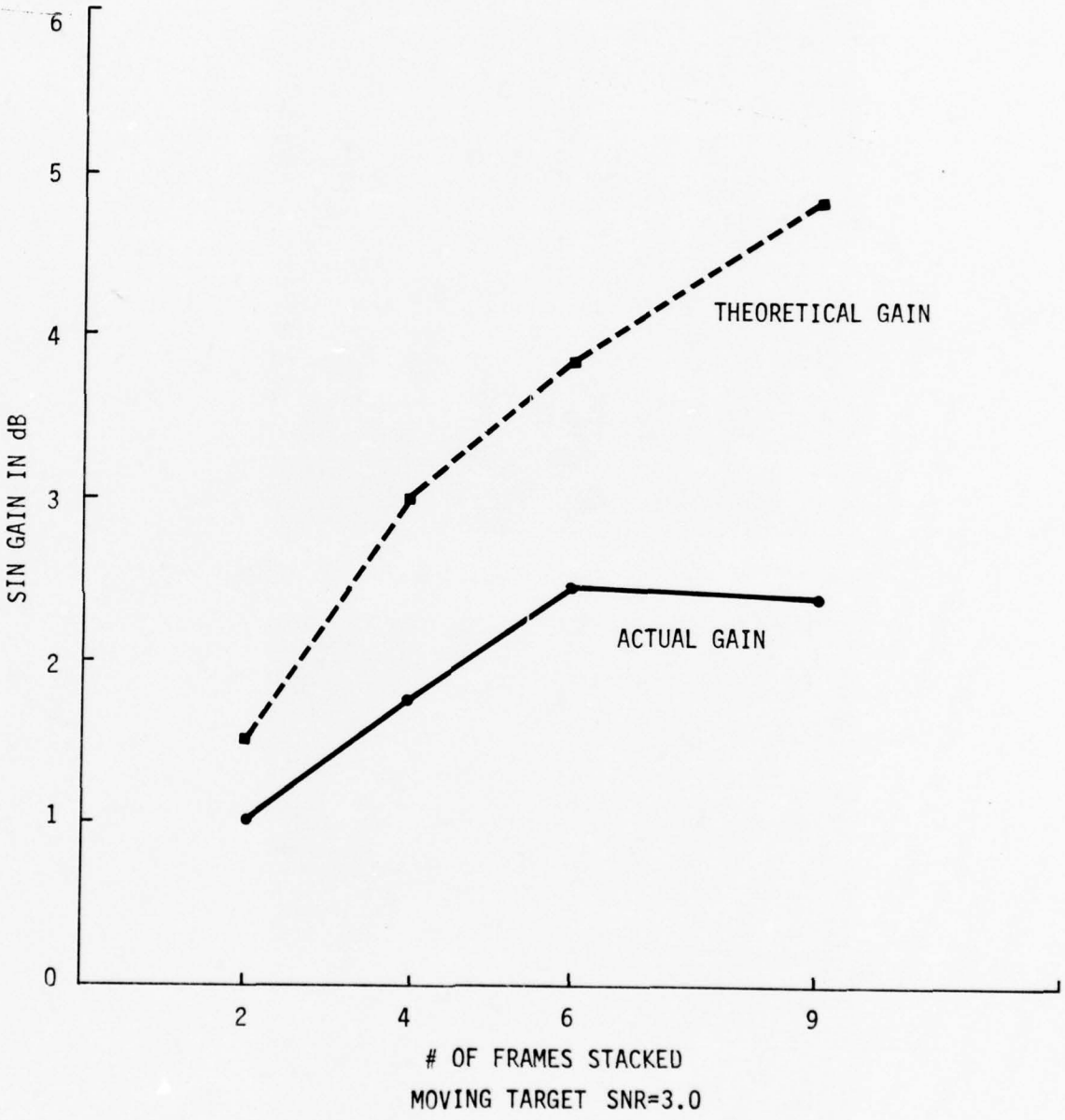


Figure 20. Signal-to-Noise Ratio Gains vs. the Number of Frames Averaged for the Moving Target Case

The above examples show that simple translation registration is indeed adequate for real-time stacking of FLIR frames to achieve MRT enhancement at reduced data rates. The Canonical Sum of Products correlator was found to be satisfactory for the purpose.

SECTION IV

RESOLUTION RESTORATION

INTRODUCTION

The task of resolution restoration for thermal imagers can be divided into two areas:

- Full-frame focus restoration for equalizing resolution out to the optical Rayleigh diffraction limit, and
- Super-resolution to extend system resolution beyond the Rayleigh diffraction limit.

In this reporting period, both the full-frame resolution restoration and partial frame iterative super-resolution algorithms were coded, simulated, and production runs were begun to enhance FLIR images evaluation.

FULL-FRAME FOCUS RESTORATION

This task consisted of modeling the space variant system modular transfer function (MTF) blur and developing piecewise linear inverse filters to restore the blur over the full frame.

The space variant transfer function of the diffraction-limited optics was assumed (following common practice) to be Gaussian:

$$H(k_x, k_y \mid r_o) = \mathrm{Exp}\left[- \frac{k_x^2 + k_y^2}{2\sigma^2(r_o)} \right]$$

where k_x and k_y are wave numbers.

Also,
$$\sigma(r_o) = \sigma_o + \sigma_2 r_o^2$$

where r_o is the radian distance from the optical axis, σ_2 was chosen so that the on-axis $\sigma_o = 4.2$, and the worst case diagonal $\sigma_{max} = 3.54$. In the space domain, this corresponds to a worst off-axis blur wide diameter equal to 1.19 times the on-axis blur. Obviously, this blur is not very space variant. However, it is representative of the actual ratios encountered in current FLIR optics and expected in future FLIRs.

Following the OTF, the detector induces a further blurring because the sensed output is the convolution of the image at the focal plane and the rectangular detector surface. The detector width was chosen so that it has the same area as the on-axis optics Gaussian blur (Figure 21). This corresponds to a detector width W given by

$$W = \frac{\sigma_o}{\sqrt{2\pi}}$$

This model essentially assumes that the detector induces approximately the same blur again as the optics. Figure 22a shows the combined OTF and detector MTF frequency response on axis and is seen to be very close to a Gaussian.

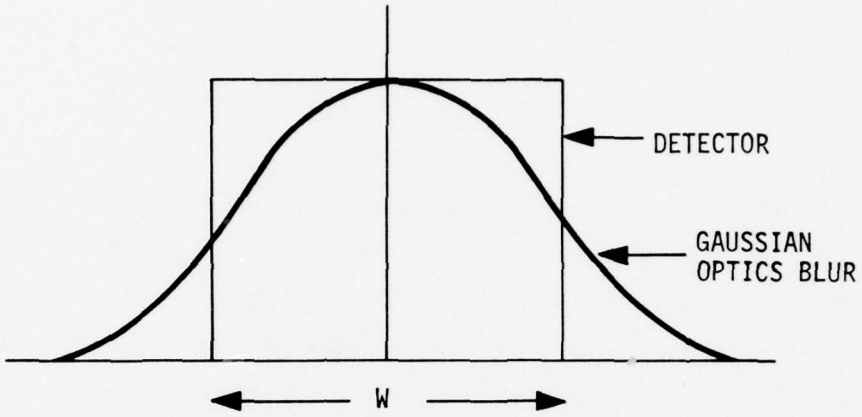


Figure 21. Detector and Optics Blur (Gaussian)
Proportions Assumed

Inverse Filtering

Given an MTF as shown in Figure 22a, we want to get back the attenuated higher spatial frequencies close to the system cutoff (first zero on the MTF). Obviously, the simple inverse filter has the form

$$H'(f) = \frac{1}{H(f)}, \quad H(f) \neq 0$$

where H(f) is the degrading (blur) MTF of the system. The problem with the straight inverse filter is that it boosts the broad band noise at high frequencies where the signal power is essentially zero. Therefore practically all linear restoring filters from Wiener on down have provision for making the inverse filter response small at frequencies where the signal

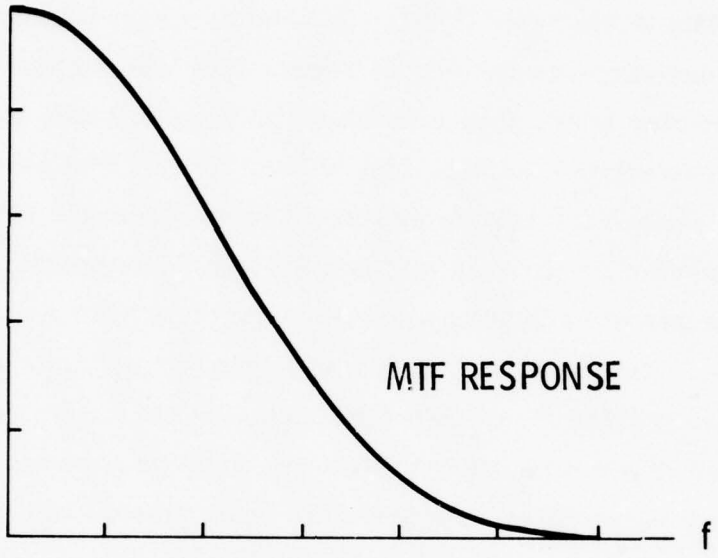


Figure 22a. Combined Optics and Detector MTF

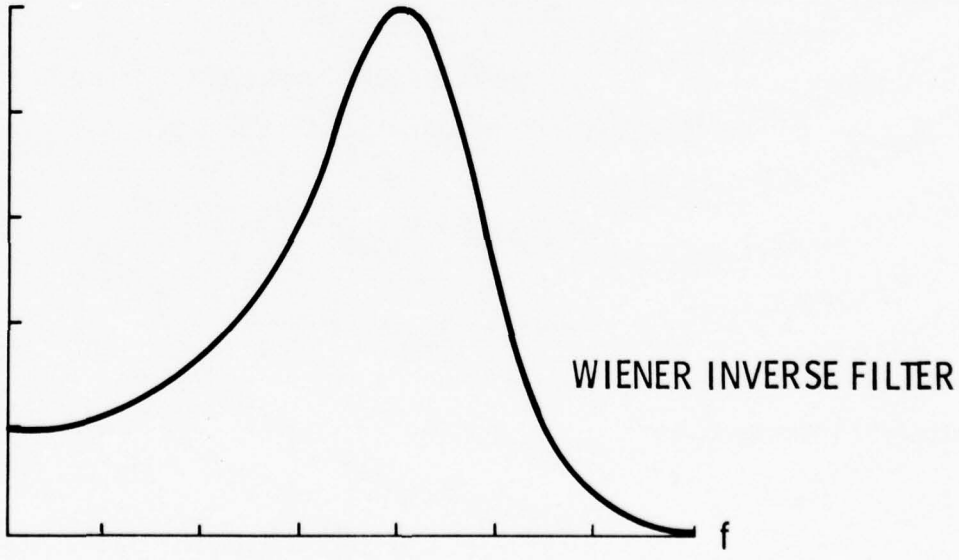


Figure 22b. Wiener Inverse Filter Response

energy is low. The Wiener (MSE) filter and Hunt's Least Square filter were studied on one-dimensional test patterns. For the simple analytic frequency limited blur MTF, they were found to be practically identical, and therefore we simulated the classical Wiener filter. For purposes of restoration, the space variant blur was modeled as piecewise space invariant. The rectangular image area was divided into 16 segments. Each of the four subsegments of each quadrant was assumed to have a space invariant blur function. This turns out to be a very reasonable approximation since there are very little significant differences in the on- and off-axis blurs encountered in practice. Each of the segments was restored individually using the inverse filter calculated for the segment, and resulting segments were mosaiced together to yield the full-frame resolution restored image (see Figure 23).

Wiener Restoration Filter

We take advantage of the fact that the degrading impulse response is separable (the detector rect function and the Gaussian OTF), i.e.,

$$H(f_x, f_y) = H_x(f_x) \cdot H_y(f_y)$$

and

$$h(x, y) = h_x(x) \cdot h_y(y)$$

The straight inverse filter

$$H_I(f_x, f_y) = \frac{1}{H_x(f_x)} \cdot \frac{1}{H_y(f_y)}$$

is also separable. But the two-dimensional Wiener filter,

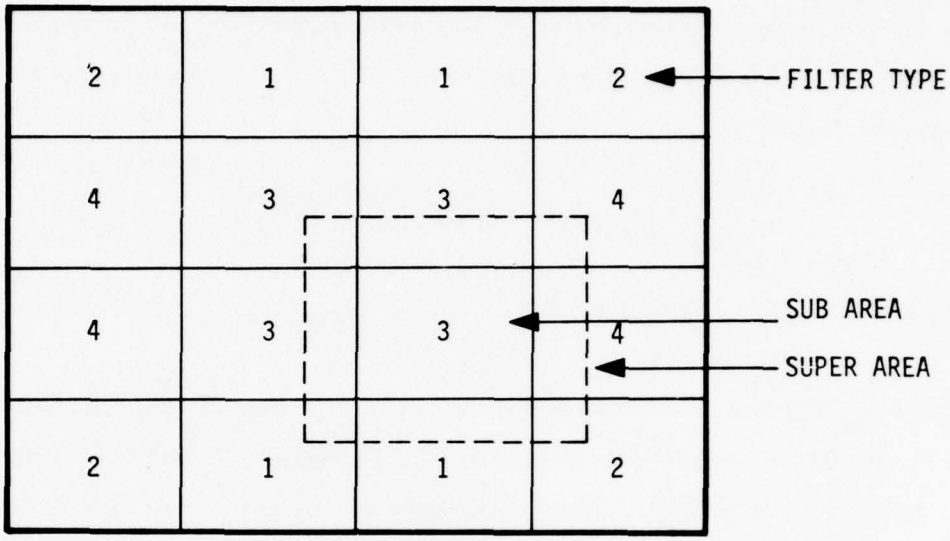


Figure 23. Geometry Used for Restoring the Space Variant Blur by Piecewise Space Invariant Filters

$$R(f_x, f_y) = \frac{H^*(f_x, f_y)}{|H|^2 + \Phi_n(f_x, f_y)/\Phi_o(f_x, f_y)}$$

is not separable, although H itself is separable. But we approximate $R(f_x, f_y)$ by a separable restoring filter:

$$R(f_x, f_y) = R_x(f_x) \cdot R_y(f_y),$$

the one-dimensional Wiener filter,

$$R_x(f_x) = \frac{H_x^*(f_x)}{|H_x(f_x)|^2 + \Phi_{nx}(f_x)/\Phi_{ox}(f_x)}$$

where $H_x(f_x)$ is known from the blur model, $\Phi_{nx}(f_x)$ is the noise power spectrum assumed to be white and the same horizontally and vertically. $\Phi_{ox}(f_x)$ is the object row power spectrum obtained by the periodogram of the rows, i.e.,

$$\Phi_{ox}(f_x) = \frac{1}{Nrow} \sum_{rows} |X(f_x)|^2$$

where $X(f_x)$ is the DFT of each row of the image.

Ideally $\Phi_{ox}(f_x)$ should be the power spectrum of the object, but in practice the image power spectrum is substituted as above. The column quantities are defined analogously. The advantages of using the separable filters are that

- They are easier to implement because the two-dimensional separable filter can be reduced to two successive one-dimensional filtering on rows and then on columns.
- The row and column power spectra are easier to estimate with just one image because the joint two-dimensional power spectrum needs an ensemble of images to get a reasonable estimate.

Figure 22a shows a frequency domain Gaussian MTF (assumed for purposes of simulation to be equal to seven pixels equivalent diameter) of the FLIR image model. The inverse Wiener filter frequency response computed using a simple noise model is shown in Figure 22b. We see that the Wiener filter is simply the inverse filter at frequencies where the object power spectrum is finite and nonzero, and it tends to the blur response (of system MTF) at frequencies where the object power spectrum is null. Figure 24a is a reasonably sharp FLIR image, blurred with the linear space variant OTF

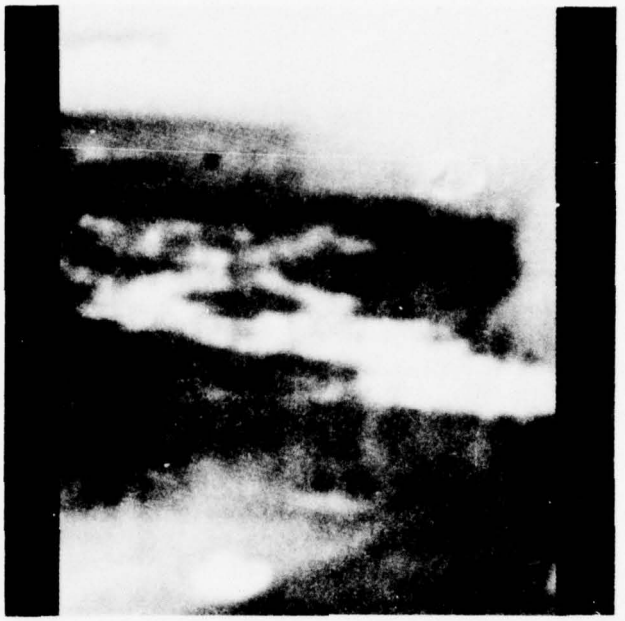


Figure 24a. FLIR Image Blurred with a Space Variant MTF Shown in Figure 22a (Five Pixels Blur On Axis and Seven Off Axis)

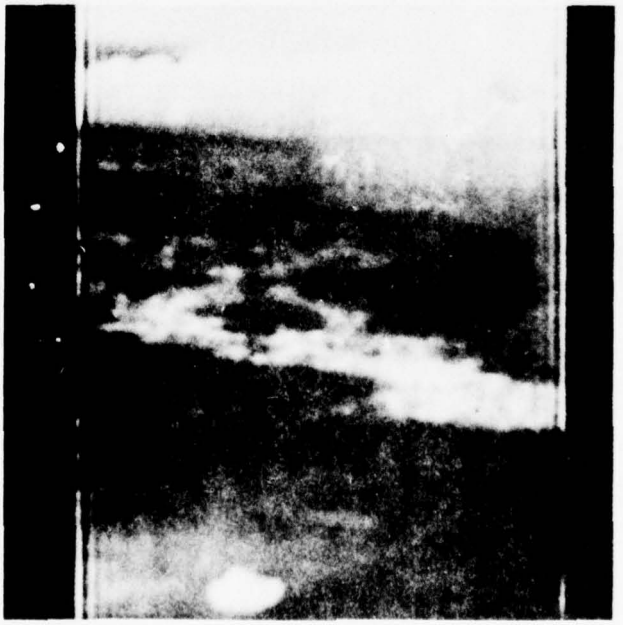


Figure 24b. Restored by Piecewise Space Invariant Wiener Filter

with a large blur circle (1.2 times) off-axis to model the space variant OTF. The on-axis blur circle was assumed to be five pixels in diameter. This image was restored by piecewise linear Wiener filter similar to that shown in Figure 22b, by dividing the image into sixteen segments, each being restored by one of four filters approximating the different blurs from on-axis to off-axis. Figure 24b is the resultant image, and we see that it has been restored adequately.

Preliminary conclusions of the above study are that the optics diffraction blur can be assumed to be space invariant for all practical purposes. Frequency domain inverse filtering is problematical for piecewise linear full-frame focus restoration because the narrow PSF translates into a broad MTF which requires a large area, two-dimensional discrete unitary transform (e.g., Fourier or Hadamard) to implement. Since the PSF is linear shift variant over that area, linear frequency domain inverse filtering is likely to be meaningless. Perhaps the best approach is to implement the inverse filter in a CCD analog spatial filter configuration (similar to the nonrecursive high frequency emphasis filter discussed in the section on contrast enhancement). Further, the space variant blur can be approximated by an average space invariant blur so that one analog filter can be used instead of four, for example. Also, we note that the inverse Wiener filter shape in Figure 22b is very close to the high frequency emphasis filter in Figure 1a. Therefore, simple high frequency emphasis could take the place of adaptive resolution enhancement filtering in a simple image enhancement design.

SUPER-RESOLUTION ALGORITHMS

The linear, frequency domain resolution restoration techniques above can only restore the high frequencies out to the Rayleigh limit--the first zero on the OTF. Theoretically, at least, extrapolation beyond the Rayleigh limit is possible due to the recovered information in the side lobes of the diffraction pattern. Several iterative techniques can use the positivity and boundedness constraints on the filtered output to accomplish this. These schemes are generally termed "super-resolution" algorithms, and the most successful techniques for super-resolution have been iterative processing algorithms.

Two iterative super-resolution algorithms were investigated: the Gradient Projection (GP) algorithm (Huang, et al.) and the Honeywell-developed Stochastic Approximation (SA) algorithm. These algorithms were used to enhance NVL supplied thermoscope and FLIR imagery. Since we are trying to improve the system resolution close to the diffraction limit, the sampling grid should be fine enough so that aliasing of the restored image spectrum does not occur. This implies that the sampling frequency should be at least twice as high as the highest frequency being restored. But these images were not over sampled; i.e., they were sampled approximately once per detector width. This also reflects the general configuration expected in second generation FLIRs. Therefore, the section of the image containing the target in each image was magnified by a factor of two or three by linear interpolation, super-resolved, and displayed in the magnified format. Magnification decreases the sampling grid size and therefore extends the sampling frequency. Unfortunately, magnification by linear interpolation, which was attempted first, proved inadequate.

Figure 25a shows an original one-dimensional spectrum with the foldover at $f_s/2$. Simple linear interpolation has a frequency transfer function shown with the broken curve in Figure 25b. Hence the resultant spectrum of the magnified linear interpolated image will have spurious frequency components corresponding to the aliased part of the original spectrum as shown in Figure 25b. Ideally, we should have an interpolating function that corresponds to a strict low-pass filter with the cutoff at the Nyquist frequency as shown in Figure 25b (dotted line).[1] A Dolph-Chebycheff low-pass filter which is maximally flat in the pass band and equiripple in the stop band, with a limiting attenuation in the stop band of -60 dB, was chosen to perform the required interpolation (Figure 26). The corresponding seven-pixel impulse response was used as the interpolating function. The interpolation was accomplished by linear filtering an image matrix with zeros embedded in the positions to be interpolated. The two-dimensional interpolation was performed by taking advantage of the separability: by one-dimensional interpolation along the rows first, followed by interpolation along the columns.

The magnified and interpolated sections were then super-resolved using the Gradient projection and the Stochastic Approximation algorithms as follows.

---

1. R. W. Schafer and L. R. Rabiner, "A Digital Signal Processing Approach to Interpolation," Proceedings of the IEEE, June 1973.

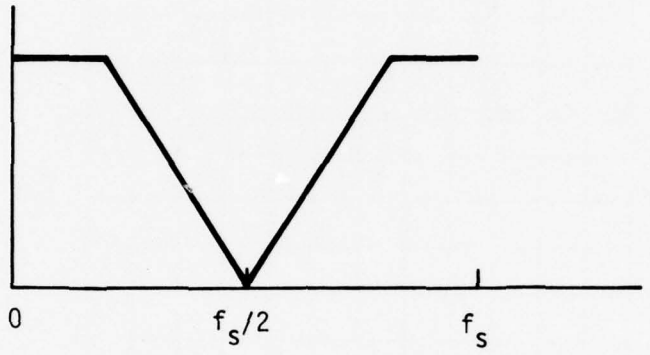


Figure 25a. Original Spectrum

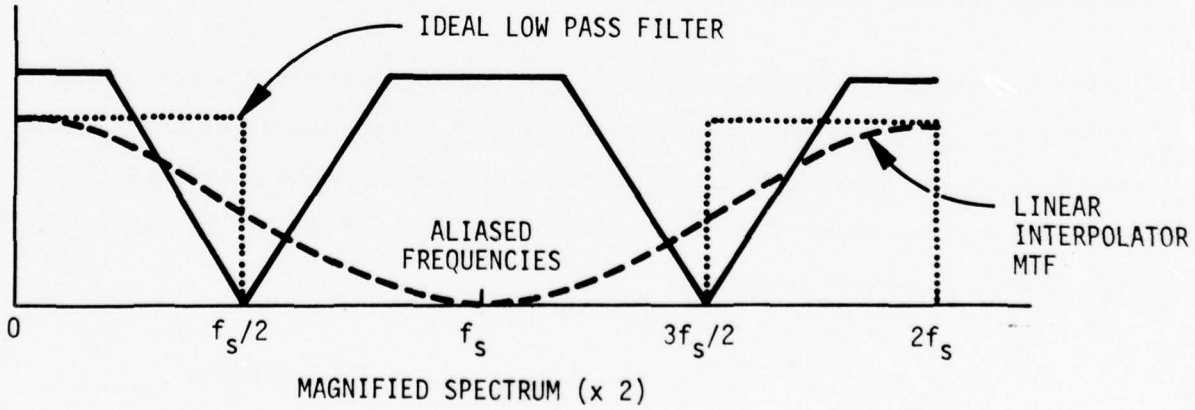


Figure 25b. Magnified Spectrum Showing the Aliased Frequencies Resulting from a Linear Interpolator

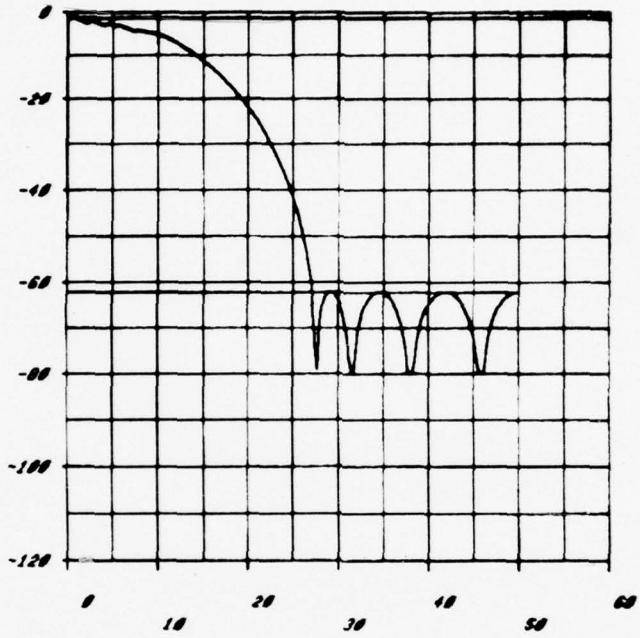


Figure 26. Dolph-Chebycheff Low Pass Filter Used for the Digital Interpolation

The Gradient Projection Algorithm

The GP algorithm developed by Professor T.S. Huang and his coworkers at Purdue University was one of the two super-resolution algorithms coded and evaluated. The algorithm is iterative with successively refined estimates of the object f made at every iteration. At the $K + 1^{st}$ iteration, the estimate is expressed as:

$$f_{k+1}(i+p, j+q) = f_k(i+p, j+q) + [g_M(i, j) - h * f_k(i, j)]\, h(p, q)$$

$$|p|, |q| \leq L$$

where $g_M(i, j)$ is the magnified image and h(m, n) is the L x L blur function. In essence, the algorithm converges when the convolution of the object estimate $f_k(i, j)$ with h coincides with the blurred image $g_m(i, j)$. Each iteration requires first a convolution of h with f_k (N^2L^2 multiplies and adds) and a further N^2L^2 MADs to compute the new $f_{k+1}(i, j)$ as above. The blur function h is assumed to be known a priori. The estimates $f_k(i, j)$ at each iteration are constrained to lie between f_{min} and f_{max} (positivity and boundedness constraints).

Stochastic Approximation Algorithm

This Honeywell-developed super-resolution algorithm is also iterative. The simple iterative formula is as follows:

$$f_{k+1}(i, j) = f_k(i, j) + \mu_k[g_m(i, j) - h*f_k(i, j)]$$

$$\mu_k < 2$$

where μ_k is an iteration constant that is made progressively smaller at every iteration to speed up convergence.

The idea behind the SA algorithm is very simple. At every iteration, the estimate f_{k+1} is updated by a fraction of the difference between the old estimate f_k convolved with h, and the input image. When the two coincide, the algorithm has converged. Again, the positivity and boundedness constraints are applied at each iteration. In contrast to the GP algorithm, the SA algorithm needs just <u>one</u> convolution $h*f_k$, a saving of N^2L^2 MADs per iteration.

Point Spread Function

To apply the above iterative super-resolution algorithms to the magnified and interpolated thermal images, we must first estimate the point spread function of the original optics that degraded the FLIR images that were supplied to us.

Resolution enhancement production runs are underway to enhance 30 FLIR targets for evaluation by imagery statistics and by the Human Factors add-on study. These images are being processed in a manner designed to simulate the scenario in which super-resolution enhancement algorithms could be used. A box in the image containing the target is magnified by two or by three, depending on its size, to fit the working space in the processor; this section is then super-resolved, and the resultant box is reembeded in the background that has also been magnified by the same factor and displayed. Of course, the resultant display will cover only 1/4 or 1/9 the original scene. In each case a guess at the PSF is made a priori and supplied to the algorithm.

Two PSF shapes, a rectangular PSF of width W and a triangular PSF of equivalent width W, were tested and were found equivalent. In keeping with current generation FLIR optics, we also hypothesized blurs W of three, five, and seven pixels on the magnified images. This corresponds to an optics blur of approximately 1.5 pixels (detector width) diameter before magnification.

Figure 27a is a 2 x 1 magnified FLIR image, magnified by bilinear interpolation, and the aliasing effect of the bilinear interpolation results in the spotty edges around the target. The SA and GP algorithm applied to this

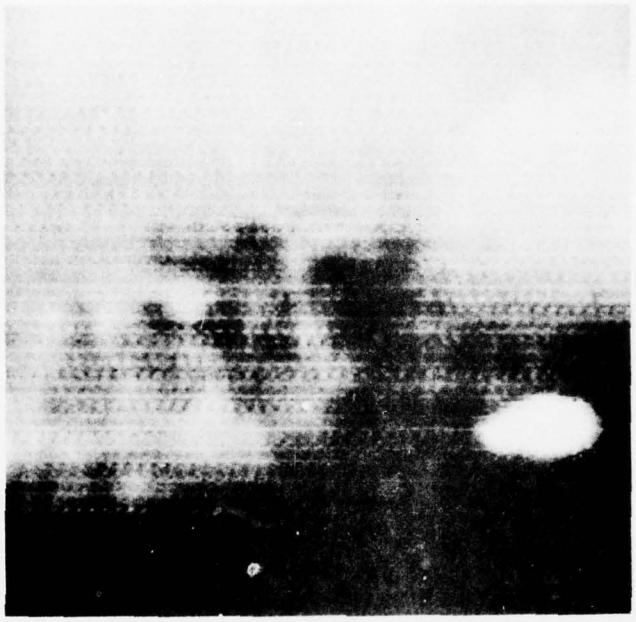


Figure 27a. Thermal Image Magnified by Bilinear Interpolation

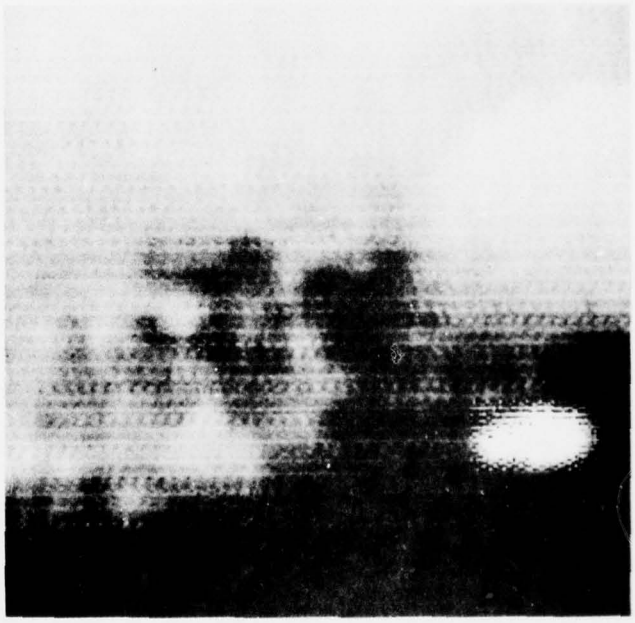


Figure 27b. Gradient Projection Algorithm Applied to the Linear Interpolated Target

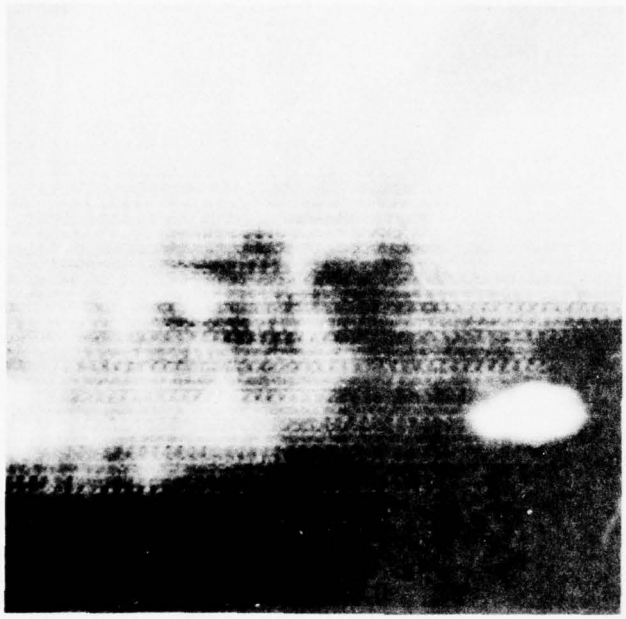


Figure 27c. GP Algorithm Applied to Chebycheff Interpolated Target

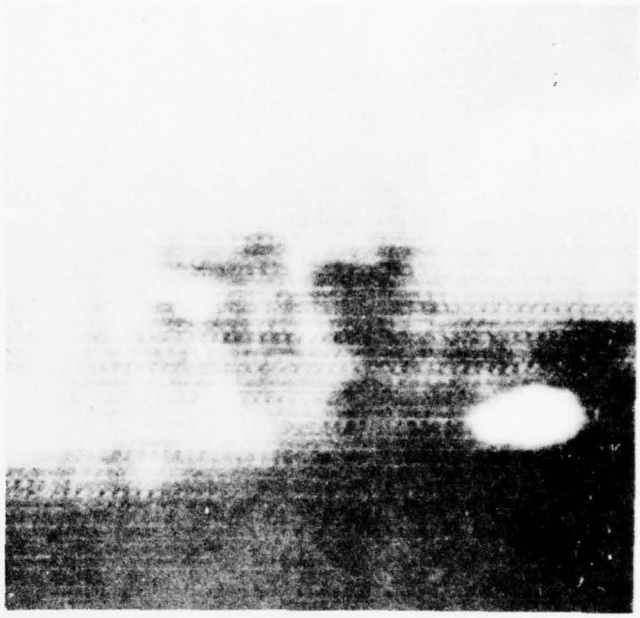


Figure 27d. SA Algorithm Applied to Chebycheff Interpolated Target

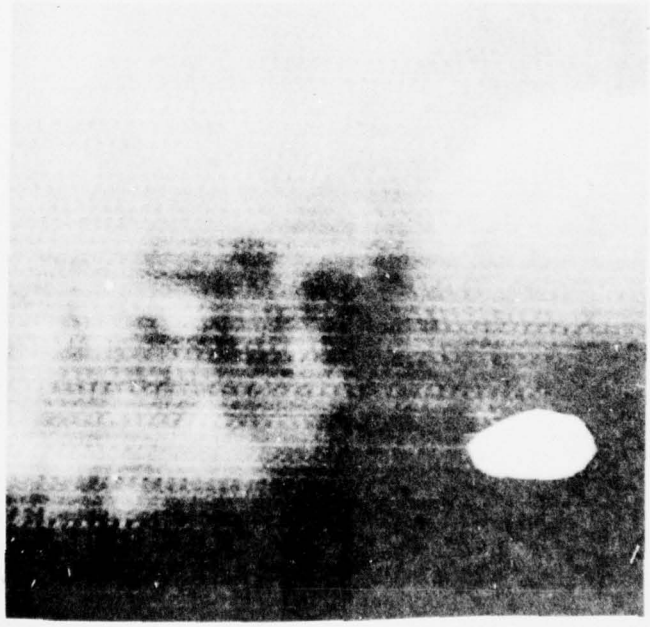


Figure 27e. Resolution Restoration Algorithms Failed (Too Large a PSF, = 5 Pixels was Assumed)

interpolated image resulted in Figure 27b where the deleterious effects of aliasing are even more apparent. By substituting the Dolph-Chebycheff low pass filter for the bilinear interpolation around the target area, better results were obtained (see Figures 27c and 27d for an assumed triangular PSF of width = 3). The resultant images show better details within the target on a CRT display, but because of saturation in the reproduction process used here, they do not appear very prominent. When a larger PSF (= 5) was assumed with the SA and GP algorithms, Figure 27e resulted. We see that all points within the target saturate to the maximum allowable value of the f_{max}. Therefore, selection of the exact size of the PSF is crucial.

We find that the SA and GP algorithms are roughly equivalent (although the SA algorithm appears to perform slightly better) in effectiveness, but the SA algorithm is easier to implement in a CCD processor because it requires one convolution versus two for the GP algorithm. These algorithms are iterative and therefore lack real-time capability. But one scenario that envisages their use in tactical imagers would be as follows. The operator cues a target area, which is either sampled at a higher rate or interpolated in two dimensions to obtain a finer grid to resolve the target. This area is then super-resolved iteratively off-line and displayed magnified to the operator either on a split screen or separate display. Because of the amenability to CCD processing, these algorithms can be implemented in near real time. The SA algorithm, for example, takes less than ten iterations to converge--at normal TV rates used for clocking the CCDs (30 cycles per second) this would mean 1/3 of a second--which is near enough real-time to be useful. This has been simulated at Honeywell, with FLIR imagery already blurred from the optics, by guessing at the point-spread

function of the original optics. These results indicate that, when augmented with automatic target cueing, iterative resolution enhancement schemes would indeed be very useful for improving operator performance under tactical situations.

SECTION V

STATISTICAL ANALYSIS OF IMAGERY

Statistical analysis of FLIR imagery is being performed to better quantify the characteristics of thermal imagery and to provide quantifiable measures of the quality of enhancement processes. The statistics being extracted can be categorized as shape, intensity, and texture features. These measures being made on the FLIR imagery were described in the previous interim report.

Following is a list of the measures:

1. Shape Statistics (on target and target-like objects)
   a. Perimeter/Area
   b. Number of edges
   c. Histogram of the edge lengths (normalized by the perimeter)
   d. Histogram of differential slopes of successive edges

2. Contrast Measures (on target and its background)
   a. Intensity histogram of target
   b. Intensity histogram of its background
   c. Average contrast of target/background
   d. Peak contrast of target/background

e. Histogram of Soebel gradient operator on the edges

f. Histogram of the gradient across the edge--edge contrast measures

3. Texture Features

Gray-level difference histograms[2] in four principal directions, and measures computed from these histograms. Texture features are measured on target and background.

These features are being measured on the original and enhanced FLIR imagery. A repertoire of analysis and display tools has been developed to analyze the statistics. They are histogram routines to generate and display the distribution of the statistics, and two-dimensional scatter plots to display the cluster plots of the features taken one pair at a time. The scatter plot programs are interactive and versatile enough to plot and display any pair of up to 80 features on an ensemble of 120 targets with annotated symbols delineating the target and background classes. Figures 28 and 29 show examples of the scatter plots of the brightness statistics: the mean vs. standard deviation, and skewness vs. the excess of the target and background intensities over an ensemble of 10 targets and their backgrounds. It is interesting to note that there is considerable separation of the background and target classes, not only in the mean and standard deviation of their intensities, but also in the skewness and excess (corresponding

---

[2] Wezka, J. S., Dyer, C. R. and Rosenfeld, A., "A Comparative Study of Texture Measures for Terrain Classification," IEEE Trans. Systems, Man, and Cybernetics, Vol. 6, April 1976.

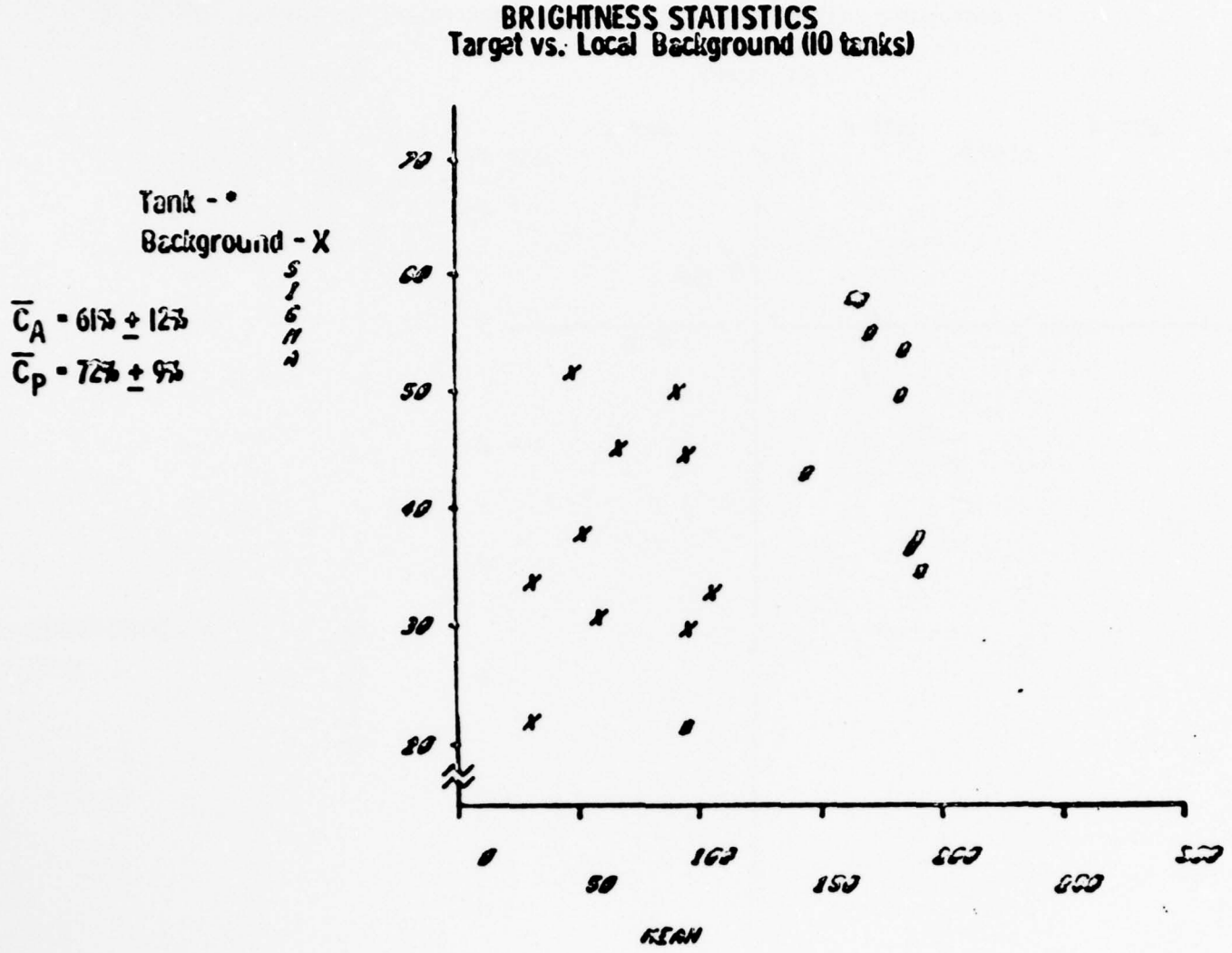


Figure 28. Mean vs. Standard Deviation of Brightness on Target and Background

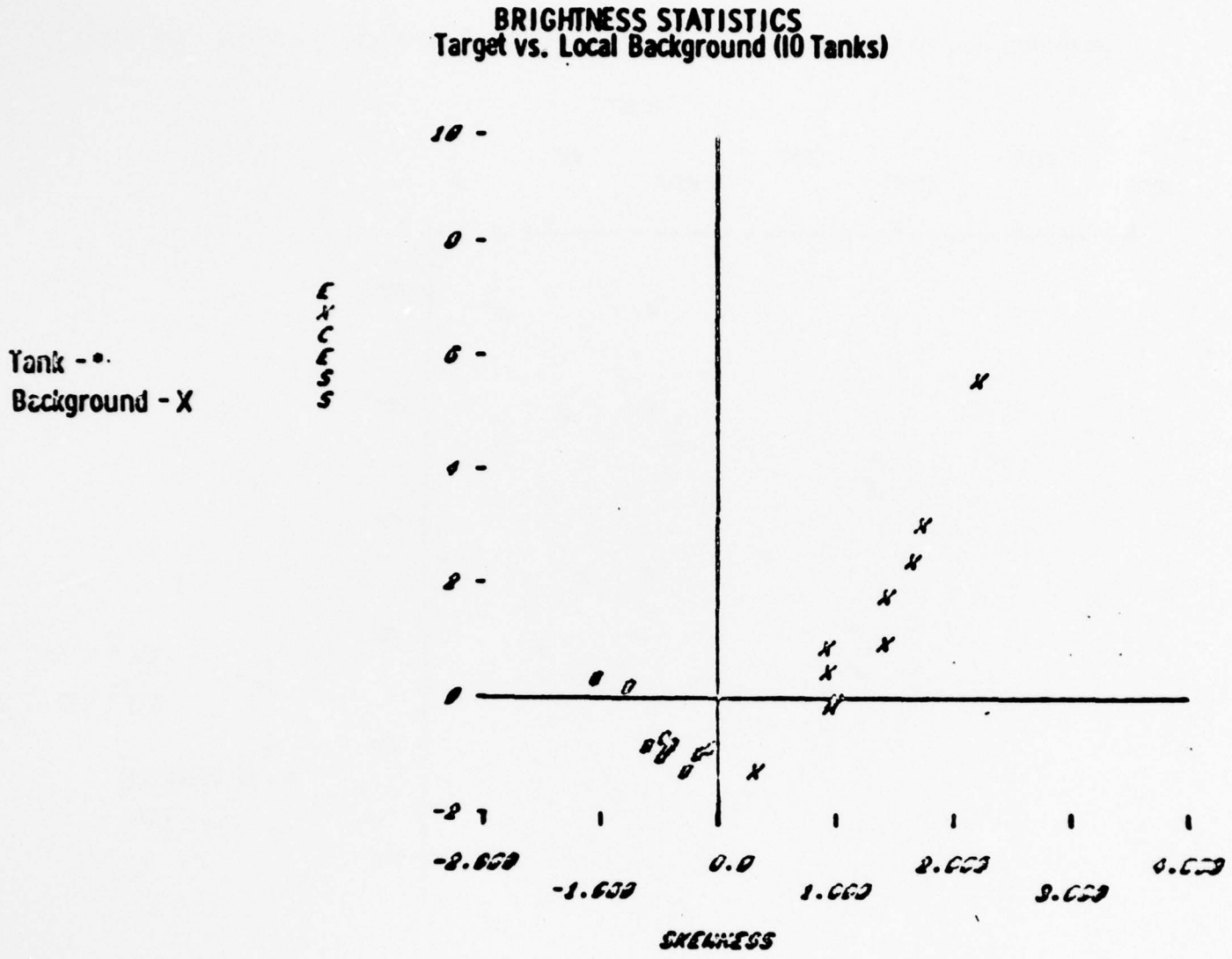


Figure 29. Skewness vs. Excess of Brightness Statistics

to higher moments of the intensity distribution). As the statistics of the original and enhanced images are accumulated, these cluster plots will enable us to display and analyze how these image quality features are transformed through the image enhancement. For example, the contrast enhanced (CE) images will show a greater separation in the normalized intensities between target and background; the high frequency emphasis and resolution restoration of images will be reflected in the plots of edge contrast measures, et cetera.

SECTION VI

AN INTEGRATED APPROACH TO IMAGE ENHANCEMENT*

PROBLEM STATEMENT

We present an integrated approach to image enhancement which aims to achieve simultaneously the following goals: noise reduction, deblurring, contrast enhancement, and crispening. A specific implementation of this integrated approach was simulated on the computer and applied to a noisy blurred FLIR image.

AN INTEGRATED APPROACH TO IMAGE ENHANCEMENT

A general block diagram of an integrated system for image enhancement is shown in Figure 30. The function $f(x, y)$ represents the noisy blurred image, and $\hat{f}(x, y)$ represents the enhanced image, where x and y are spatial coordinates. The processing proceeds in two parallel channels.

In the upper channel we have a low-pass filter with frequency response $H_1(u, v)$, where u and v are spatial frequencies. The output of this low-pass filter contains the slowing-varying background of the image where most of the noise is smoothed out. The contrast enhancer C_1 could be just a global histogram modifier.

---

*This work was performed by Professor T. S. Huang of Purdue University under subcontract to Honeywell Systems and Research Center.

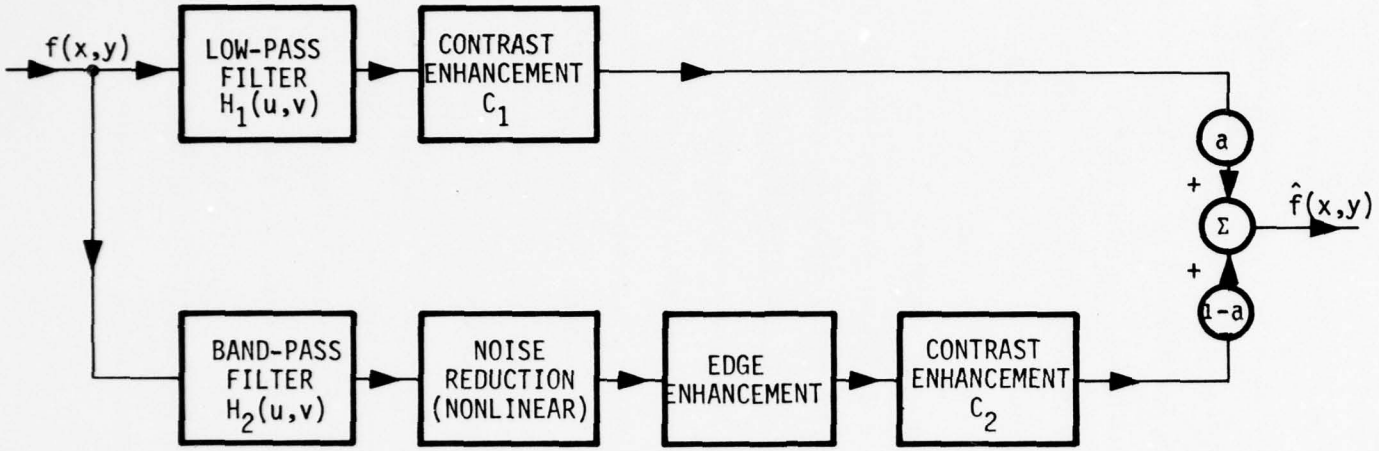


Figure 30. An Integrated Image Enhancement System

In the lower channel we have a band-pass filter with frequency response $H_2(u, v)$. If we ignore all the boxes in the system except H_1 and H_2, the total system frequency response is $H = aH_1 + (1-a)H_2$, where a is a constant. We can choose H_1 and H_2 such that $H(u, v)$ is equal to the frequency response of some desired inverse filters for deblurring. The output of the band-pass filter contains mainly edges and noise. A nonlinear noise reducer eliminates most of the noise but preserves most of the edges. In simple cases, this would be a thresholding operation. Everything above the threshold is preserved, whereas everything below the threshold is set to zero. The edge enhancer reconditions the edges, for example, by making them sharper. The contrast enhancer C_2 could be either global or locally adaptive and could be used, for example, to bring up the weaker edges. Recent work of Dr. Schreiber at MIT (private communications) indicates that the amount of edge enhancement should be roughly proportional to local brightness and inversely proportional to local contrast. Therefore, one possibility for C_2 is to apply a global brightness transfer curve in the shape of a square-root function and to multiply each picture element by a number roughly proportional to the brightness value of the corresponding picture element in the low-passed version of $f(x, y)$.

The final output $f(x, y)$ is a weighted sum of the response from the upper and lower channels. The main motivation behind the two-channel approach is as follows. Low pass filtering provides an easy way of smoothing out noise. However, it blurs the edges in the image. Therefore, the lower channel is created to extract edges from the noisy blurred image and add them to the low-passed image to get the final results.

Note that, in the system of Figure 30, deblurring is done by making $aH_1 + (1-a)H_2$ equal to the frequency response of some desired inverse filter, noise reduction is done by the low-pass filter and the nonlinear noise reducer in the lower channel, contrast enhancement is done by C_1 and C_2, and crispening is done by C_2 and the edge enhancer.

A SPECIFIC EXAMPLE

A specific implementation of the integrated image enhancement system of Figure 30 was simulated on a PDP-11/70 computer and applied to a noisy blurred FLIR image shown in Figure 31 (512 x 512 samples, 8 bits per sample). The noise in the image is signal-dependent, and the average signal-to-noise ratio is approximately 10.

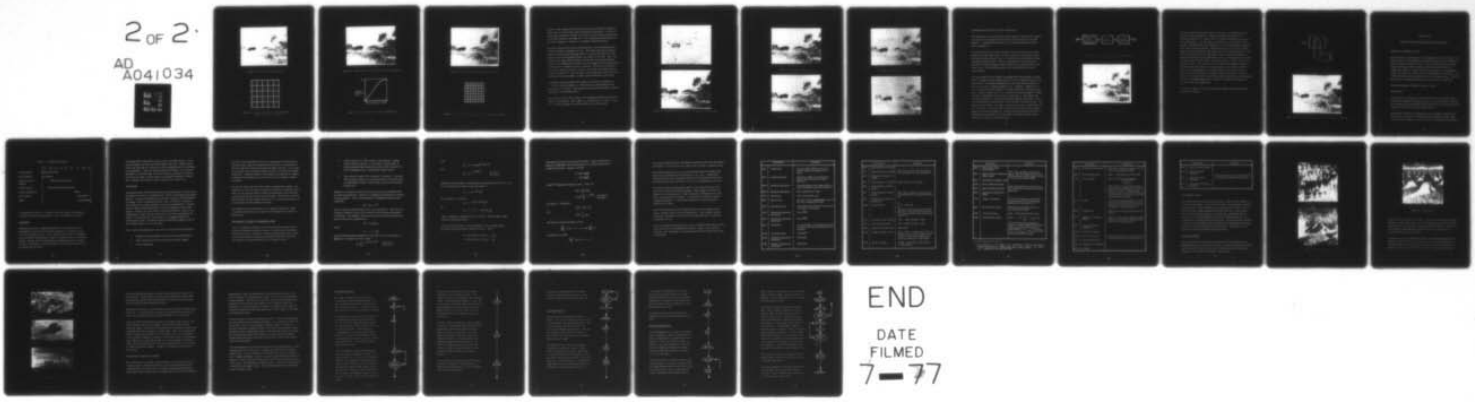
The low pass filter H_1 was implemented by a spatial convolution with a 5 x 5 window. The relative weights of the coefficients is the window shown in Figure 32. The low-passed image is shown in Figure 33. The contrast enhancer C_1 was a nonlinear memoryless spatially-invariant operation with a piecewise linear transfer curve (Figure 34) which aimed to stretch the brightness range of the low-brightness areas in the image. The output of C_1 is shown in Figure 35.

The band-pass filter was realized by a smoothed gradient operation. At each picture point (pel), we look at a 7 x 7 window centered at the given pel (Figure 36). The smoothed gradient at the pel X is given by

$$g_x = |a-b| + |c-d|$$

AD-A041 034 HONEYWELL INC MINNEAPOLIS MINN SYSTEMS AND RESEARCH --ETC F/G 17/5
AUTOMATED IMAGE ENHANCEMENT TECHNIQUES FOR SECOND GENERATION FL--ETC(U)
MAY 77 P M NARENDRA, L G WILLIAMS DAAG53-76-C-0195
UNCLASSIFIED 77SRC49 NL

2 OF 2
AD A041034



END

DATE FILMED
7-77

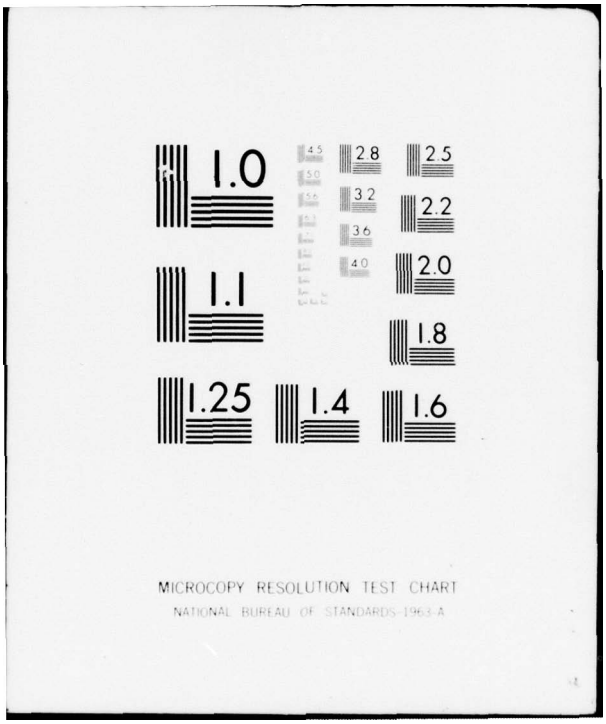

1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Figure 31. Noisy Blurred FLIR Image

.5	.5	.5	.5	.5
.5	.75	.75	.75	.5
.5	.75	1	.75	.5
.5	.75	.75	.75	.5
.5	.5	.5	.5	.5

Figure 32. Relative Weights of Impulse Response Coefficient of Low-Pass Filter

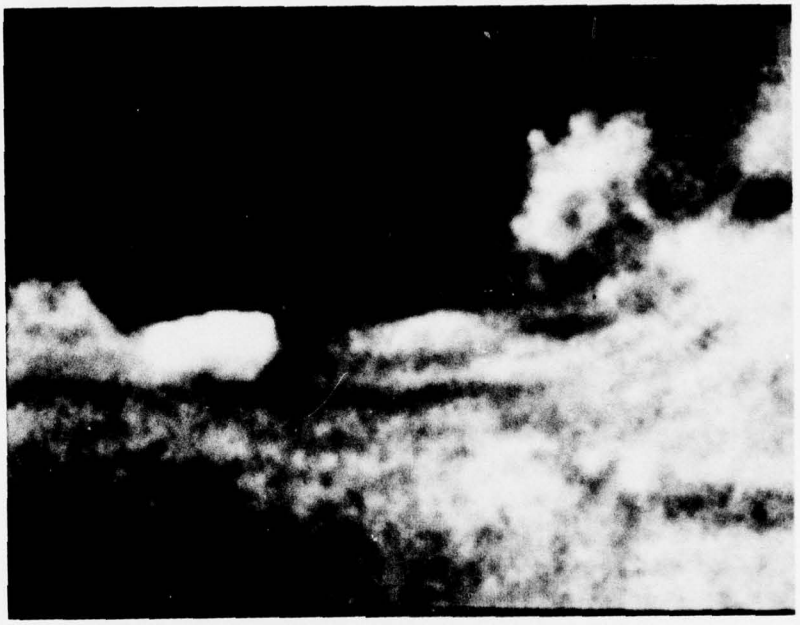


Figure 33. Low-Passed Version of Image Found in Figure 31

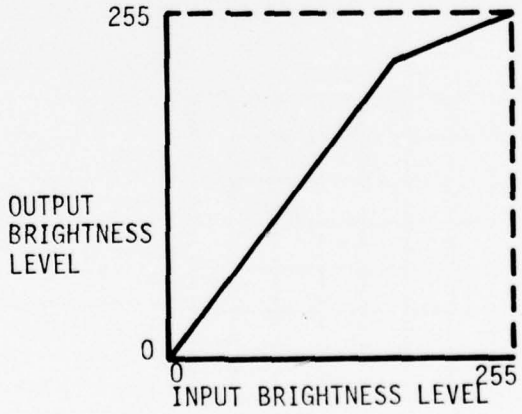


Figure 34. Transfer Curve of Contrast Enhancer C_1

Figure 35. Contrast Stretched Version of Figure 33

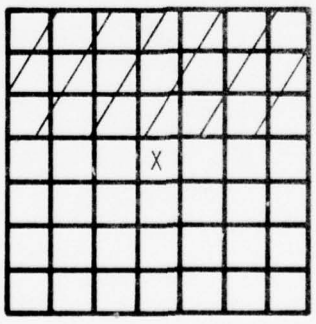


Figure 36. A 7 x 7 Window for Calculating the Smoothed Gradient

where a is the average of the 21 pels in the three rows above X (shaded in Figure 36), b is that of the 21 pels in the three rows below X, c is that of the 21 pels in the three columns to the left of X, and d is that of the 21 pels in the three columns to the right of X. This smoothed gradient operator is, of course, nonlinear. However, it resembles a band-pass filter.

The noise reduction was done by setting all pels in the smoothed gradient, which were below a threshold, to zero. The output of the noise reduction process contains mainly the edges of f(x, y). Due to the large size of the window used in the smoothed gradient, the edges are too wide. They were thinned by the edge enhancer which was basically a local gain controller. Each pel was multiplied by a number that was roughly inversely proportional to the local (7 x 7 window) variance. However, if the local variance was very small, nothing was done to the pel. The output of the edge enhancer is shown in Figure 37. This local gain control operation in fact also served to bring up the weaker edges; no other contrast enhancer C_2 was used.

A weighted sum of the outputs of the upper and the lower channels with $\alpha = 0.875$ is shown in Figure 38. (The picture was obtained by filming the display on a Ramtek. Unfortunately the quality of the image is severely degraded by the filming process.)

The results for some other values of the weighting coefficient are shown in Figures 39 through 42. The weights are, respectively, 9, 0.85, 0.8, and 0.7. For these images, the contrast enhancer C_1 was not used.

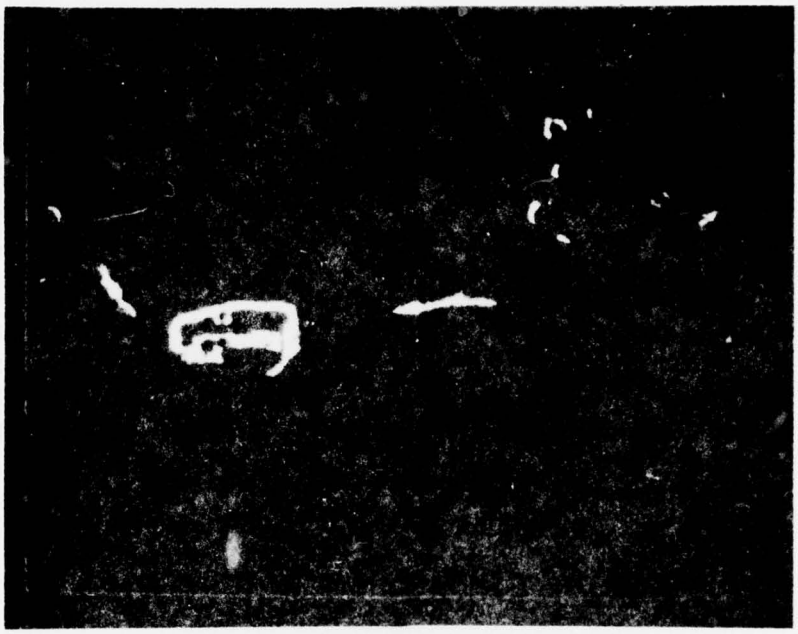


Figure 37. Output of Edge Enhancer

Figure 38. Weighted Sum of Low Passes and Edge Enhancer Outputs

Figure 39. Enhanced with $\alpha = 0.9$

Figure 40. Enhanced with $\alpha = 0.85$

Figure 41. Enhanced with $\alpha = 0.8$

Figure 42. Enhanced with $\alpha = 0.7$

COMPARISON WITH THE CASCADE APPROACH

The advantages of the two-channel parallel approach cannot be fully appreciated unless one investigates the alternative, more conventional cascade approach. A typical block diagram of a cascade enhancement system is shown in Figure 43.

A major difficulty here is how to design an inverse filter that will smooth out noise but keep edges sharp. A linear spatially-invariant filter trades off image sharpness against noise; it cannot do noise smoothing and image sharpening simultaneously. We have experimented with several linear spatially-varying and nonlinear filters which aim to smooth noise but keep edges sharp. We shall present below some results (which are rather disappointing).

We experimented with an adaptive smoothing filter whose impulse response width is roughly inversely proportional to the local smoothed gradient of the input image. To simplify implementation, only four different impulse responses were used. Impulse response #1 is an impulse. Impulse #2 is a 3 x 3 window with relative weights 1 at center and 0.25 for the remaining 8 pels. Impulse response #4 is a 5 x 5 window with relative weights as shown in Figure 32 except that 0.25 was used instead of 0.5. The range of the smoothed gradient (5 x 5 window) of the input image is normalized to 0-255 and divided into four subdivisions: 0-180, 181-220, 221-250, 251-255. For these four subdivisions, the impulse responses used were, respectively, #4, #3, #2, and #1. The result of this adaptive filtering of the FLIR image of Figure 31 is shown in Figure 44. We observed that indeed the noise in the background was smoothed out and the edges were kept relatively sharp. However, the edges are extremely noisy.

Figure 43. A Cascade Image Enhancement System

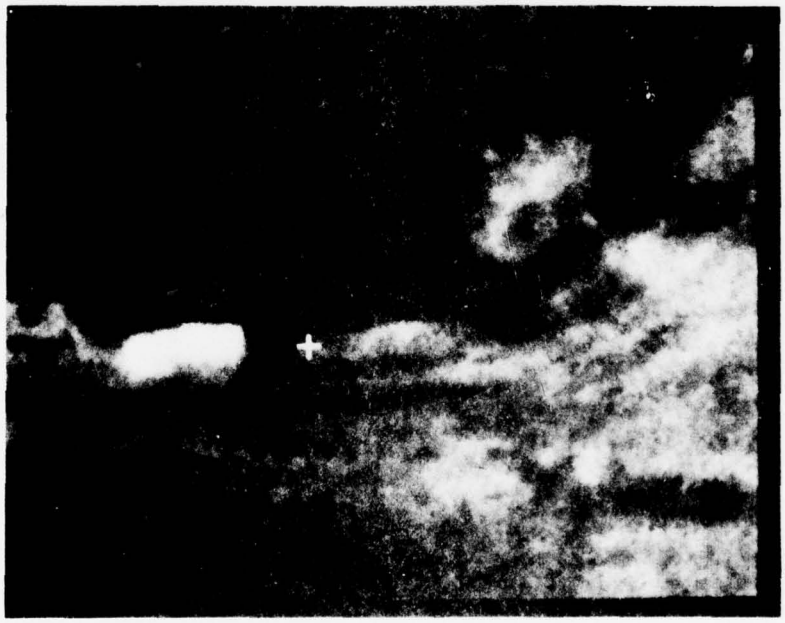


Figure 44. Result of Adaptive Filtering H2

In order to obtain sharp edges which are not noisy, we experimented with the idea of edge interpolation. Regions in the input image where the smoothed gradient (5 x 5 window) exceed a specified threshold were considered edge regions. The remainder of the image was considered the background. The background was smoothed by a 5 x 5 window (impulse response #4 of the preceding paragraph). When the window covered part of an edge region, only pels outside the edge region were used in the convolution. Then the values of the pels in the edge regions were obtained by interpolation. Figure 45 illustrates the interpolation procedure. The edge region is shaded. To calculate the value of a pel X, in the edge region, we look at the 15 x 15 window around X. The value of X was set equal to the weighted average of the pels in the window but not in the edge region. The weight attached to a pel is inversely proportional to the Hamming distance from that pel to X. The result of the background smoothing and edge interpolation is shown in Figure 46. We observe that now we have indeed achieved the aim of smoothing out noise and keeping edges sharp. However, the edges are distorted in that the corners are rounded off.

In summary, then, we have not succeeded in finding a good cascade image enhancement system.

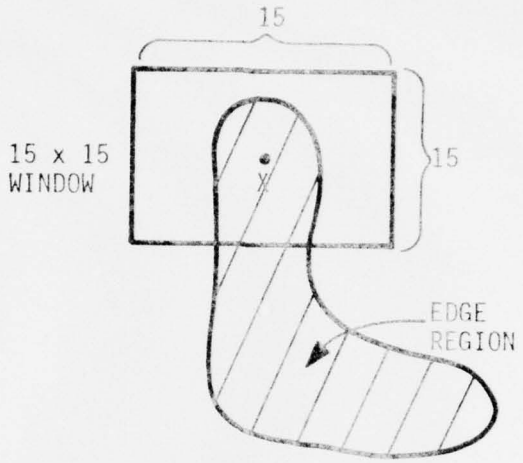


Figure 45. Edge Interpolation

Figure 46. Result of Background Smoothing and Edge Interpolation

SECTION VII

SEARCH EFFECTIVENESS MODEL DEVELOPMENT

SUMMARY--CURRENT STATUS

We have defined the approach for the development of the search model. Our approach places major emphasis on the partitioning of the search field into different regions of uniform texture. The framework focuses on finding targets within these specific regions and then treating the field as a composite of regions. We have defined the stimulus set, the background scenes, the target set, and the means for producing the search fields. The targets and scenes have been digitized and are on magnetic tape. A part of the software for embedding targets has been tested and is operating.

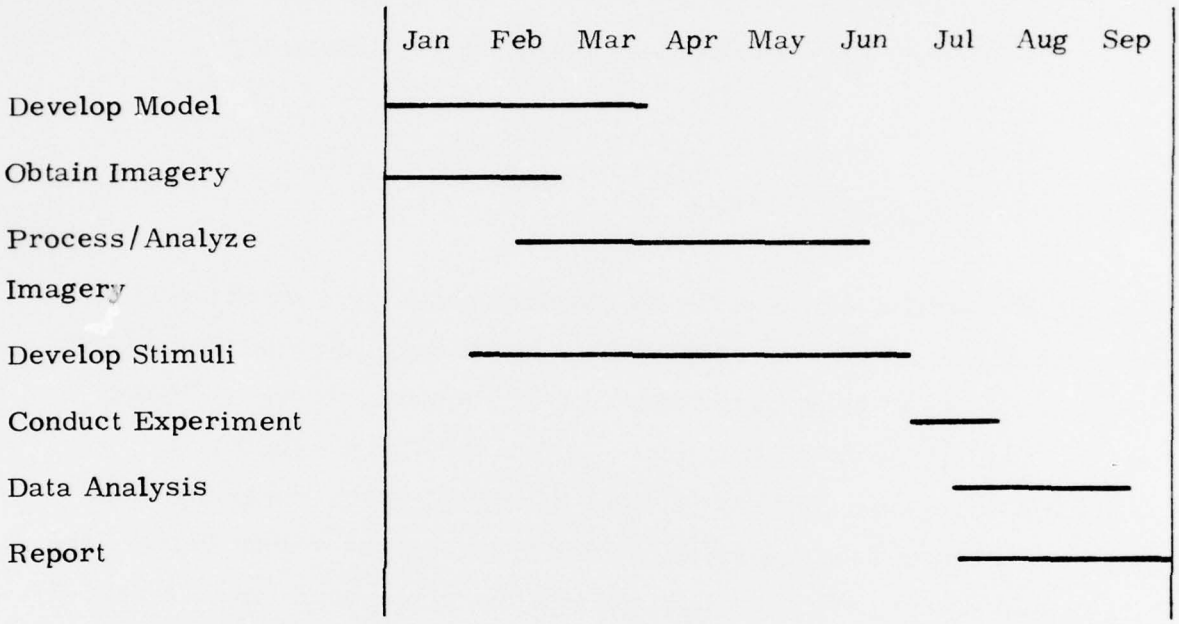
The present program schedule is shown in Table 1.

OBJECTIVES

The objective is to develop a model and procedure for predicting search performance using electro-optical sensors in tactical situations. We want to go further than current efforts which simply take into account target size, contrast, and search area, with limited success.

Specifically, we know that a large part of the variability in performance is due to characteristics of the target and the background, size, shape, texture, internal details, and so on. We want to incorporate these factors

TABLE 1. PROGRAM SCHEDULE



in our predictive system. In addition, since the quality of the image has an important effect on performance, it must be included in the model.

Constraints

Someday we may have a model which predicts to all situations we can contemplate, but for now we must limit ourselves in a number of ways. First, the situation is static. Over a period of time there is no change in the search field--no perspective change, no target movement, no distance change. Second, there is no potential for changing the field; the target is always in the field or, for some purposes, the target is out of view and will never come into view.

The target will be relatively free from contact with other objects. It will not be occluded, nor will it abut other things in the field. Another limitation is that the fields are achromatic with no immediate plans for extending to colored fields. There are presently other restrictions. It is assumed that there is no pre-briefing. Each search field is assumed to contain no more than one target, and the targets are assumed to belong to a limited class, namely, small military objects (trucks, guns, vans, etc.).

THE MODEL

The general nature of the search process is well understood. We have often described it in terms of a combination of a central and peripheral perceptual process. An object is detected extra-foveally; the gaze is directed toward that object, and it is examined foveally. Thus, to construct a model we must predict extra-foveal conspicuity and foveal recognizability. However, there is another consideration which is often recognized but universally ignored. Namely, the search field is divided into different, more or less homogeneous, regions with the conspicuity of the target depending on the region. For the observer, each region has an a priori probability of containing the target. The observer then decides which regions to search first and how long to stay in each region.

Thus, when searching a given scene, there are three primary processes:

1. Strategy in selecting region to search and remaining within it
2. Extra-foveal detection and discrimination within a region
3. Foveal recognition

Our view is that the region effects have a large impact on performance. We believe that target and background characteristics are significant in that they affect both selection strategy and target visibility. In any event, we have some intuitive feelings about the process, but obviously nothing quantitative. During this phase of the program our intention is to begin a mathematical development, but not to carry out any supportive experimentation.

For now we shall study search for targets in homogeneous regions. By doing this we will have removed selection strategy as a significant factor. Thus, we can study the effects of target and background characteristics uncontaminated by this facet of search strategy. We feel that we must be able to predict for this simpler situation if we are to have any hope at all of predicting to the more general situation.

The problem then is to determine, for a range of backgrounds, which measurable characteristics of targets and backgrounds are related to performance.

Searching for a Target in a Homogeneous Field

We first attempted to develop a new statistical math model for search based upon continuous Markov processes. In particular, the model for Brownian motion was examined. Notwithstanding the difficulties associated with the solution of complicated stochastic differential equations, this approach was found deficient for two essential reasons:

1. Visual search is not really random in the sense of a random walk or Brownian motion. If we assume a random walk, we come up against experimental data which clearly contradict the basic assumptions (e.g., equiprobable "jump" states).

2. This cannot be repaired by assuming a "force field." We would then encounter velocity and acceleration effects in contradiction to experimental evidence, which shows a more or less constant fixation rate.

However, an extensive survey of the literature on search models has led us to the conclusion that a simple and straightforward approach is adequate for our purposes. That is, the probability of finding the target by time t is given by

$$p(t) = K(1 - e^{-\lambda t})$$

where λ and K are both functions of the scene parameters. This was concluded by noting that almost all the models surveyed reduced essentially to this form. For example, the General Research Corporation Model A is given by (with some simplification)

$$p(t) = P_1 P_2 P_3 P_4$$

where

$$P_1 = 1 - (1 - P_g)^n$$

P_g is the single glimpse probability of detection, and n is the number of glimpses in a single time frame, where

$$P_2 = \begin{cases} 1 - e^{-(N-3.2/11)^2} & \text{for } N \geq 3.2 \\ 0 & \text{for } N < 3.2 \end{cases}$$

and

$$P_3 = 1/1 + (M/29T^{0.93})^{1.29}$$

and

$$P_4 = \begin{cases} 1 - e^{-(S/N-1)} & \text{for } S/N \geq 1 \\ 0 & \text{for } S/N < 1 \end{cases}$$

Ignoring, for this purpose, the definitions of the parameters N, M, T, and S, we note that we can rewrite these equations as

$$P_1 = 1 - e^{n\ 1n(1-P_g)}$$

Or, setting nT = t, we have

$$P_1 = 1 - e^{t[1/T\ 1n(1-P_g)]}$$

or

$$P_1 = 1 - e^{-\lambda t},\ \lambda = -\frac{1}{T} 1n(1-P_g)$$

and $K = P_2P_3P_4$ is a function of N, M, T and S. Thus, the GRC model reduces to the standard form.

To use this form of p(t), we must determine, for each experimental parameter set (each scene), values for K and λ. Thus,

$$K = F(p) \text{ where } p = (p_1, p_2, \ldots, p_m)$$

$$\lambda = G(q) \text{ where } q = (q_1, q_2, \ldots, q_n)$$

and where $\{p_i\} \cap \{q_j\}$ is not necessarily empty. These systems will, in general, be nonlinear and overdetermined so we cannot solve them in closed form directly. Instead, we define

$$\epsilon^2 = \overline{[K - F(p)]}^2$$

$$\mu = \overline{[\lambda - G(q)]}^2$$

where $\overline{[\ .\ ]}$ denotes the expected value. Thus, for

$$F(p) = \prod_{i=1}^{m} F_i(p_i)$$

$$F_i(p_i) = \begin{cases} 1 - e^{-f_i(p_i)} & \text{, for } f_i(p_i) \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

for $f_i(p_i) \geq 0$. Otherwise,

$$f_i(p_i) = A_i p_i + B_i$$

and

$$G(q) = \sum_{i=1}^{n} a_i q_i$$

We solve the derived nonlinear system

$$\frac{\delta \epsilon^2}{\delta A_1} = 0 \text{ for } i = 1, \ldots, m \text{ and } \frac{\delta \epsilon^2}{\delta B_i} = 0$$

and the linear system

$$\frac{\delta \mu^2}{\delta a_i} = 0 \text{ for } i = 1, \ldots, n$$

It may be that these forms for the unknown functions of the scene parameters will suffice, or we may have to try other forms to obtain an adequate model.

The systems themselves are to be solved using the Fletcher-Powell algorithm (or its equivalent) for the constants A_i, B_i, and a_i. Splitting the experimental data into two groups allows us to use one group to derive the constants and the other to test their accuracy in modeling the search task.

The specific target-background parameters to be used will be extracted from the set listed on the following pages. Some changes in the list are expected as work progresses. Linear regression techniques will be used much in the manner that "features" are determined in automatic pattern recognition problems. A three-letter mnemonic is used to name the measures to be taken.

Many of the following measures depend on the values of additional parameters relating mainly to the surrounding area. For example, contrast calculations will depend on the size of the background area for which the calculation is made.

In the description of the parameters, "mean" represents the ordinary arithmetic mean of the corresponding distribution, "sigma" is the standard deviation, and "skewedness" is the fourth moment minus three times sigma to the fourth.

Parameters		Definition
TAR	Target Area	Area of target defined in terms of number of picture elements in the scene.
PER	Target Perimeter	Perimeter length of the target (the edge is extracted using the Sobel gradient).
MXD	Maximum Dimension	Long dimension of the target defined in terms of number of picture elements.
MND	Minimum Dimension	Short dimension of target.
FAR	Field Area	Total search field area.
SAR	Search Area	The area of that "homogeneous" part of the field containing the target.
NAR	Normalized Area	TAR/SAR that proportion of the search area which is the target.
NXD	Normalized Maximum Dimension	$MXD/\sqrt{SAR}$
NND	Normalized Minimum Dimension	$MND/\sqrt{SAR}$
RES	Resolution	The resolution of the display measured in terms of the smallest visible acuity elements.
ARR	Area/Resolution	TAR/RES^2
XDR	Maximum Dimension/ Resolution	MXD/RES
NDR	Minimum Dimension/ Resolution	MND/RES

Parameters		Definition
MIT	Mean Intensity of Target	MIT, SIT, and KIT are derived from the distribution of target intensities.
SIT	Sigma Intensity of Target	
KIT	Skewedness Intensity of Target	
PIT	Peak Intensity of Target	Peak intensity of the target.
MIB	Mean Intensity of Background	MIB, SIB, and KIB are derived from the distribution of background intensities.
SIB	Sigma Intensity of Background	
KIB	Skewedness Intensity of Background	
CON	Scene Contrast	$\sum_{k=0}^{N-1} (k - G)^2 P(k)$ where G is the mean overall intensity and P(k) is the probability of level k. CON may be estimated by substituting MIB for G.
ATC	Average Target Contrast	$\lvert MIT - MIB \rvert / Max(MIT, MIB)$
PTC	Peak Target Contrast	$\lvert PIT - MIB \rvert / Max(PIT, MIB)$
PRA	Perimeter to Root Area	$PER/\sqrt{TAR}$
HWA	Height by Width to Area	HWA = (MXD · MND)/TAR, where the height and width of the target are defined in terms of the minimum and maximum dimension of the target.
NME	Number of Edges	(NME) = (the number of perimeter straight edges).

Parameters		Definition
MEL	Mean Edge Length (normalized)	MEL, SEL and KEL are the mean, sigma, and skewedness derived from the distribution of edge lengths normalized by the perimeter length.
SEL	Sigma (standard deviation) Edge Length	
KEL	Skewedness of Edge Length	
MDS	Mean Differential Slope	MDS, SDS and KDS are derived from the distribution of slopes of the target perimeter.
SDS	Sigma Differential Slope	
KDS	Skewedness Differential Slope	
TCS	Target Coarseness	TCS will be computed on the target using a variation of the techniques developed by Rosenfeld et al.[3]
SCS	Surround Coarseness	SDS will be computed on the area surrounding the target.
CSR	Coarseness Ratio	CSR = |TCS-SCS|/Max(TCS, SCS)
TDR	Target Directionality	$TDR = 1 - 8/n^2 \sum_{i=p}^{q} (i - k_{max}) H(i)$ where H denotes the histogram of normalized edge number by directional angle; k_{max} is given by $H(k_{max}) = \max(H9k)$ for $k = 1, \ldots, n$; $p = k_{max} n/2 + 1$; $q = k_{max} + n/2$; and n is the quantizing number of angle.

3. K. C. Hayes, Jr., A. N. Shah, and A. Rosenfeld, "Texture Coarseness: Further Experiments," IEEE Trans. Syst., Man, Cybern., Vol. SMC-4, No. 5, September 1974, pp. 467-472.

Parameters		Definition
SDR	Surround Directionality	SDR will be similarly computed on the area surrounding the target.
DRR	Directionality Ratio	$DRR = \lvert TDR-SDR \rvert / Max(TDR, SDR)$
LLK	Line Likeness	$(LLK) = \sum_{i}^{n} \sum^{n} \frac{[\cos(i-j)2/n]P(i,j)}{\sum_{i}^{n} \sum_{j}^{n} P(i,j)}$ where P(i, j) is the relative frequency with which two neighboring cells separated by a distance d along the edge direction occur, one with direction code i and the other with j, and n is again the quantization of the angle π.
CLT	Clutter	CLT will be defined psychometrically by having a group of subjects rate the pictures on a ten-point scale with the end points (1 and 10) specified.
HOM	Homogeneity	HOM will be defined psychometrically as CLT.
NCO	Number of Confusing Objects	NCO is a count of the number of objects whose area is within a threshold absolute value of the target area.
AID	Average Intensity Difference	These parameters are described in the last quarterly report on this project.
BNG	Boundary Gradient	
GLM	Gray-Level Mean	
GLC	Gray-Level Contrast	
ASM	Angular Second Moment	
ENT	Entropy	

Parameters		Definition
MXC	Mean Extremum Count	These parameters are described in the last quarterly report on this project.
MXS	Mean Extremum Smoothed	
SXS	Sigma Extremum Smoothed	
KXS	Skewedness Extremum Smoothed	

EXPERIMENTATION

Search performance data will be measured for a range of small targets on a variety of backgrounds. In our previous studies, obtaining adequate stimuli has always been a major problem, and it continues to be one now. We had hoped to use the imagery developed in the main contract. Unfortunately, these cannot be used because they do not present a sufficiently difficult search task. Even when degraded, the task is reduced to basically one of recognition for various levels of image quality. We have decided, therefore, to create the search fields using our computer and digital image processing facilities to embed the targets in selected scenes. The process will be described in detail in the subsections which follow.

Scenes and Targets

We have selected 23 scenes from an initial file of more than 400 diverse photographs used in an earlier NVL contract. These 23 scenes were selected for their distinct contrast and for their suitability for target embedding. Figures 47, 48, and 49 are examples of these scenes.

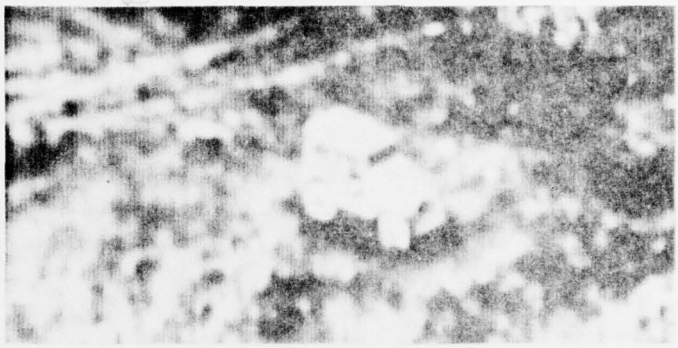
Figure 47. Forest Scene

Figure 48. Mountain Scene

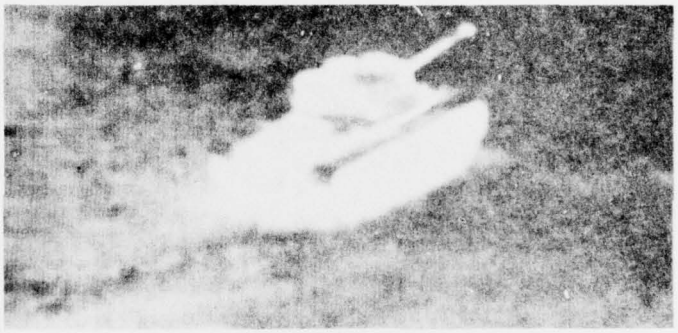
Figure 49. Field Scene

There are two available sets of targets. One set consists of about 60 photographs of a jeep, truck, tank, van, and artillery pieces taken from various aspect angles. Three of these targets are shown in Figure 50. The second set consists of 60 photographs of the monitor from an infrared (3 to 5 microns) sensor of the same targets.

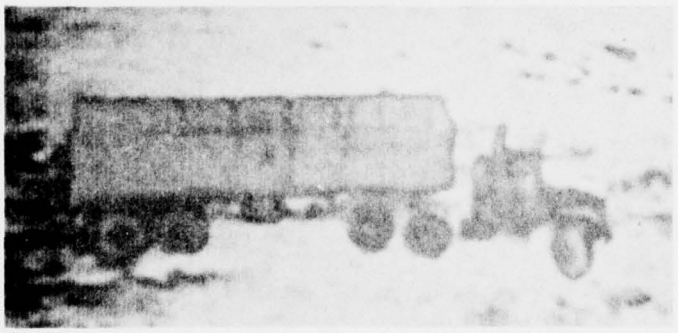
We plan to create about 500 search fields by embedding approximately 20 targets in each of the scenes. We hope to use both the photographic and infrared targets. However, if the latter targets, either as a positive or a negative, do not work, we will be restricted to the photographic targets.



a) Jeep



b) Tank



c) Truck

Figure 50. Targets

Each search field will contain an indicated region which shows where the target will be. We will devise a procedure for outlining that region which will minimally interfere with the perception of the entire scene and especially with that part which contains the target.

Most likely, the subjects' task will be to locate and classify the target (jeep, truck, etc.). Since the search area will be considerably smaller than the total scene, the time for each trial will be limited to 5 to 10 seconds.

Each of the 500 scenes will be presented at two image quality levels. Data will be obtained from 20 to 25 subjects for each of the 1,000 (2 x 500) presented scenes. The scenes will be presented as projected transparencies subtending about 20 x 20 degrees of visual angle. Since we are planning to collect a large amount of data, 20 to 25 thousand trials, we are hoping to run many subjects together, possibly in a classroom setting. This presents some difficulties in collecting the time data. One possibility is simply to collect success-failure data for each trial and have different trial durations for different groups of subjects. Another is to implement some form of self-timing.

EMBEDDING TARGETS IN SCENES

The requirement is to produce a large number of search fields, each containing a target. The most feasible way to meet this requirement is by embedding the target in the scene. Computer embedding is preferred to photographic processing because of its high degree of control and flexibility.

Embedding is simply the substitution of selected picture elements in the scene by values corresponding to the target. This is a trivial computation for any computer. However, one must take care that the embedded target matches the overall scene with respect to size, intensity range, blur, orientation, and shadow position. Further, it is important to remove any artifacts associated with the embedding process such as light or dark edges surrounding the target.

We are presently writing and testing the software. The main computations are done using the man-computer lab facility; we are using two nine-track tape drives, a disc, a plasma display to display targets, scenes, and numeric intensities of target elements, and a keyboard for control, setting parameter values, outlining the target, and locating the target in the scene. The data tapes produced in this laboratory are then taken to the digital image processing facility to be made into the film transparencies used in the search experiments.

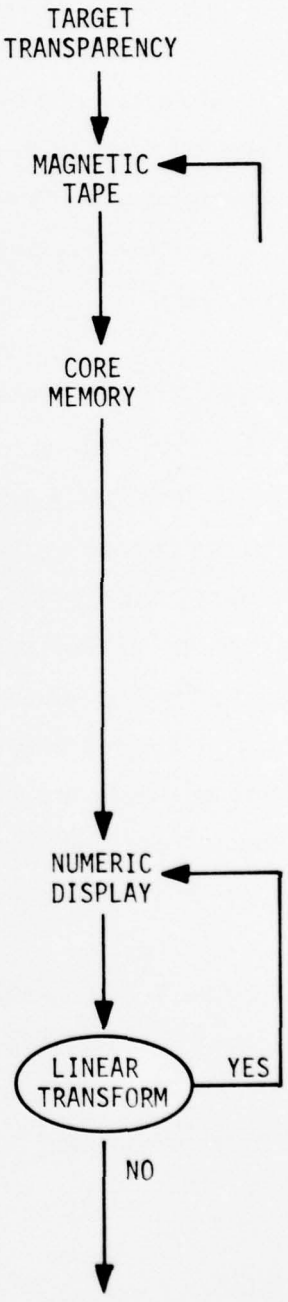
The embedding process starts with target and scene transparencies. These transparencies are analyzed by a scanning microdensitometer and put on magnetic tape. Then a target file and scene file are created on magnetic disc. During embedding, a target is selected, target parameter values are defined, a scene is selected, and the target is placed in a specified location in the scene. The output goes to magnetic tape. At the end, a transparency is made from the magnetic tape. The following flowcharts describe these events in greater detail.

Creating Target File

The target transparency (25 x 50 mm) is scanned with a 100 analyzing aperture. The data on magnetic tape consist of 250 records with 500 bytes per record. Each byte contains the density (0.2 to 2.5) at each picture element.

The target file is then read into core memory by collapsing the picture from 250 x 500 to 27 x 54. Since the embedded target consists of a small number of cells, there is no loss of essential information. The collapsing is accomplished by first converting the density information to intensity information and then calculating the mean value for each 9 x 9 block of picture elements. The calculated value ranges from 0 to 100.

The 27 x 54 picture is presented pictorially and numerically on the plasma display. The numeric display is a 27 x 54 array of integers from 0 to 9 corresponding to the original calculated values. However, a linear transformation on the original values can be called at the keyboard to produce a more usable picture. For example, if the original range was from 0 to 20, one might multiply by 5 to expand the range.

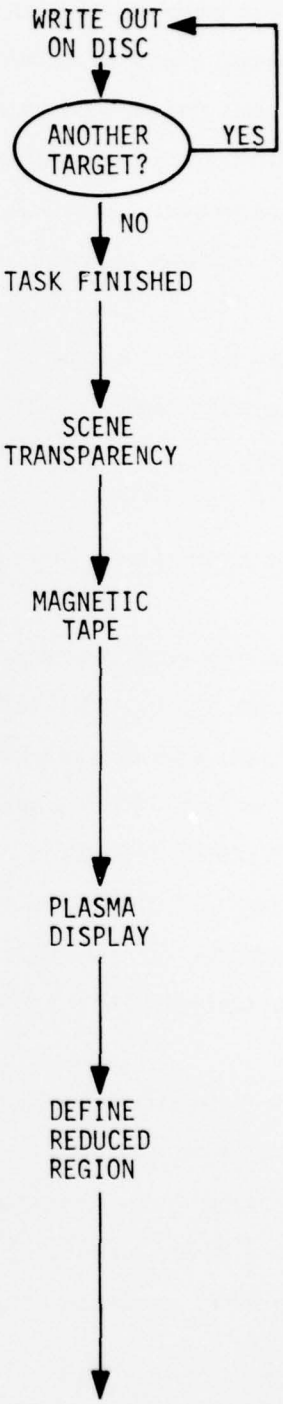


The pictorial display presents the intensity (luminance) at each picture element. Ten levels are used, corresponding to the 10 integer values. Since the plasma display is a 512 x 512 array of dots, each being either on or off, each picture element in the 27 x 54 picture is displayed using a 3 x 3 array of dots. Here the 10 intensity levels are displayed using from 0 to 9 dots.

The next step is to outline the target, including its shadow. The operator uses the keyboard to position the cursor on the numeric display. When the cursor is positioned at the target boundary, that part of the row to the left or right of the cursor can then be eliminated by pushing the appropriate control button. The pictorial display helps one to see where the target is, while the numeric display is more useful for precise location of the target boundary.

Some intensity and shape features of the target can now be calculated. This calculation is facilitated since the target data are now isolated from the remaining picture.

The target array along with the calculated features are then written out on disc. This process is repeated for the entire target set.



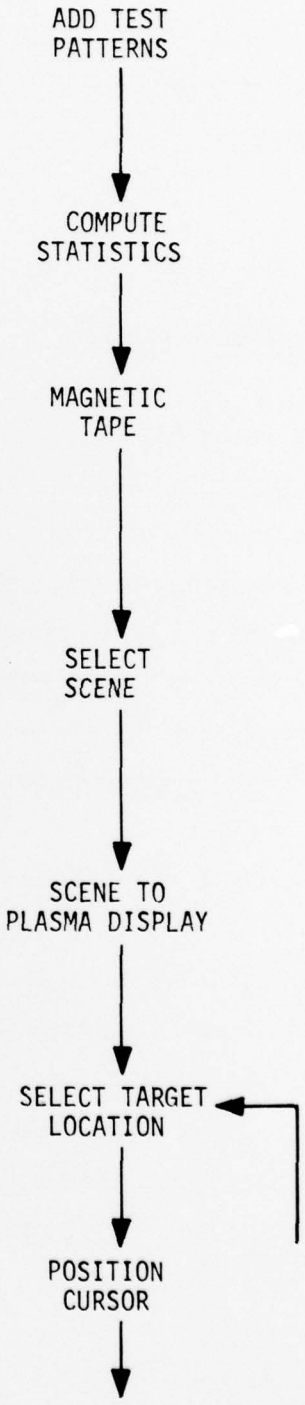
Creating Scene File

The scene file is constructed much more simply. The scene transparency (55 x 55 mm) is scanned with a 50μ analyzing aperture. The data on magnetic tape then consist of 1100 records with 1100 bytes per record. Later we may find that a 550 x 550 array suffices to portray the scene. We would then modify the processing accordingly.

The digitized scene is then transformed to magnetic disc in order to reduce it to 800 records with 800 bytes/record. The reason is to produce a projection transparency 40 x 40 mm which can be used in ordinary 35 mm slide projectors. This reduction is accomplished by displaying the scene on the plasma display and defining the reduced region to be retained.

A test pattern (checkerboard or tri-bar) at several sizes is added to the scene so that system resolution can be measured during the experiment. In addition, some intensity and texture statistics will be calculated for specified regions in the scene.

The reduced scene is then written out on magnetic tape for the embedding which occurs later.



The Embedding Process

The embedding process consists of selecting a scene, selecting a target, transforming the target, and placing it in a specified location in the scene. First, one of the 23 scenes is selected. It is then collapsed to a 160 x 160 array and displayed on the plasma display. Each cell in the 160 x 160 array requires a 3 x 3 array for display.

The intention now is to embed several targets in each scene as a means for trying out different target transformations to see which is best for a given scene. First, a location in the scene is selected by positioning a cursor.

Then a target is selected and its size, intensity, range, rotational angle, and degree of blur are specified. This is repeated for as many targets as derived for that scene.

The resultant picture is written on magnetic tape. After a number of pictures are put on magnetic tape, the tapes can be made into pictures in three ways. The picture can be viewed immediately by transferring the information first to a scan converter and then to a video monitor. It takes slightly longer to produce a picture on paper. This is a degraded representation of the picture. Finally, a high quality film transparency can be made. Since we have some degree of control over the picture on the monitor (for example, a part of it can be enlarged), we expect that the monitor will be used to decide which target transform to use.

Then, the entire process is repeated using the proper target transformation to create scenes containing single targets.

This time a slight blur function is added to the scene to help eliminate any target boundary artifacts. A film transparency is then made from each file on magnetic tape.