NRTC 79-41-R

(2) LEVEL II

PASSIVE DETERMINATION OF RANGE AND DEPTH FROM MOTION STEREO ANALYSIS

AD A083620

December, 1979

William B. Lacina

William Q. Nicholson

NORTHROP RESEARCH AND TECHNOLOGY CENTER
One Research Park
Palos Verdes Peninsula, California 90274
Telephone: (213) 377-4811

Sponsored by

Defense Advanced Research Projects Agency (DoD)
ARPA Order No. 3501
Monitored by MIRADCOM under Contract No. DAAK40-78-C-0047

NORTHROP
Research and Technology Center

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 14 NRTC-79-41R
2. GOVT ACCESSION NO.: AD-A083 620
3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): 6 Passive Determination of Range and Depth from Motion Stereo Analysis,
5. TYPE OF REPORT & PERIOD COVERED: 9 Final Technical Report, Jan[redacted]78 — Sep[redacted]79
6. PERFORMING ORG. REPORT NUMBER: NRTC 79-41-R
7. AUTHOR(s): 10 William B. Lacina, William Q. Nicholson
8. CONTRACT OR GRANT NUMBER(s): 15 DAAK40-78-C-0047, ARPA Order-3501
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Northrop Research and Technology Center, One Research Park, Palos Verdes Peninsula, California 90274
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: ARPA Order 3501
11. CONTROLLING OFFICE NAME AND ADDRESS: Defense Advanced Research Projects Agency, 1400 Wilson Blvd., Arlington, Virginia 22209
12. REPORT DATE: 11 Dec[redacted]79
13. NUMBER OF PAGES: 48

12 56

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): U.S. Army Missile Research & Development Command, Redstone Arsenal, Alabama 35809

407696

15. SECURITY CLASS. (of this report): Unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:
16. DISTRIBUTION STATEMENT (of this Report)

Distribution is listed at the end of document. Other requests for document should be referred to DARPA, Attn: Cmdr. Wiener

This document has been approved for public release and sale; its distribution is 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)

Image processing, motion stereo, passive ranging, depth aided target acquisition, correlation tracking, dynamic imagery.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report describes results of passive depth and range determination from processing multiframe imagery obtained with a moving sensor platform, by exploitation of motion stereo analysis. Results obtained from synthetic and real imagery are presented. Application of the present technique to autonomous target acquisition is discussed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF FIGURES
(Continued)

LIST OF FIGURES
(Continued)

1.0 SUMMARY

An objective of high priority for the Department of Defense is the development of a reliable, accurate and economical Cruise Missile weapon system having operational capability within three to five years. The DARPA Advanced Cruise Missile Program is exploring those technologies which will provide significant improvement in performance for the next generation of cruise missiles. A major objective of the DARPA Autonomous Terminal Homing Program is the development of precision guidance techniques which will enable the effective destruction of fixed, high value strategic and theater targets using non-nuclear munitions.

A critical problem for the cruise missile is the development of image processing techniques applicable to target acquisition for an autonomous terminal homing system which depends upon an on-board comparison of a sensed scene with a stored replica of a predesignated target area. Extensive efforts are currently in progress to develop algorithms based upon area correlation and feature matching techniques for accurate registration of sensed and reference imagery. However, image intensity matching depends upon several unpredictable factors such as time of day or year; weather; changes in scale, viewpoint, and perspective; spectral and sensor characteristics, etc. In contrast, one of the most invariant properties of a scene is its geometric form, defined by a sensed height distribution of a target scene, which can be determined passively from dynamic imagery by exploitation of the concept of motion stereo.

Autonomous target acquisition based upon exploiting geometric form by matching a passively sensed reference height map offers an attractive approach either as a supplement or as an alternative to conventional scene matching techniques. Since scene matching tests performed during Phase I showed improved results when augmented with range information and since the existence of enemy defenses may dictate passive operation, passive ranging and height matching techniques should be further developed. A reference height image in planimetric form offers an all-azimuth capability using only a single reference image, and the ability to attack from any direction has obvious military significance.

During Phase I of the DARPA Autonomous Terminal Homing Program (ATHP), Northrop Research and Technology (NRTC) proposed and demonstrated a method for measuring range (and/or depth) to a target using multiframe imagery from a passive sensor. This document presents a brief technical discussion of the motion stereo concept, as well as a summary of the results obtained. An accuracy of 0.2% of flight height was obtained using both terrain simulator imagery and synthetically generated imagery, and better than 2% of flight height was obtained with real imagery even when perturbed by lack of sensor stabilization. A preliminary study of the proposed hardware implementation of the present motion stereo processing algorithms is discussed.

2.0 INTRODUCTION

One of the critical requirements for an autonomous terminal homing system is the development of image processing algorithms for onboard comparison of a sensed scene with a stored replica of a predesignated target area. Many techniques, both active and passive, are being explored to achieve this goal. An obvious disadvantage to active guidance is that it may alert enemy defense and be vulnerable to countermeasures. In the case of passive techniques, extensive efforts have been undertaken to develop algorithms based upon area correlation and feature matching for accurate registration of sensed imagery with a given reference image.

Target acquisition by image intensity correlation suffers from several disadvantages that relate to the unpredictable or changeable factors in sensed and reference imagery prepared at different times or under different conditions:

- Time of day, year
- Weather conditions
- Scale factor, perspective, viewpoint
- Illumination angle, shadows
- Different sensor characteristics
- Ease of camouflage
- Lack of azimuthal capability

In contrast to image intensity, one of the most invariant properties of a scene is its geometric form -- unlike the radiated or reflected intensity distribution, the elevation distribution of the target scene is relatively permanent. Thus, techniques for determination of the three-dimensional form of observed scenes should be applicable to depth-aided target acquisition. Instead of determining target location by correlation matching of grey level intensity values over the two-dimensional image coordinates, elevation values defined over the ground coordinates could be used as shown schematically in Fig. 1. This approach would have the following advantages:

- Passive (immunity to detection and countermeasures)
- Independent of sensor wavelength and observation conditions
- Improved information extraction from multiframe integration
- Three-dimensional form fitting in ground coordinates
- Reference map could be a wire-frame model

Comparison of a sensed elevation distribution (rotated and translated as necessary) with a reference model of the target elevation data may be sufficient in itself for attainment of accurate registration, or it could at least supplement the use of image intensity correlation for location of the target.

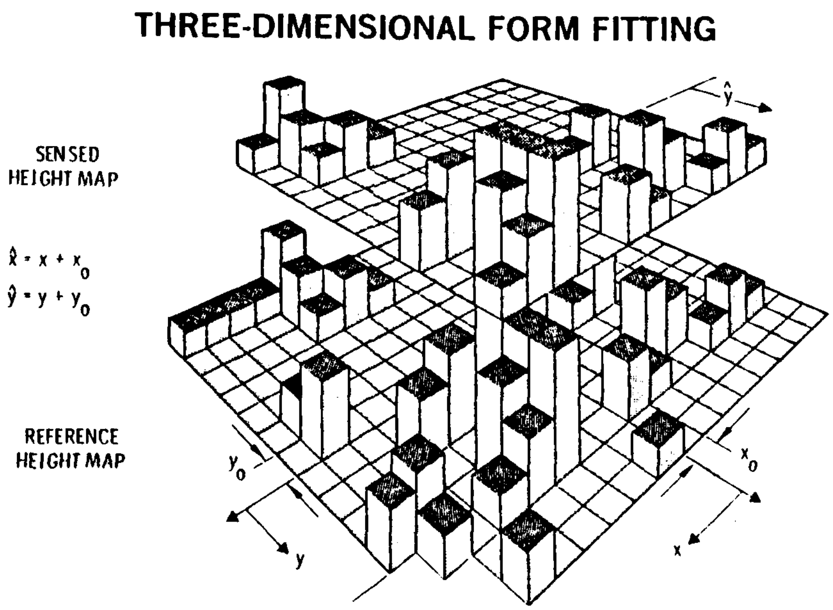


Fig. 1. Registration of a sensed and reference elevation distribution, defined over a ground coordinate system by three-dimensional form-fitting.

The form fit concept offers the potential of an all-azimuth target acquisition capability through address rotation of a single reference height (or height gradient) array not available with either a slant range and/or intensity reference. This potential can provide the vital military element of surprise for low altitude target penetration from any direction, countered only by unacceptable enemy cost of an all-azimuth defense. Successful and practical implementation of the motion stereo concept will have potential application in tactical as well as strategic weapons, in military communication systems (image bandwidth compression) as well as in advancing other important military technologies.

During Phase I of the Autonomous Terminal Homing Program, the DARPA Strategic Technology Office initiated, advanced and evaluated the technology required for autonomous target acquisition and terminal homing. Both active and passive sensors were evaluated and both are considered to be suitable candidates for the cruise missile application. It is anticipated that a passive sensor capability will become more important as vehicle penetration problem becomes more acute and as military adversaries develop more effective defenses against the cruise missile. The need for active sensors has been based upon their capability to sense range and, in turn, range rate. It has been demonstrated that use of range parameters improves terminal homing accuracy and thus helps achieve the precision guidance objective. More important, however, is the value of range information for autonomous acquisition, since terminal homing cannot take place without first achieving target acquisition.

The three major areas of the ATHP effort consist of scene measuring, reference preparation and scene matching. The objective of these three areas of effort has been to determine the optimal sensor and scene matching technique which could best accommodate the inevitable changes that occur between the reference and sensed scene. Thus, a substantial portion of ATHP effort has been directed toward evaluation of scene matching accuracy (match error) and reliability (false fix, no match) for each of several techniques over many conditions causing scene variation to occur in an operational weapon system. Changes due to sensor type and spectrum, approach angle, time of day, weather, time of year, range, and reference image synthesis were all expected to affect accuracy and reliability of scene matching. It was in this context that Northrop Research and Technology Center proposed a scene matching technique based on geometric form-fitting, since three-dimensional form is a more invariant property of a scene than image intensity.

From preliminary results of scene matching tests, it became apparent that all proposed scene-matcher techniques worked well given a satisfactory reference and sensed image. Both accuracy and reliability satisfied evaluation criteria using real imagery from the same or similar wavelength sensor given adequate resolution. Except for hardware complexity, there was little to differentiate in the selection among scene matching techniques. However, as scene matching tests became more realistic and included more of the likely opeational variations between reference and sensed scene, the more robust techniques showed less variation with time of day, azimuth and range differences. Nevertheless, overall performance of the matchers was still consistently good given a real, sensor-based reference. It is when a synthetic reference image is used that reliability degrades. Significant degradation results with a synthetic intensity reference, and overall reliability drops to an unacceptable level. However, contractors who presented results at the Sixth Autonomous Terminal Homing Program Technical Interchange Meeting at AFAL in March 1979 confirmed that scene matching algorithms for which synthetic intensity data is augmented with range or wire frame data perform much better.

Based upon the planned operational concept of synthetic reference preparation, the above background results of the ATHP to date clearly support the technical advantage for using range and/or three-dimensional form (wire frame) data in advanced target acquisition techniques for the cruise missile. The anticipation of a more acute vehicle penetration problem emphasizes the technical need for a passive cruise missile sensor. These two conditions support the need for a passive ranging capability, and Northrop's results to date, described in Section 3.0, confirm that high accuracy passive ranging can be achieved.

3.0 ACCOMPLISHMENTS

This section describes results obtained by Northrop under contract DAAK40-78-C-0047, ARPA Order No. 3501, initiated on 24 January 1978 for Depth-Aided Target Acquisition for the Cruise Missile.

The feasibility of exploiting dynamic imagery obtained from a moving platform for passive determination of the three-dimensional form of an object scene using motion stereo analysis has been examined. Preliminary results have indicated that the accuracy of depth determination (relative to vehicle flight altitude) can be $\leq 0.2\%$. Accuracies significantly less than the sensor resolution have been achieved by fractional pixel interpolation and multiple frame processing, which provides the opportunity for statistical refinement of the derived elevation data. These results have been limited by the available data bases and by the simplified approach of tracking discrete points, and no systematic investigation has been made of the sensitivity or dependence of the accuracy on flight parameters or image characteristics. Further work is required to evaluate the performance of these techniques as a function of flight geometry, sensor parameters, and image statistics, and to extend them to area processing algorithms which are suitable for implementation by real-time hardware.

3.1 Introduction

The motion stereo concept is illustrated schematically in Fig. 2, which shows a moving vehicle (platform) containing a framing sensor. As the vehicle approaches the target scene, its framing sensor generates a sequence of images corresponding to observation of the target from a succession of spatially separated vantage points. The sample frames shown in Fig. 2 illustrate the stereoscopic advantage provided by the changing perspective. The geometry of the motion stereo system depicted schematically in Fig. 2 can be conveniently described in two coordinate systems shown in Fig. 3. The platform, moving with velocity v, contains an oblique forward-looking imaging system with its optical

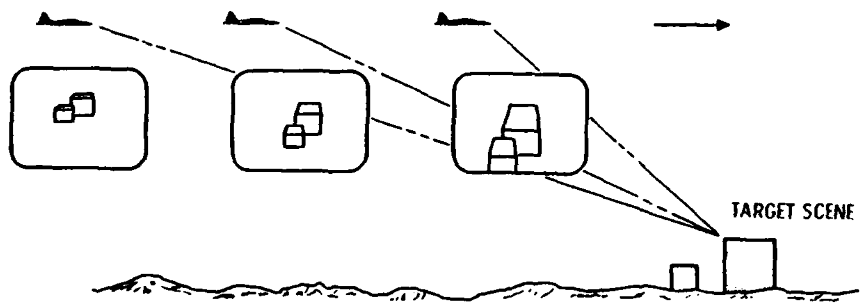


Fig. 2. Stereoscopic observation of an object scene obtained from dynamic imagery generated by a framing sensor mounted on a moving vehicle.

axis oriented downward at an angle θ in the vertical plane containing the flight trajectory. If the flight velocity is constant and the trajectory is parallel to the ground, the equations of motion for an image point are[1,2]

$$dX(t)/dt = (v/fd) \cos^2 \theta \; X \, [Y + f\tan \theta] \qquad (1)$$

$$dY(t)/dt = (v/fd) \cos^2 \theta \, [Y + f\tan \theta]^2 \qquad (2)$$

where f is the sensor focal length and d is the depth.

From these equations, expressions can be obtained[2] for the address shift (ΔX, ΔY) of an image point between two successive frames separated by a time interval Δt; to first approximation,

1. W. B. Lacina and W. Q. Nicholson, "Concept Validation of Depth Aided Target Acquisition for the Cruise Missile," Northrop Rpt #NRTC-78-42R, Nov. 1978.
2. W. B. Lacina and W. Q. Nicholson, "Passive Determination of Three-Dimensional Form from Dynamic Imagery," in Digital Processing of Aerial Images, SPIE Vol. 186, pp. 178-179, May 1979.

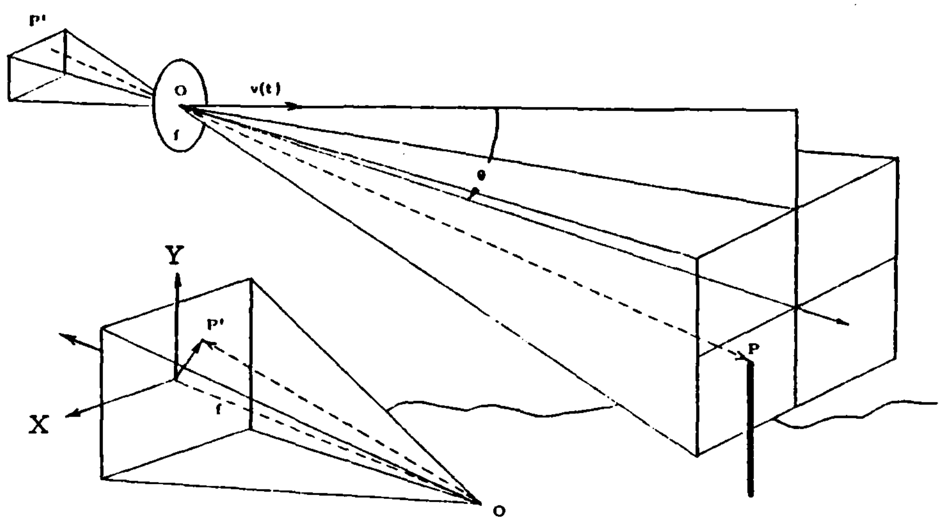


Fig. 3. Ground and sensor referenced coordinate systems for motion stereo analysis, with focal plane of the sensor shown enlarged in the inset.

$$\Delta X = (v \Delta t/fd) \cos^2 \theta \; X [Y + f\tan \theta] \tag{3}$$

$$\Delta Y = (v \Delta t/fd) \cos^2 \theta \; [Y + f\tan \theta]^2 \tag{4}$$

For any given frame, the set of address shift vectors

$$\vec{M}(X,Y) = [\Delta X(X,Y), \Delta Y(X,Y)] \tag{5}$$

which specify the change of location of every pixel (X,Y) that will result in a subsequent frame defines a Motion Vector Field. Note that the Motion Vector Field only partially describes the change of imagery that will occur in the succeeding frame; in general, the image transformation defined by the motion vector field is not a one-to-one mapping of frames, since occlusion of (old) or appearance of (new) objects will inevitably occur.

The fundamental problem of motion stereo analysis is the development of techniques for automatic computation of the motion vector field $\vec{M}(X,Y)$ from a given frame pair, and the subsequent transformation of that data into an elevation distribution in ground coordinates. In principle, any single pair of frames is sufficient for the determination of the range and depth of every object point which lies in the _common field of view_ of the two frames, assuming of course that the parameters (e.g., distance to target scene, spatial separation of the stereo views, etc.) lie within suitable ranges for which the concept is valid. In practice, however, there may be imperfect knowledge of the vehicle flight parameters (velocity, altitude, trajectory, orientation) as a function of time, and there will be changes in scale factor and geometrical perspective as the sensor platform approaches the target scene. Furthermore, the imagery may be corrupted by noise or characterized by poor resolution, fluctuations in intensity, or abnormal gradient statistics. Finally, there will be inherent errors in the computation of the motion vector field associated with spatial and/or grey level quantization effects and the discrete nature of "tracking" or other algorithms. Thus, range and depth data derived from different frame pairs will not necessarily produce identical numerical results. Statistical techniques can be used to filter the data to determine the "best estimate" of the three-dimensional form from observations over several frame pairs. In effect, multiframe processing of a sequence of frames makes it possible to exploit the high redundancy of dynamic imagery for refinement of the estimated range and elevation data.

Fig. 4 shows two successive frames of imagery. Any point in Frame n which remains visible in the subsequent Frame (n + 1) -- for example, the corner of the building -- will be shifted by a motion vector $(\Delta X, \Delta Y)$ as illustrated. From Eq. (3) - (5), the magnitude of this motion vector is inversely proportional to the object depth d, while its direction $\Delta Y/\Delta X$ is independent of depth. In principle, it is possible to obtain both the range and the depth of the object point from knowledge of ΔY.

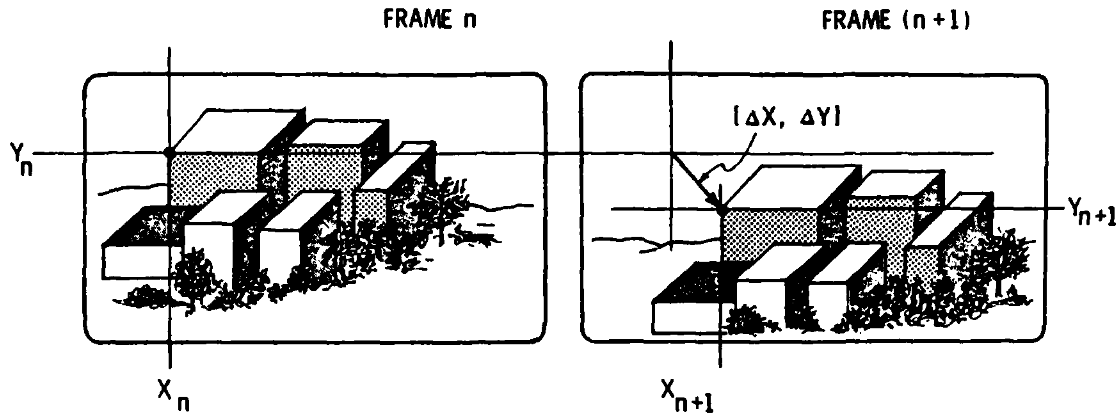


Fig. 4. Interframe address shift [ΔX, ΔY] illustrated for a sample pixel (X_n, Y_n) of Frame n, defines a Motion Vector Field.

Motion stereo processing of a sequence of frames is depicted schematically in Fig. 5 where data obtained by computation of the motion vector field is shown transformed into successive approximations to a height map. The first pair of frames (0 and 1) produces an estimate of the elevation distribution which has been labeled "Frame 1". Subsequent pairs of frames can be used to generate a succession of updated estimates of the ground elevation distribution, improved by two advantages. First, for those object points which have remained within the field of view over a long sequence of frames, a running average of the processed data permits a continuous statistical refinement of the derived elevation data. Second, as the sensor platform approaches the target scene, new information about the three-dimensional form becomes available as previously obscured structures enter the field of view.

A straightforward approach to calculation of object depth proceeds as follows. Assume that an object, initially imaged at (X_o, Y_o), follows a coordinate trajectory

$$(X_o, Y_o), (X_1, Y_1), (X_2, Y_2), \ldots (X_n, Y_n), \ldots (X_N, Y_N)$$

over a sequence of frames. From Eq. (2), a determination d_n of the corresponding object depth can be made for each of the frame pairs [n-1,n] in terms of the observed image point coordinates:

$$d_n = (v \Delta t/f) \cos^2 \theta \, [Y_{n-1} + f\tan\theta] \, [Y_n + f\tan\theta]/\Delta Y_n \qquad (6)$$

where

$$\Delta Y_n = Y_n - Y_{n-1} \qquad (7)$$

For several reasons discussed earlier, the values d_n defined by Eq. (6) will fluctuate from frame to frame. A simple definition of the statistical best estimate of the object depth d would be to use the running average over a sequence of such determinations d_1, d_2, d_3,....d_N:

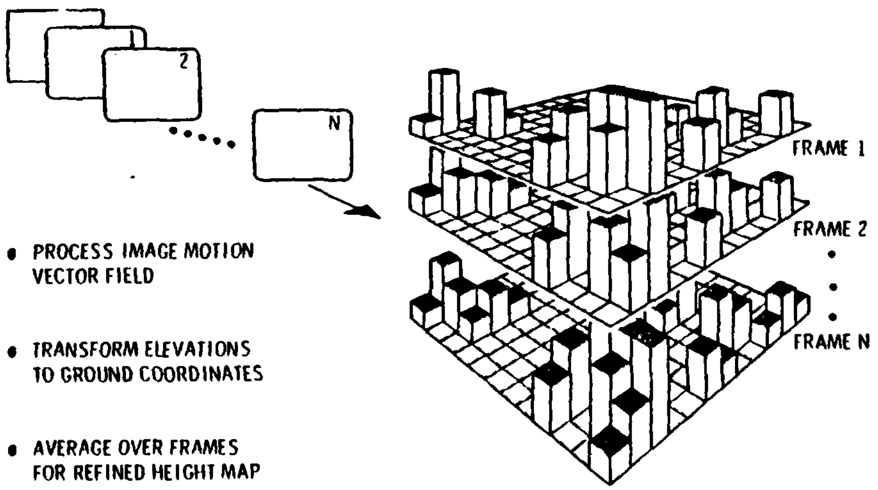


Fig. 5. Successive frame pairs from a sequence of dynamic imagery obtained from the moving sensor are processed for computation of the motion vector field, which is subsequently transformed into an estimated elevation distribution in ground coordinates. Multiframe integration permits a statistical refinement of height map.

$$d = \langle d_i \rangle = (1/N) \sum_n d_n, \tag{8}$$

or,

$$d = (v \Delta t \cos^2 \theta / fN) \sum_n [Y_{n-1} + f\tan\theta][Y_n + f\tan\theta]/\Delta Y_n. \tag{9}$$

Such an estimate assumes that all of the determinations d_i should be equally weighted, which may not be the optimum approach, since the later determinations (made at closer range) may be uniformly more accurate. Thus, we could alternatively define (for example) the best estimate for depth to be that value of d which minimizes the squared error of the actual line shifts ΔY_n from their predicted values,

$$S = \sum_n [\Delta Y_n - (v \Delta t/fd) \cos^2 \theta \, (Y_{n-1} + f\tan\theta)\,(Y_n + f\tan\theta)].^2 \tag{10}$$

Setting $\partial S/\partial d = 0$ gives

$$d = (v \Delta t \cos^2 \theta / f) \frac{\sum_n (Y_{n-1} + f\tan\theta)^2 \, (Y_n + f\tan\theta)^2}{\sum_n (Y_{n-1} + f\tan\theta)\,(Y_n + f\tan\theta)\,\Delta Y_n}, \tag{11}$$

which is equivalent to the weighted average

$$d = \sum_n (d_n \Delta Y_n)^2 \quad \sum_n d_n \Delta Y_n{}^2 = \langle d^2 \Delta Y^2 \rangle / \langle d \Delta Y^2 \rangle. \tag{12}$$

3.2 Numerical Results

In general, the implementation of motion stereo analysis requires the development of computationally efficient area processing algorithms for calculation of the motion vector field $M(X,Y) = [\Delta X(X,Y), \Delta Y(X,Y)]$ for any given frame pair. However, a preliminary validation of the concept can be demonstrated by calculating the motion vector for selected discrete points. Frame-to-frame "block tracking" using normalized product (or minimum absolute difference) correlation matching is used to obtain the image point coordinate trajectory that corresponds to some fixed object point in the target scene.

Computation of the image point trajectory for specific object points was accomplished by frame-to-frame block tracking. A small (2N+1) x (2N+1) "reference image", centered at the pixel nearest to the address (X_n, Y_n) is extracted from Frame n and correlated over a larger (2M+1) x (2M+1) subarea of Frame (n+1) to determine the new location (X_{n+1}, Y_{n+1}) of the image point in Frame (n+1). Subpixel accuracy was attained by means of algorithms for prediction and correction of the tracking, using quadratic interpolation of the (2L+1) x (2L+1) correlation function (L=M-N). In general, the success of frame-to-frame block tracking depends upon both the geometrical and statistical characteristics of the imagery.

If a scene contains significant detail, accurate tracking can be achieved by using a very small block size for the reference image. Conversely, if the scene contains very little contrast, a much larger block size may be required. It is apparent, therefore, that the degree of spatial averaging defined by the block size, and thus the success of the tracking, can be scene-dependent. The statistical distribution of gradient values is an important factor to which the successful computation of interframe motion vector shift is intimately related. The "real imagery" of the first and third data base was characterized by a Gaussian distribution of grey levels, and a corresponding Rayleigh distribution of gradient values. In contrast, the statistical distribution of grey levels and gradient values for the "synthetic imagery" was quite artificial: the grey level histogram contained three discrete sharp peaks, and the gradient histogram was sharply peaked at zero. As would be expected, therefore, more difficulties were encountered with frame-to-frame tracking using the latter data base.

Other factors can also play an important role in the determination of an image point trajectory. For example, if there are significant inter-frame changes in scale factor, image intensity distribution, or geometric perspective, errors in correlation tracking may occur. The most reliable type of point for accurate tracking is a corner point, since the correlation algorithm will tend to lock onto the three intersecting edges. The use of real imagery minimizes the difficulties encountered from changes in geometrical scale or perspective since there is generally much more detail (gradient information) in the image for correlation tracking. The general problems associated with computation of the motion vector (by correlation tracking or otherwise) are anticipated to be much less severe with the more continuous statistical characteristics of real imagery.

Numerical results were obtained using four different dynamic imagery bases for which some ground truth is known. Parameters for these data bases are summarized in Table I.

Table I. Dynamic Imagery Data Base Parameters.

Parameter	Fuel Dump Complex	Hughes Culver City Complex	ERIM Down	(Lockheed-Sunnyvale) Forward
Flight Altitude (ft)	2500	1000	1000	685
Velocity (ft/frame)	140	50	24	12
Optical Depression Angle θ (^{o})	14.5^{o}	20^{o}	90^{o}	20^{o}
Field of View (H x V) (^{o})	5 x 2.5	20 x 20	29.6 x 38.8	18.6 x 24.8
Pixels/Frame (H x V)	400 x 160	256 x 256	400 x 512	400 x 512
Intensity Quantization Bit/Pixel	256 8	256 8	256 8	256 8

3.2.1 Correlation Algorithm

Registration of a sensed (S_{ij}) and reference (R_{ij}) image can be achieved by correlation techniques which are based upon determination of the shift (I,J) which corresponds to the extremum point of some metric function M(I,J) which measures the departure (distance) between the two images. For any point (i,j) of the reference image R, normalized product correlation locates the registration point as that shift (I,J) which maximizes

$$M_{ij}(I,J) = \sum_{xy} R_{x+i,\ y+j}\, S_{x+i+I,\ y+j+J}\, W_{yx} \Big/ \left(\sum_{xy} R^2_{x+i,\ y+j}\, W_{xy}\right)^{1/2} \left(\sum_{xy} S^2_{x+i+I,\ y+j+J}\, W_{xy}\right)^{1/2} \qquad (13)$$

while for minimum absolute difference correlation (MAD), the shift (I,J) is defined by the minimization of

$$M_{ij}(I,J) = \sum_{xy} \left| R_{x+i,\ y+j} - S_{x+i+I,\ y+j+J} \right| W_{xy} \qquad (14)$$

where the sum indices (x,y) may be regarded to range over all values for which the "window function" w_{xy} is nonzero.

Structurally, these two algorithms are quite similar, inasmuch as they both require that some specified binary operation (●) be applied to every point in the two arrays (R and S) which are relatively shifted by some vector (I,J). Aside from the additional normalization required in Eq. (14), this binary operation (●) is either subtraction or multiplication, which is applied <u>point-by-point</u> over the two (relatively shifted) arrays to form a "product" array P(I,J):

$$P(I,J) = R \bullet S. \qquad (15)$$

The elements (i,j) of this product array P (which is merely labeled by the shift parameters (I,J)) are given explicitly by

$$P_{ij}(I,J) \;=\; R_{ij} \bullet S_{i+I,\; j+J} \tag{16}$$

If it is intended to track every point in the initial (reference) frame R to its new address in a subsequent (sensed) frame S by using a (2N+1) x (2N+1) correlation block, as was done previously for software concept validation, the array P(I,J) must then be spatially filtered by a (2N+1) x (2N+1) mask W defined by

$$W = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & 1 & 1 & \dots & 1 \\ & & & \vdots & \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix} \tag{17}$$

to give a correlation array (labeled by parameters I,J)

$$M(I,J) \;=\; W * X(I,J), \tag{18}$$

where * denotes convolution. The array elements $M_{ij}(I,J)$ are the values of the correlation metric for every point (i,j) in the reference image when correlated with the sensed image shifted by (I,J) and weighted by a window function defined by the matrix W. Of course, the mask W need not be square (as was assumed), nor must its coefficients be equal to unity. Other windows may, in fact, be preferable for optimizing the performance of the algorithm.

For every point (i,j) in the reference image R, the components of the motion vector field $[\Delta X(i,j)\ \Delta Y(i,j)]$ are determined as local extremum points (maxima for normalized product, minima for MAD) of the correlation matrices $M_{ij}(I,J)$ over all (I,J) shifts spanning some search region. In order to obtain subpixel accuracy, functions $M_{ij}(X,Y)$ of continuous coordinates (X,Y) can be defined (e.g., using quadratic interpolation over the search region spanned by the discrete shifts (I,J).) It then follows that the roots (X_o,Y_o) to the equations

$$\partial M_{ij}(X,Y) / \partial X \,|_{X_o,Y_o} = 0,$$

$$\partial M_{ij}(X,Y) / \partial Y \,|_{Y_o,Y_o} = 0 \tag{19}$$

define the image motion vectors for the reference points (i,j):

$$[\Delta X(i,j), \Delta Y(i,j)] = [X_o,Y_o]. \tag{20}$$

If the range of vectors (I,J) is taken to be $(-L \leq I,\ J \leq L)$ (i.e., a correlation search over a square $(2L+1) \times (2L+1)$ region), then the straightforward approach that was taken in the original software validation would be simulated.

The symbolic computational structure just described for area correlation over the entire frame can be depicted schematically as shown in Fig. 6. Aside from the additional normalization required in Eq. (14), the binary operation ($\bullet$) is either a subtraction or multiplication, which must be applied point-by-point over two relatively shifted arrays. The normalized product algorithm has the advantage that two images which differ only by a scale factor will be correctly registered. However, this invariance is not expected to be an important consideration for dynamic imagery, since the successive frames are generated by the same sensor and would be approximately histogram equalized in grey level distribution. Therefore, the computationally simpler MAD algorithm is likely to be preferable for a hardware mechanization.

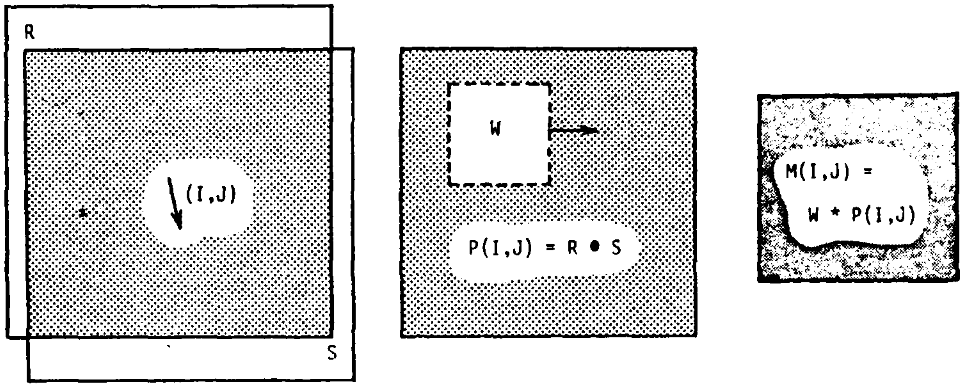
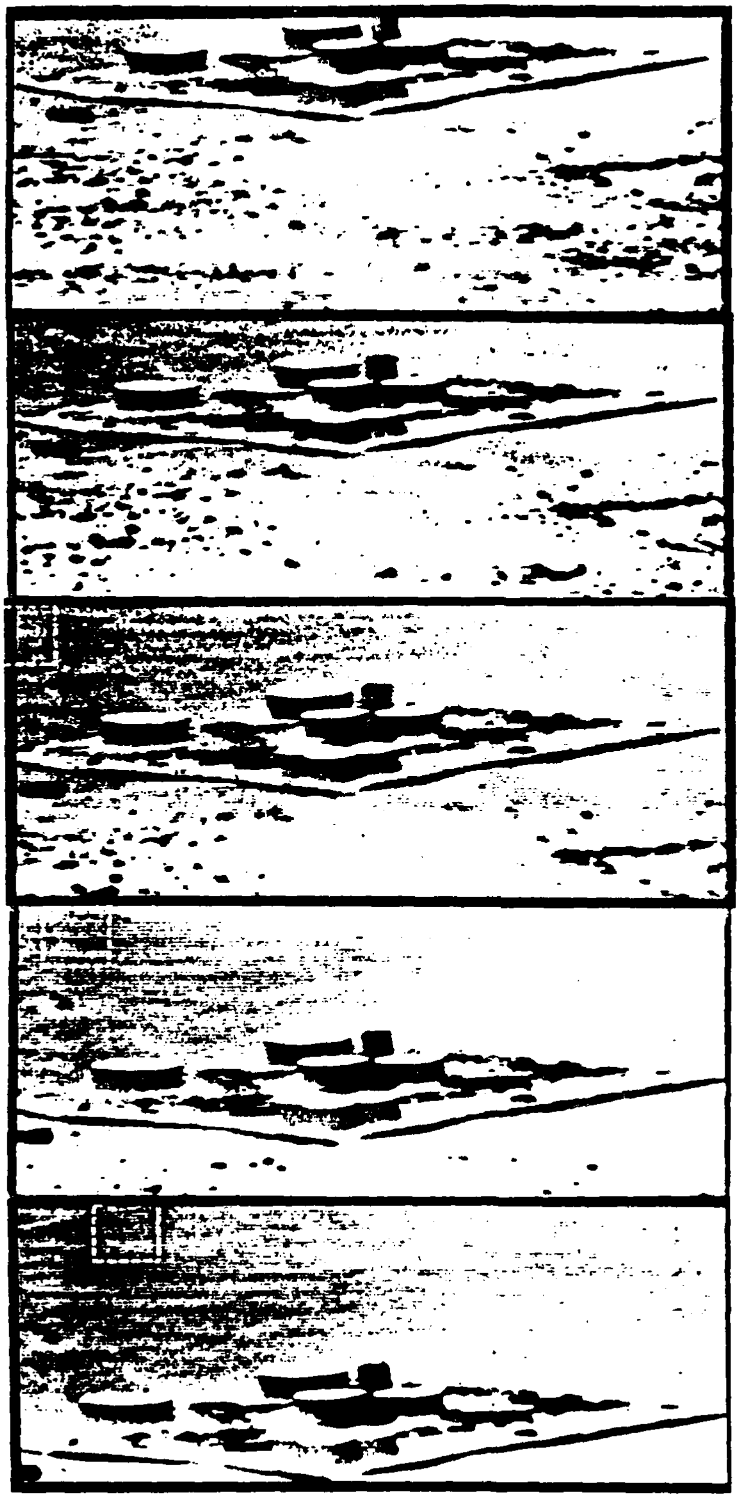


Fig. 6. Symbolic structure of correlation between a sensed array S, shifted by a vector (I,J) relative to a reference array R. A pointwise operation (•) is performed over the intersection to form a product array P(I,J) = R • S, which is then convolved with a spatial mask W to form the correlation array M(I,J) = W * P(I,J).

3.2.2 Terrain Simulator Imagery

A multiframe sequence of dynamic (30 frames/second) imagery was obtained from a silicon vidicon sensor using a 3D Terrain Board at Northrop's Ventura Division and digitized. Ten frames from the sequence, called the Fuel Dump Scene, were selected for processing. Simulation parameters and sample frames from the scene are displayed in Fig. 7. A point on the ground and a point at the top of one of the fuel tanks were tracked by frame-to-frame correlation. Fractional pel resolution for the image address of the tracked points was used in the tracking algorithm. Both product correlation and minimum absolute difference algorithms were used for tracking, with no significant difference in performance observed between the two algorithms.

Two different estimation algorithms (9) and (11) were used to obtain depth from the sequence of image address changes of a tracked object point, and the difference between the two methods (37 feet vs 36 feet) was negligible. Fig. 8 shows a plot of the tank elevation (as computed for each frame from the motion vector) and the running average versus frame number. The height variation due to attitude changes (vibration) of the camera is filtered by the



Simulation Parameters

- Northrop Ventura 3D Terrain Board
- Scale: 1:1000
- Altitude: 2500 ft
- Depression angle of optical axis below the horizon: $\theta = 14.5^{\circ}$
- Slant range to center of picture: 10,000 ft
- Horizontal field of view: 5°
- Horizontal width in field of view: 871 ft
- Vertical field of view: 2.5°
- Velocity: 140 ft/frame
- Quantization: 8 bit/pel 400 pel x 160 lines

Fig. 7. Sample frames (86, 88, 90, 92, 94) and parameters for the Fuel Dump Scene, prepared from RPV flight simulation with a model 3D terrain board.

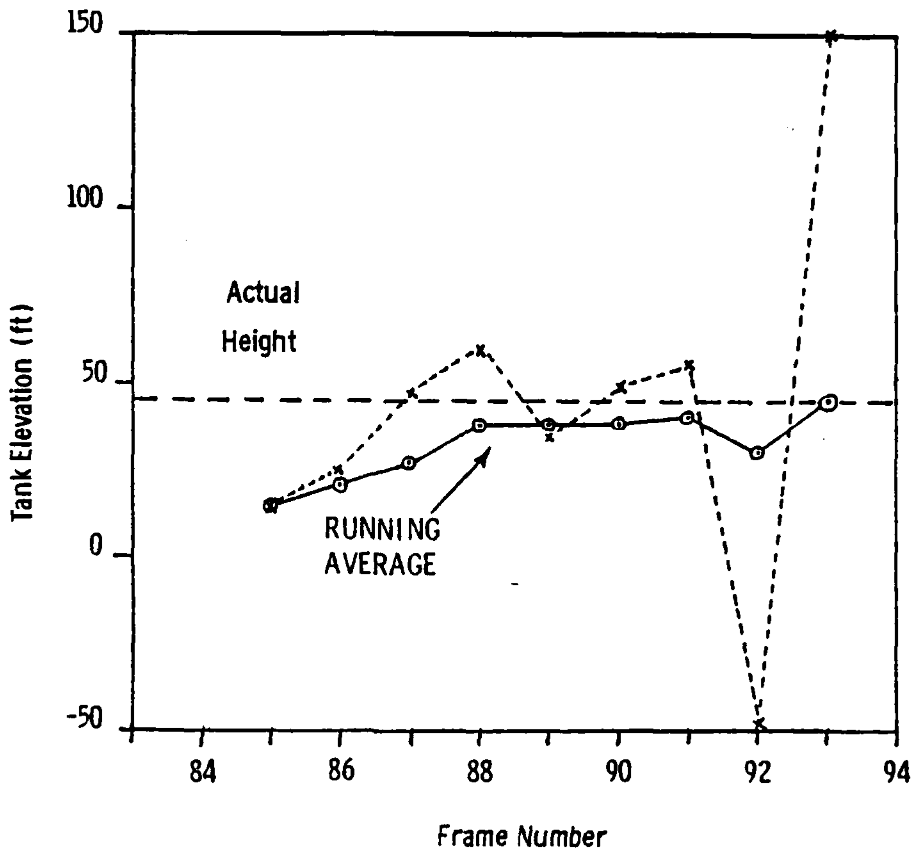


Fig. 8. Interframe determinations (x) and running average (o) of tank elevation (ft) over a ten-frame sequence of imagery from the Fuel Dump Scene.

running average and converges to the actual height of the tank at the end of the sequence. The result of passive height measurement using terrain simulation imagery was accurate to within five feet from a height of 2500 feet, or 0.2% of flight altitude, better than the sensor resolution.

3.2.3 Synthetic TSC Imagery of Hughes Culver City

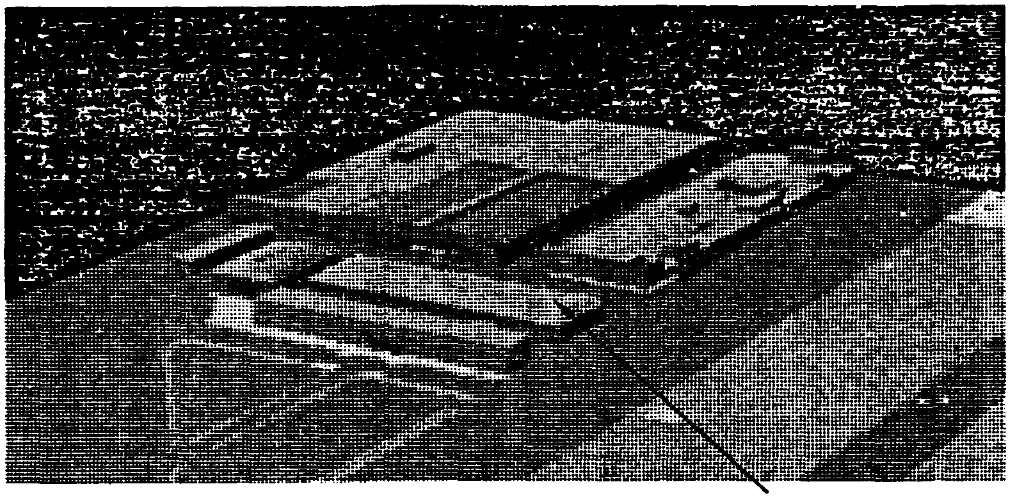
The second data base consisted of "synthetic" imagery of the Hughes Culver City facility, prepared by TSC. In contrast to the real imagery of the first data base, the second data base was generated mathematically using geometric perspective transformations and a computer model of the ground elevation distribution. Grey levels for the latter data base were assigned on the basis of a simple illumination model, and thus, the imagery was characterized by surfaces which contained no contrast or texture.

The results of a sample calculation of object depth are presented in Fig. 9 for a point on the top of building No. 2 of the Hughes Culver City facility. (Although it is difficult to see from the reduced size of Fig. 9, the tracked point corresponds to a corner.) The running average of the elevation determinations over a sequence of twelve frames (cf. Eq. (9) is displayed in Fig 9. The final value obtained for height, using an 11 x 11 tracking window, is seen to be 29.5 ft. This value is to be compared with the known elevation of 31.5 ft, obtained from a wire-frame plan view that was supplied by TSC with the imagery. Although the error is ~ 2.0 ft, it should be remarked that some of the error may be due to the 1.5 ft/resolution of the wire-frame printout. Thus, the sensed elevation accuracy is less than ~ 0.2% of the vehicle flight altitude, or a factor of two better than the vertical resolution of the sensor, 1/256.

3.2.4 ERIM Sunnyvale Imagery

During the August 1978 flight over the Sunnyvale area, Northrop provided ERIM with a TEAC video tape recorder for obtaining imagery from the airborne TV camera (Sanyo 1620X) used as a viewfinder. Approximately two hours of flight imagery was obtained at various altitudes, depression angles, and focal lengths.

Fig. 10 shows three frames (1, 20, 40) from a digitized down-looking sequence of frames taken over the Ames area from an altitude of 1000 feet. Nine DMA survey stations in the Ames area were observed during this sequence and are annotated in the figure. The actual ground speed during this pass was 105 knots, or 180 feet per second, and 160 frames of (30 frame/second) imagery were recorded. By sampling and digitizing one out of four frames, a ground speed of 720 feet per second for a 30 frame/second is simulated in the 40-frame digitized sequence. The forward motion per frame, 24 feet, is shown in the side elevation of Fig. 11. Also shown is the relative position



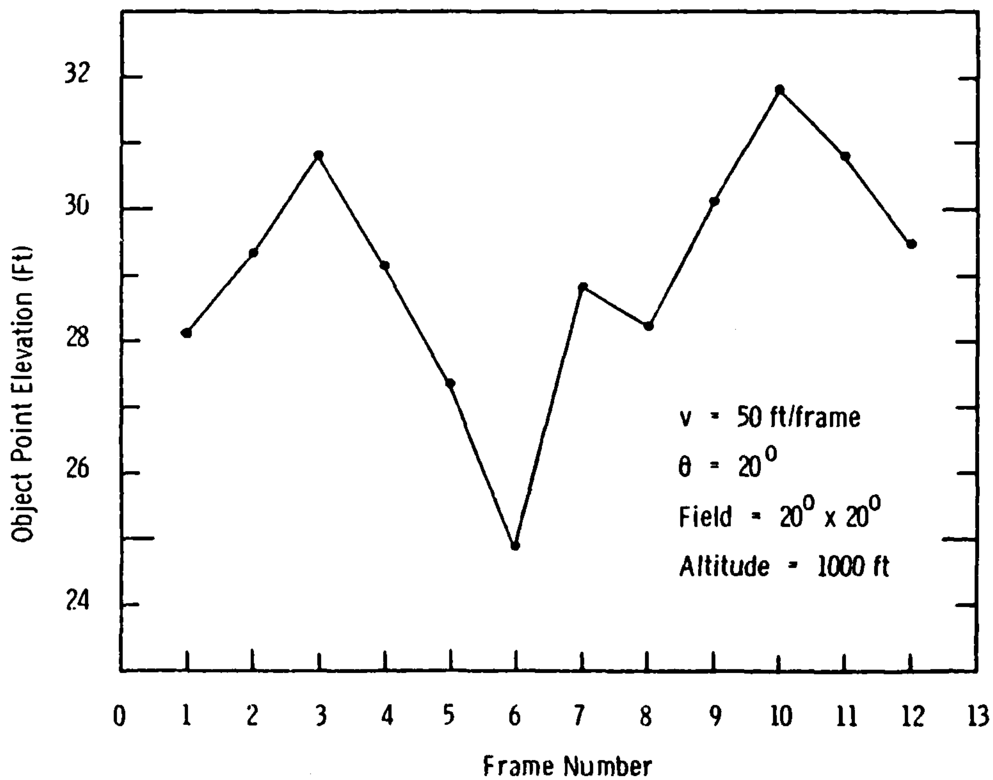
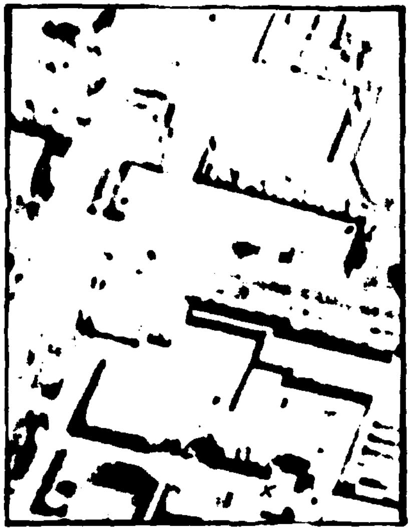
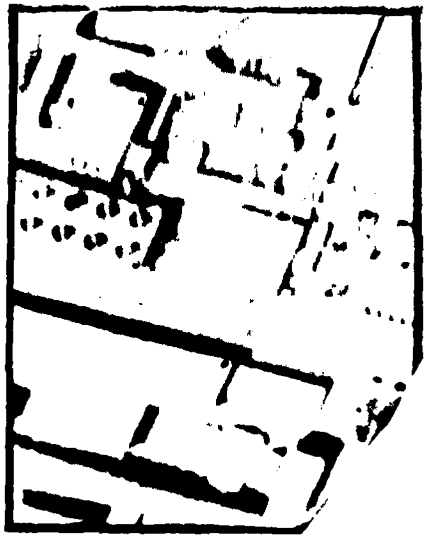


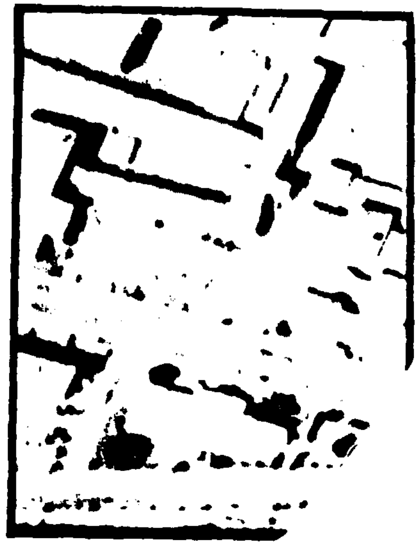
Fig. 9. Sample frame of the synthetic Hughes Culver City Data supplied by TSC, with results of the statistical determination (running average) of the height of a point (shown designated) on Building 2.



Frame # 1

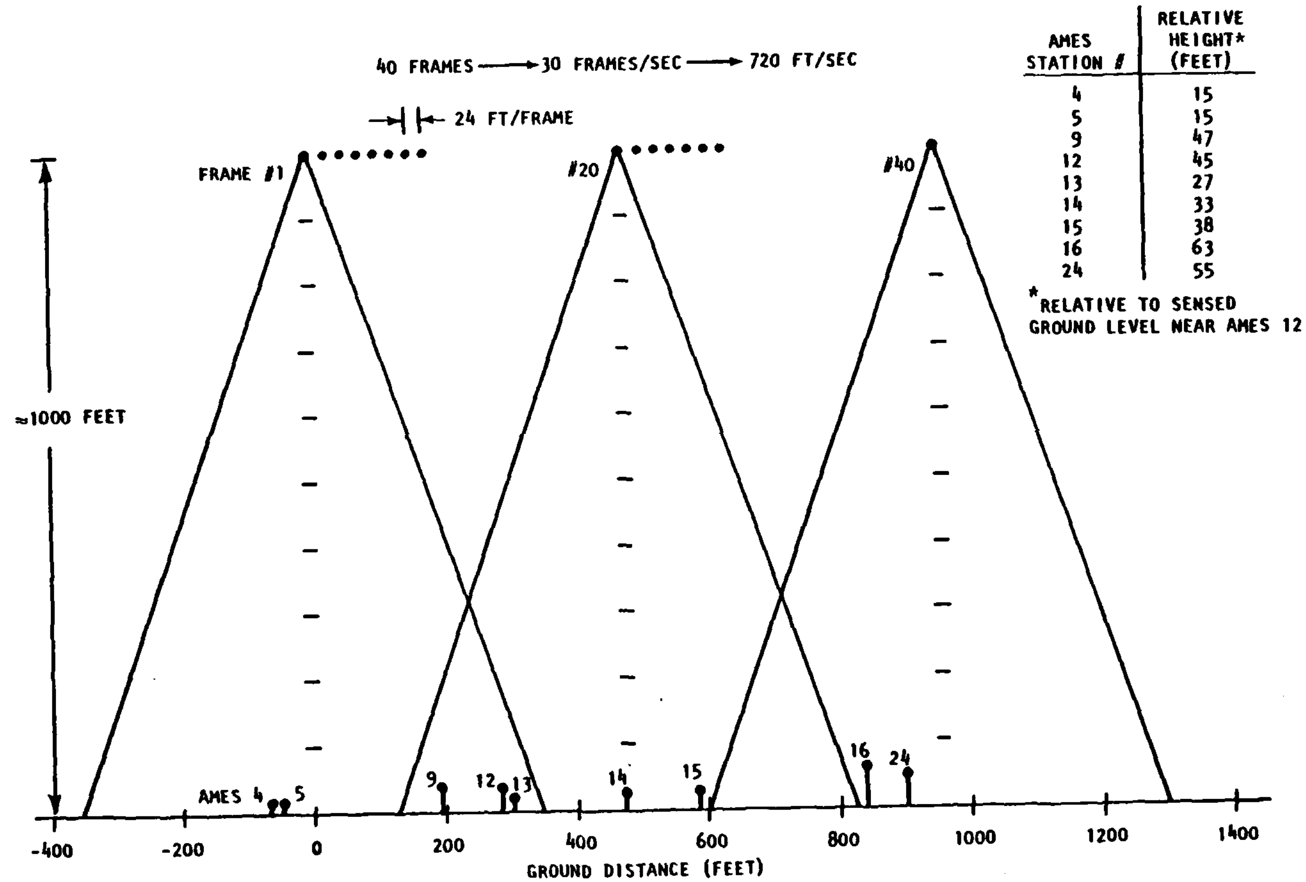


Frame # 20



Frame # 40

Fig. 10. Three sample frames (1,20,40), digitized from a down-looking sequence of imagery obtained by ERIM during flights over the Ames-Sunnyvale area, altitude 1000 ft.



AMES STATION #	RELATIVE HEIGHT* (FEET)
4	15
5	15
9	47
12	45
13	27
14	33
15	38
16	63
24	55

*RELATIVE TO SENSED GROUND LEVEL NEAR AMES 12

Fig. 11. Geometry of the ERIM overflight imagery of the Ames-Sunnyvale area, with vehicle location over the frame sequence indicated.

along the ground track of the nine Ames stations surveyed by DMA. Fig. 12 is a plan view plot of the nine Ames stations showing their location relative to Ames 1 (this station is located on the roof of a 33.5-meter high tower atop Building N242 on Moffett Field Naval Air Station) in local retangular space coordinates.

The flight path is 160^{0} relative geodetic North. Image plane coordinates (x,y) and frame number are shown for each station when it first appears in the sequence. Height values are shown for each station relative to Ames 12. Ames 12 is 48.07 feet below the local DMA origin Ames 1.

Because of the large separation between stations, not all stations crossed the entire field of view, limiting the number of frame pairs for some measurements. The calculated depth values obtained from stereo processing for each of the Ames stations are summarized in Table II.

Table II. Ames Sunnyvale Data.

Ames Station #	Number of Frames Tracked	Sensed Depth
4	12	960
5	11	951
9	22	972
12	26	957
13	26	975
14	25	961
15	26	963
16	19	957
24	23	1029

Since the sensor was unstabilized, roll, pitch, and yaw variations during the actual 5 1/3 second (180 feet/second) flight influence the

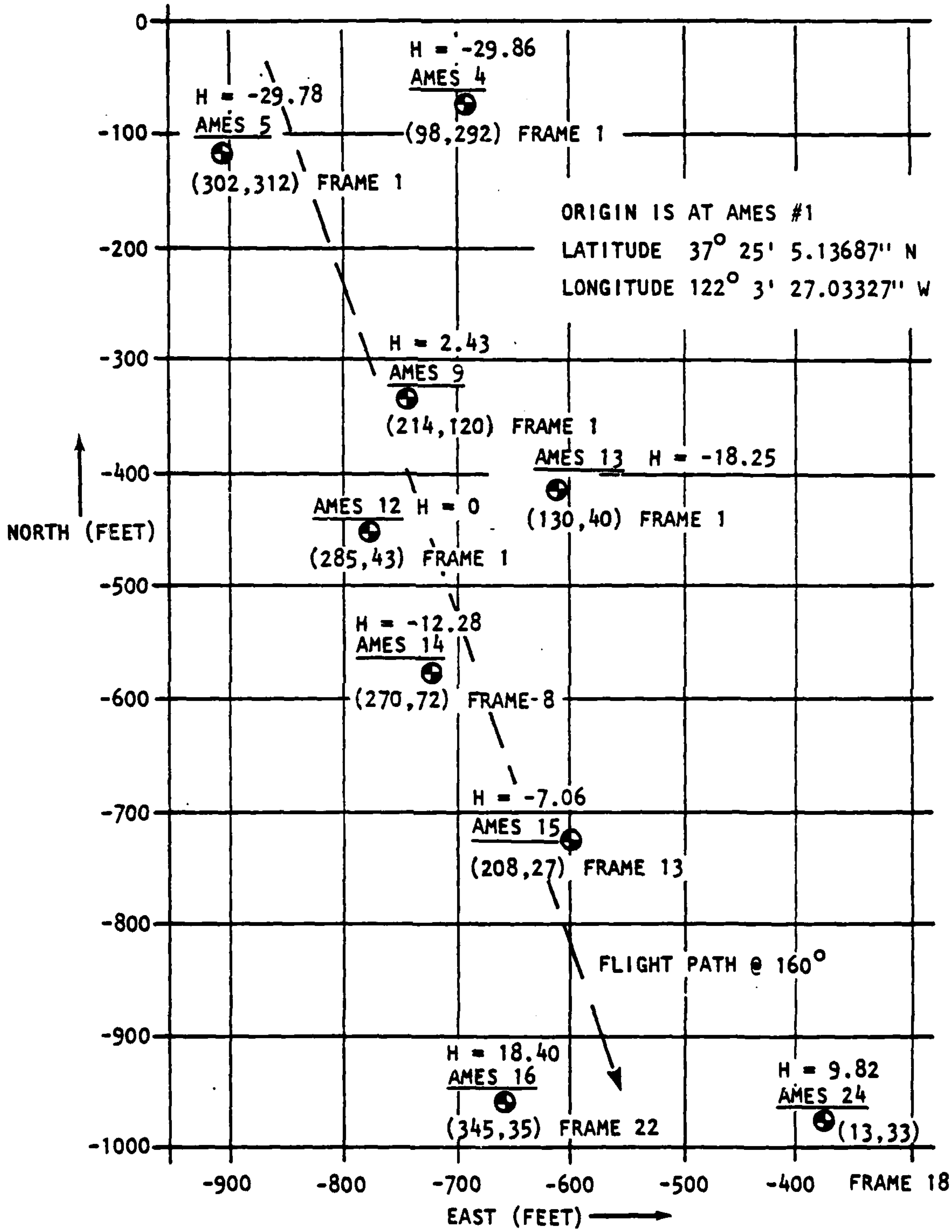


Fig. 12. Plan view of nine Ames stations showing their location relative to Ames 1, located on Building N242 of the Moffett Field Naval Station.

depth. However, even with the unstabilized data and the assumption of a constant attitude for the sensor, RMS height error for the nine stations was 1.7% of flight altitude. The data showed an average sensor tilt from true vertical of 4.7^{0} forward and 6.5^{0} roll to left. The 1.7% value indicates the possibility of even using an unstabilized sensor for a 200-300 ft altitude flight.

A reference height map was prepared from the fourth test set data provided by ERIM. The planimetric view range image in file 6 of tape #1 was selected. This image is listed as a down looking reference for scene matchers. It is an array of size 110 x 150 representing height samples in the vicinity of Ames 14 based on measurements obtained by ERIM with the 1.06 μ laser sensor. The sampling interval is approximately six feet in both the down track and cross track dimensions. This image is not exactly a planimetric view but is close enough to the vertical (viewing angle is tilted 1.2^{0} to the North and 4.0^{0} to the east of the vertical) to use as a plan view reference. Fig. 13 shows the file 6 image from tape #1 of the fourth test set.

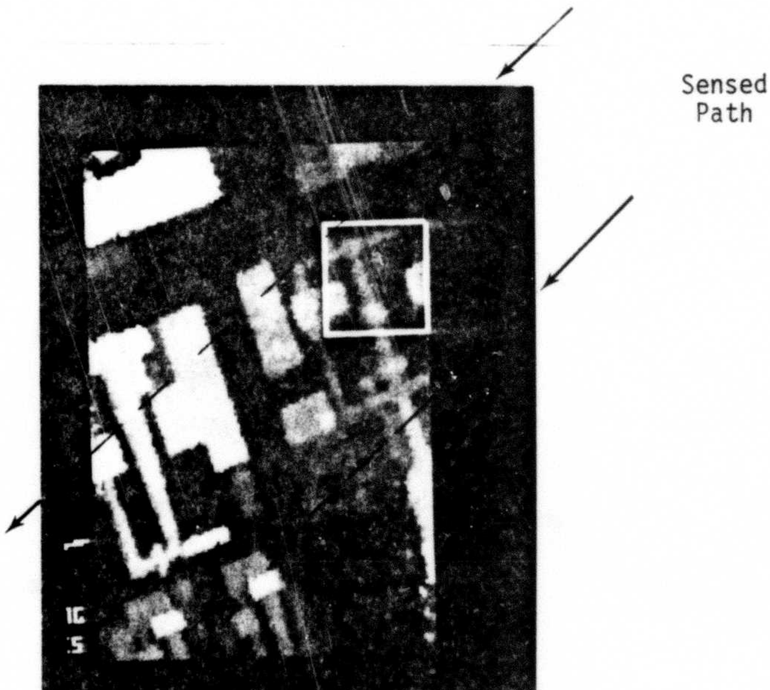


Fig. 13. File 6 Range (Height) Reference Image.

The box shows a portion of this image in the vicinity of Ames 12 where a sensed height array was obtained. The direction of the sensed path is shown. Address rotation of file 6 through 249^{0} CCW was used to provide rotation of the reference to the direction of the sensed path as would normally be provided by the heading gyro in a missile navigation system. A 49 x 49 array from this rotated reference was selected as the reference height map for cross correlation with the passively sensed height map of the area.

Fig. 14 shows the intensity image of the region processed for a sensed height map. The crosses represent the points in Frame 1 selected for processing to obtain the sensed height map. The points are separated

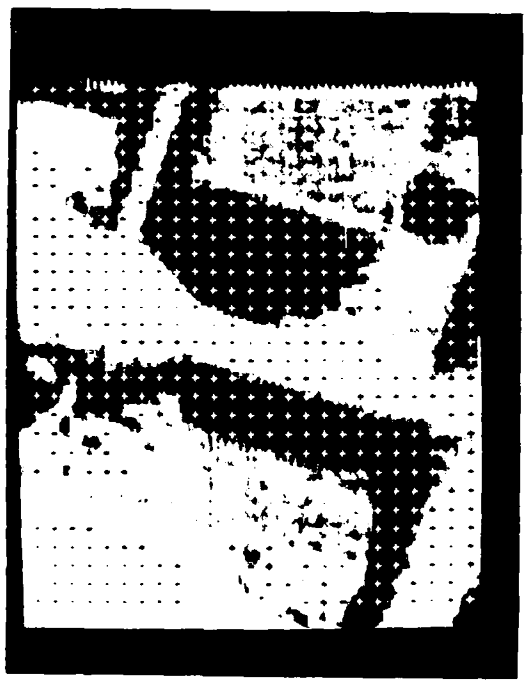


Fig. 14. Intensity Image of Ames Region Processed for Sensed Height.

about 5.5 feet apart in each direction on the ground and correspond to an increment of five pels in X and four pels in Y. The 5/4 ratio results from the product of the aspect ratio of the TV camera's field, 4/3, and the ratio of the number of samples in the X vs Y direction, 480/512, of the field of view. The sensed height map is 175 feet on each edge.

Fig. 15 shows the height array values obtained from processing the flight video tape image sequence as the points crossed the field of view from a flight altitude estimated at 1000 feet. The numbers are relative height above the local ground level in the region. This ground level was sensed at 1002 feet below the aircraft. Points not measured are noted by zero in the sensed map.

As can be seen by inspection of Fig. 14, this sensed height is also not a planimetric view. Since the height values are ground referenced to the viewing aspect near the top of the first frame, this sensed height map is rotated from the vertical by about 1/2 of the 40^{o} forward viewing angle of the camera. Although this corresponds to some 20^{o} angular offset from the reference, this known correction was not made in the data before initiating a sensed/reference height match.

Fig. 16 shows the normalized cross correlation matrix resulting from this initial match. The location of best fit (2,0) corresponding to an offset of two pels in the height array. Since registration was based upon visual match of the Ames 12 coordinates in both sensed and reference map, we would have expected best fit at (0,0).

Although a location error of two pels is greater than desired, we are encouraged by this initial test of the form fit approach to autonomous target acquisition considering that unstabilized flight sensor imagery and an angular offset of some 20^{o} were present. We believe that ground location accuracy will improve by use of a sensed array larger than 32 x 32, stabilization of the imagery, correction of the planimetric view offset, and by down looking from altitudes less than 1000 ft.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1:	0	0	0	9	0	-9	1	34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	0	0	19	0	19	9
2:	0	0	0	-1	0	3	20	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	10	8	10	13	17	0
3:	0	0	0	-2	0	-2	14	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	14	15	14	13	21	20	9
4:	0	0	0	0	-3	-5	15	0	0	26	0	0	0	0	0	0	0	0	0	0	0	0	0	2	13	10	9	11	34	0	0
5:	0	0	0	0	0	-19	24	0	0	47	0	0	0	0	0	0	0	0	-70	0	0	0	23	18	19	10	16	12	0	0	9
6:	0	0	0	-6	-6	0	17	0	0	42	46	49	0	0	0	0	0	0	0	0	0	0	9	13	13	15	17	20	0	0	0
7:	3	0	0	-5	0	12	38	0	-21	38	0	0	41	0	0	0	0	0	0	0	0	3	11	6	4	7	18	23	0	0	33
8:	17	26	22	22	21	27	0	0	15	31	0	0	0	0	0	0	0	0	0	0	0	15	0	11	0	11	22	0	30	43	41
9:	30	0	33	35	33	37	33	23	22	0	0	0	0	0	0	0	0	0	0	0	36	18	0	7	4	17	0	19	0	20	0
10:	42	0	45	39	29	30	36	22	16	0	0	0	0	0	0	0	0	0	23	0	34	0	2	2	5	0	0	0	0	0	0
11:	37	30	46	43	36	33	32	29	0	0	3	0	0	0	0	0	23	29	13	20	33	0	0	0	0	0	0	0	0	0	0
12:	0	37	39	40	35	33	0	20	17	0	-2	0	0	0	0	-4	22	21	4	21	0	0	0	0	26	-8	0	0	0	0	0
13:	31	37	36	39	32	0	31	0	23	0	0	0	0	0	17	19	17	12	4	17	0	0	0	0	19	3	0	4	0	0	0
14:	25	37	34	35	26	30	0	0	0	16	27	0	0	0	0	14	10	6	7	11	0	0	0	5	0	0	0	0	0	0	9
15:	32	32	27	20	21	20	0	0	0	0	0	0	0	0	0	20	17	6	6	10	13	0	0	11	0	0	0	0	0	0	9
16:	27	29	27	20	10	0	16	0	19	0	0	0	0	0	14	0	0	24	18	20	26	22	20	21	27	0	0	0	0	0	0
17:	22	26	20	23	0	0	6	8	10	0	0	0	0	0	20	0	9	0	0	20	21	10	23	26	27	24	25	0	21	17	0
18:	15	24	20	22	12	12	17	14	15	15	0	0	0	0	0	0	0	0	4	0	27	0	0	20	30	34	31	31	25	30	0
19:	15	22	25	16	14	12	0	0	11	16	0	0	0	0	0	0	0	0	0	0	20	0	34	35	37	37	35	31	25	27	0
20:	13	19	13	11	6	7	0	0	15	9	0	0	0	0	0	0	0	0	0	0	0	-20	20	32	53	34	38	34	35	39	0
21:	16	20	17	10	0	10	0	0	0	0	0	0	0	3	0	0	0	0	0	0	26	20	25	34	30	40	36	30	29	32	0
22:	24	17	13	10	12	12	0	19	0	0	0	0	0	6	0	0	0	0	0	16	10	25	0	34	0	0	34	23	22	19	0
23:	0	15	16	10	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	12	0	0	0	0	0	31	22	21	0	0
24:	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	3	0	15	0	0	0	22	22	0	0	4
25:	0	0	0	0	0	13	0	0	0	0	0	0	0	1	0	0	0	0	0	3	8	4	0	0	0	36	15	16	15	8	1
26:	0	0	0	0	0	0	0	51	0	-1	0	0	4	6	4	0	0	0	-2	0	7	0	0	0	19	12	7	0	2	0	-3
27:	0	0	0	0	0	0	0	0	0	31	0	0	12	10	4	2	3	2	5	7	19	0	0	0	2	15	13	0	0	0	0
28:	0	0	0	0	0	0	0	0	0	20	0	0	21	15	12	9	0	9	12	7	0	0	0	0	39	-1	19	15	-1	0	9
29:	0	0	0	0	0	0	0	0	0	27	-19	0	0	14	12	12	17	20	10	17	0	0	0	0	33	-4	5	7	1	-16	-1
30:	0	0	0	0	0	0	0	0	27	14	10	0	0	0	7	13	13	13	3	0	0	0	0	0	25	0	0	8	7	0	7
31:	0	0	0	0	0	0	0	0	29	10	11	12	29	0	0	0	0	15	-4	0	0	0	0	0	15	21	10	11	8	0	16

Fig. 15. Sensed height map in the vicinity of Ames 12.

	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9
-9:	0.817	0.811	0.803	0.803	0.801	0.804	0.807	0.819	0.816	0.822	0.824	0.832	0.837	0.842	0.844	0.847	0.852	0.853	0.860
-8:	0.834	0.825	0.810	0.814	0.805	0.805	0.807	0.816	0.821	0.820	0.821	0.836	0.839	0.842	[illegible]	0.850	0.855	0.854	0.858
-7:	0.835	0.829	0.831	0.824	0.821	0.824	0.823	0.831	0.831	0.826	0.827	0.837	0.839	0.844	0.845	0.853	0.857	0.856	0.858
-6:	0.839	0.831	0.828	0.830	0.829	0.828	0.835	0.833	0.844	0.842	0.836	0.845	0.849	0.849	0.847	0.855	0.859	0.855	0.854
-5:	0.837	0.827	0.828	0.824	[illegible]	0.829	0.829	0.836	0.837	0.846	0.847	0.856	0.857	0.856	0.854	[illegible]	0.859	0.858	0.849
-4:	0.820	[illegible]	0.810	0.817	0.819	0.820	0.827	0.839	0.838	0.843	0.851	0.860	0.865	0.862	0.861	0.863	0.861	0.856	0.854
-3:	0.814	0.818	0.805	0.810	0.808	0.810	0.815	0.833	0.835	0.838	0.856	0.860	0.866	0.867	0.867	0.866	0.857	0.858	0.849
-2:	0.813	0.808	0.809	0.799	0.804	0.803	0.806	0.818	0.834	0.841	0.852	0.860	0.871	0.871	0.870	0.865	0.864	0.854	0.847
-1:	0.807	0.799	0.792	0.791	0.800	0.796	0.802	0.820	0.826	0.840	0.861	0.866	0.864	0.871	0.865	0.869	0.866	0.857	0.852
0:	0.804	0.823	0.793	0.792	0.786	0.794	0.800	0.823	0.834	0.846	0.859	$\boxed{0.876}$	0.873	0.870	0.869	0.866	0.866	0.859	0.851
1:	0.807	0.806	0.801	0.807	0.795	0.793	0.805	0.823	0.831	0.845	0.859	0.867	0.875	0.874	0.867	0.865	0.862	0.858	0.857
2:	0.809	0.801	0.803	[illegible]	0.801	0.813	0.818	0.826	0.835	0.842	0.854	0.865	0.872	0.871	0.867	0.856	0.859	0.853	0.854
3:	[illegible]	[illegible]	0.801	0.804	0.803	0.812	0.823	0.832	0.836	0.836	0.844	0.859	0.861	0.860	0.858	0.853	0.854	0.847	0.850
4:	0.805	0.797	0.794	0.799	0.806	0.819	0.826	0.831	0.834	0.829	0.834	0.848	0.847	0.843	0.844	0.845	0.842	0.837	0.835
5:	0.811	0.809	0.796	0.794	0.797	0.805	0.820	0.824	0.826	0.824	0.826	0.834	0.833	0.827	0.827	0.810	0.829	0.827	0.821
6:	0.825	0.823	0.815	0.802	0.794	0.807	0.815	0.822	0.824	0.826	0.818	0.827	0.827	0.819	0.814	0.812	0.814	0.810	0.815
7:	0.815	0.820	0.823	0.815	0.816	0.824	0.822	0.822	0.821	0.811	0.808	0.813	0.814	0.808	0.810	0.811	0.810	0.805	0.802
8:	[illegible]	0.827	0.814	0.821	0.824	0.831	0.835	0.829	0.829	0.815	0.799	0.802	0.802	0.797	0.794	0.799	0.803	0.798	0.795
9:	0.842	0.830	0.818	0.817	0.820	0.833	0.835	0.834	0.824	0.820	0.805	0.802	0.796	0.792	0.790	[illegible]	0.799	0.793	0.791

Fig. 16. Normalized product correlation matrix of the sensed and reference elevation distribution in ground coordinates.

The first frame of a 40-frame sequence of digitized forward-looking imagery obtained by ERIM is shown in Fig. 17, with a 2:1 enlargement displayed in Fig. 17(b). The parameters for this data base were given in Table I. A portion of the grid also falls upon the adjacent ground area. The points of intersection defined by this grid were used for correlation tracking and the results of height determination are summarized in Fig. 18. Lockheed Building 104 is shown, with a coarse grid superimposed on the roof, penthouse, and tower area of the building.

To compare the passively sensed height values with the Defense Mapping Agency (DMA) survey of ground truth, the values were grouped into four regions for which a DMA survey height is known. These regions are identified by a dashed line boundary in the array of Fig. 1 and are labeled tower, penthouse, roof and ground. Each value in the array was obtained by block tracking on 11 x 11 size block of picture elements (pels) in the image over a sequence of some 30 to 40 frames during which the block remained in the field of view of the sensor. Height measurements were not obtained at 16 of the 126 positions in the array due to loss of track or failure to track because of an absence of contrast in that region of the image.

Height variation within a region having the same elevation, i.e., penthouse can be due to many factors contributing to measurement noise. Sync error and/or noise sources include the sensor, tape recorder, disk recorder and A/D converter used to obtain the image in digital form. The major factor, however, is believed due to uncompensated attitude variations (pitch, roll and yaw) of the aircraft during the flight pass. The sensor, a Sanyo 2/3" vidicon camera, was hard mounted to the airframe. Attitude variations during the flights can be seen upon imagery playback. Instrumentation printout from the INS platform showed peak swings of $\pm 2^{0}$ during some flights. Because of the above mentioned height variations, the results for passively sensed height in the first column of the Table in Figure 19 are the <u>average</u> for each region.

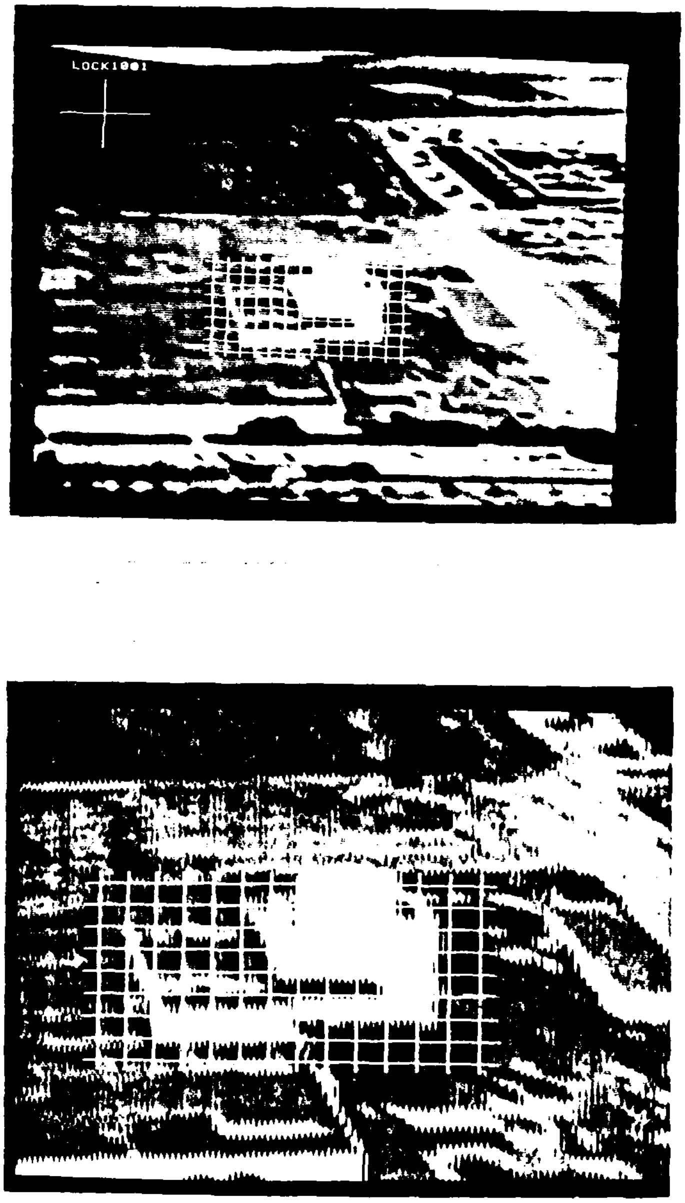


Fig. 17. Initial frame (and a 2:1 enlargement of the central region) of a 40-frame sequence of forward-looking imagery of Lockheed Sunnyvale area taken by ERIM.

- SENSED HEIGHT ARRAY -

	X = 125													255
Y = 200	45	127	150	-	-	-	-	75	143	161	128	47	89	54
										Tower				
	170	57	-	105	40	135	146	63	-	-	183	67	75	63
	-	235	114	98	114	31	57	102	57	60	264	29	140	66
										Not Surveyed				
	148	82	146	214	161	-	97	189	-	90	-	88	60	51
	-8	122	69	81	107	71	59	43	-	36	-	-	80	198
	-14	96	104	91	103	95	101	63	59	61	69	76	64	164
					Penthouse									
	96	129	108	88	89	89	132	62	90	92	92	92	48	48
	137	149	75	73	74	80	109	80	118	94	114	114	107	37
				Roof							Ground			
Y = 280	2	166	101	111	78	80	80	87	67	-	25	27	39	-

Fig. 18. Sensed height array from forward-looking Lockheed imagery.

Location on Bldg 104	Passively Sensed Height, Ft	DMA Survey Height, Ft	Δ Survey Ft	Δ Ground Ft	% $\hat{H}$ (Δ Survey)	% $\hat{H}$ (Δ Gnd)
Tower	144.00	115.48	28.52	3.47	4.2	0.51
Penthouse	109.70	84.30	25.40	0.35	3.7	0.05
Roof	89.59	67.19	22.40	-2.65	-3.3	-0.39
Ground	49.75	24.70	25.05	-	3.7	-
				RMS Error %	3.7	0.37

Fig. 19. Results of passive height determination after image segmentation and smoothing of sensed height array.

The second column shows the DMA survey height for comparison. The Δ survey column shows difference between sensed height and ground truth. The major portion of this difference is due to the offset in ground reference of ≈ 25 feet since the sensed height values are relative to a ground plane estimated from the flight log to be 684 feet below the aircraft. The Δ ground column shows the height error after correcting for this ground offset, since the actual height of the aircraft above sea level was not available. The last two columns show height error as a percent of the estimated flight height. Accuracy in height difference measurement as a percent of flight altitude was 0.37%, RMS for the flight conditions shown. Absolute altitude of the aircraft above ground truth showed an RMS accuracy of 3.7% when compared to an altitude estimate from the flight log (i.e., 2000 sin 20°).

The obvious improvement in sensed height results which occurred by segmentation of the image to obtain averages of height over specific regions illustrates the importance of developing suitable pre-processing and post-processing techniques to augment algorithms for motion vector computation.

3.3 Preliminary Conclusions

The concept of object depth determination by means of motion stereo processing of a sequence of dynamic imagery has been validated by frame-to-frame correlation tracking of discrete image points. Accuracies significantly better than the sensor resolution have been achieved by fractional pixel interpolation and multiple frame processing, which provide the opportunity for statistical refinement by filtering a sequence of derived depth values. An accuracy of ~ 0.2% of flight altitude was obtained using NRTC terrain simulator and TSC synthetically generated imagery, and better than 2% was obtained with unstabilized (real) imagery taken over the Ames-Sunnyvale complex (provided by ERIM).

It is the position of NRTC that a promising approach to the problem of autonomous target acquisition may be direct correlation matching of a sensed and reference depth (height) distribution in ground coordinates.

The current ATHP approach is based upon target acquisition by matching intensity distributions in image coordinates. However, it has been shown that the performance of the latter algorithms can also be significantly improved when range data is simultaneously available. Therefore, the capability for passive range and/or depth determination from motion stereo processing could be exploited in several ways. Inasmuch as the algorithms for image intensity correlation are presently in an advanced state of development, it is anticipated that the first application of passive ranging will be to provide supplementary (sensed) data for these algorithms.

Further work is required to evaluate the performance of the motion stereo techniques as a function of flight geometry, sensor parameters, and image statistics, and to extend them to area-processing algorithms which are suitable for implementation by real-time hardware. However, preliminary estimates to be presented in Section 3.4 for even a straightforward implementation of the present discrete point-tracking algorithms over an image area show that real-time hardware within the present ATHP constraints appears to be feasible.

3.4 Hardware Considerations*

Implementation of the discrete point correlation tracking technique for determination of the image motion vector field over an entire frame is computationally intense. Fortunately, the required computations are very highly structured, which makes it possible to consider a specialized pipeline computer architecture for a viable hardware mechanization.

3.4.1 Basic Computational Structure

An initial estimate of the overall signal processing required for mechanization of the motion stereo algorithm is shown in the block diagram of Fig. 20. A digital deblurring operation is performed if uncompensated image motion causes excessive blurring of the image, which may occur for down-looking operation with a high velocity platform.

---

* Preliminary investigation of processor design was performed by J. J. Reis under an on-going Northrop Corporation independent research and development (IR&D) program.

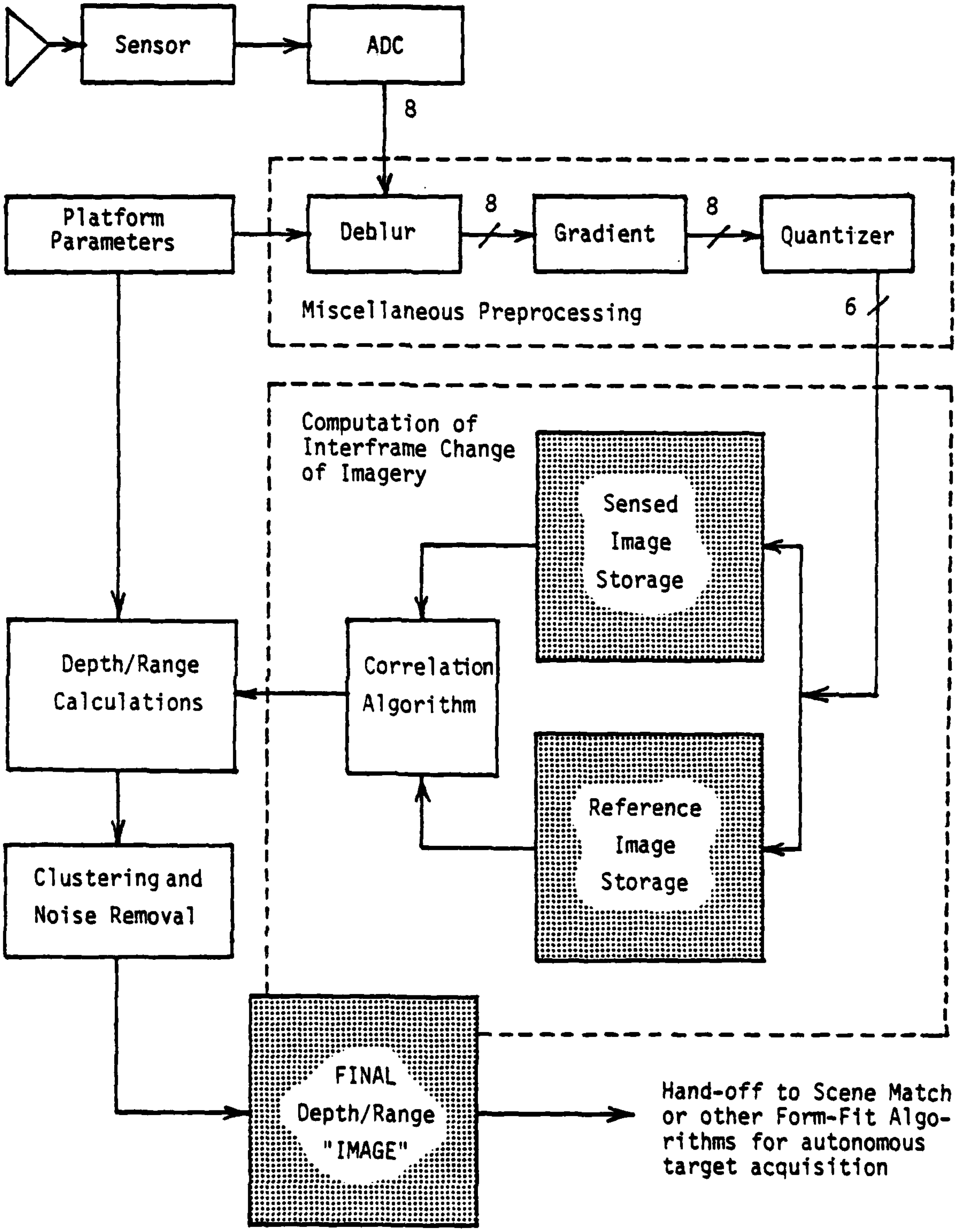


Fig. 20. Computational block diagram for passive range/depth determination obtained from motion stereo processing dynamic imagery.

The magnitude of the gradient of the deblurred imagery is (nonlinearly) quantized and stored in either the sensed or reference image memory storage as required. The correlation algorithm (normalized product correlation or minimum absolute difference) (MAD) is used for computation of the image motion vector field for every pel of the sensed image. Computationally, the MAD algorithm may be preferable for hardware implementation, and we shall assume that it has been selected in the following discussion.

It is expected that a variety of preprocessing techniques may be required to restrict the algorithm to image points characterized by edges, corners or objects of interest. Generally speaking, the successful implementation of the tracking algorithm requires that the image points be located in regions with good local gradient statistics, so a gradient block has been inserted in Fig. 20 to schematically represent some sort of edge preprocessing. The vertical component of the motion vector field is used to calculate the relative range/depth of each image point. Finally, the output would consist of a temporally filtered two-dimensional array representing the object scene depth distribution (as a function of ground coordinates) or a range image (in image coordinates) representing the range to each pel at any instantaneous vehicle position. For application of the technique to three-dimensional form fitting, depth distribution is required, while for augmenting scene intensity matching algorithms currently under development, range imagery is required. In anticipation of the necessity that the results may have to be post-processed to remove noise and/or to compensate for certain anomalies that may be inherent in the algorithms, a schematic block representing a final clustering or other spatial filtering operation on the depth or range distribution has been included in Figure 20.

3.4.2 Preprocessing Techniques

In the following discussion, we shall concentrate on description of the key hardware items: the MAD correlator, the depth/range calculations, and the clustering/noise removal processes. The deblurring, gradient, and quantization processes are well understood operations whose hardware mechanization involves relatively little risk. While such preprocessing operations are important and must be thoroughly investigated in the initial software algorithm development, they are not expected to materially affect overall hardware complexity or feasibility and will not be discussed in further detail. NRTC has had extensive experience in the design and construction of real-time hardware for spatial filtering, edge extraction, and image area convolution.

3.4.3 Parallel Pipeline Architecture

The correlation processing of an entire frame involves the computation of the MAD metric function $M_{ij}(I,J)$ defined by Eq. (14) for every point (i,j) in the image for a set of shift vectors (I,J) defined over some region. If a $(2L+1) \times (2L+1)$ search area centered about some predicted estimate is attempted, the memory requirement is $(2L+1)^2$ image-size arrays, which clearly becomes prohibitive for realistic image sizes. (For example, a search over a region which is centered $\pm$ 2 pels about an estimated location requires 25 image-size memory storages.) In a realistic system, therefore, it will be essential to take maximum possible advantage of any techniques that may be available for intelligently reducing the search region defined by the shift vectors (I,J). For example, if good inter-frame predictions of image motion can be made because of high platform stabilization, then it may be possible to limit the search to a narrow range of shifts localized in the vertical Y-direction only.

If vertical correlation is sufficient, $J = 0$. Furthermore, if the motion induced image point line-shifts can be predicted (frame-to-frame) with similar high accuracy, then it may be possible to limit I to a small range of values. For example, if platform altitude is 200 ft, and objects in the field of view are not higher than 50 ft, the magnitude of image point line-shifts over the entire frame will vary by no more than 150/200. Thus, for example, if parameters are chosen in such a way that nominal interframe line shifts are $\sim$ 10/frame, it may be possible to limit I to a range (0,1,2,3). (Furthermore, if parameters are such that the range of shifts originated from the dependence of shift on line address (cf. Eq. (4) causes greater ranges in I, the image could be segmented in the vertical direction.) Therefore, it shall be assumed that a range I = (0,1,2,3) may be reasonable for preliminary design estimates. The rectangular window W will be assumed to be (nominally) 5 pels in the horizontal extent and 11 lines in vertical extent (cf. Eq. (17)), leading to a 55-pel^2 area convolution for each image point correlation.

With the above considerations in mind, the MAD algorithm requires the determination of the minimum of four correlation arrays $M_{ij}(I,0)$ for every image point (i,j), for I = 0,1,2,3. For maximum throughput, the hardware would consist of four parallel "pipes", each pipe mechanizing the computation of one of the functions

$$M(I,0) = W * P(I,0),$$

I = 0,1,2,3, as shown in Fig. 21. The hardware associated with each pipe is shown in Fig. 22.

Further hardware reductions could, perhaps, be realized by multiplexing one pipe among the four paths, with an attendant four-to-one reduction in throughput. If it becomes necessary to expand the correlation search range to obtain satisfactory performance of the algorithm, multiplexing (time) and memory (storage) considerations may have to be determined by a trade-off analysis. Clearly, the memory required to store the sensed (S), reference (R), and intermediate correlation (M(I,J)) arrays will dominate the hardware considerations. For a nominal image size ~ 128 x 128, memory for these six main arrays would require ~ 100 kbyte of storage. Although it may be possible to employ some compression techniques, it is apparent that memory storage considerations depend critically upon image size and reduction of search. Future development efforts must therefore include algorithm refinement as well as parameter sensitivity study.

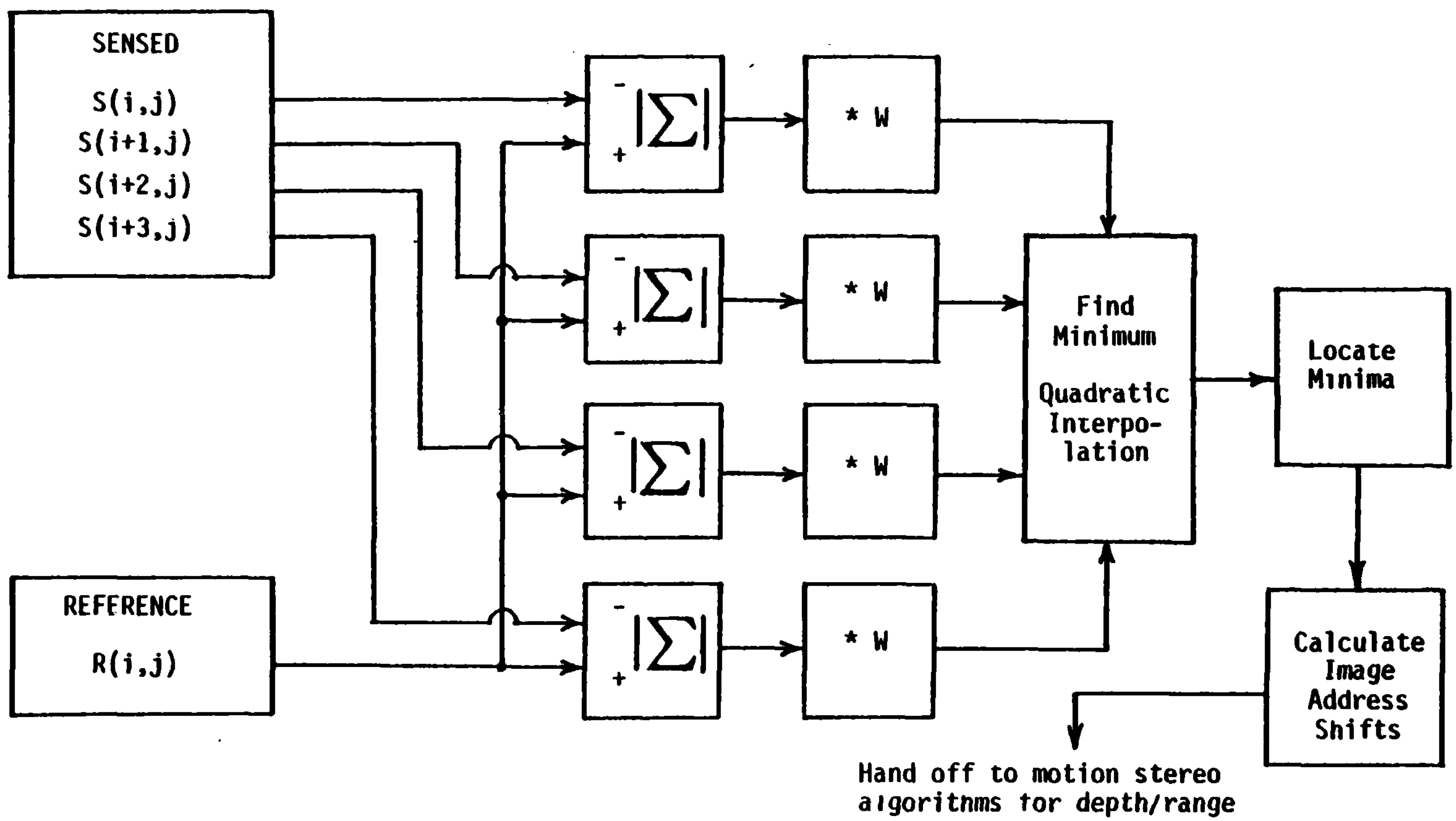


Fig. 21. Parallel pipeline hardware architecture for implementation of the MAD correlation algorithm, assuming that search can be restricted to a limited range in the Y-direction only.

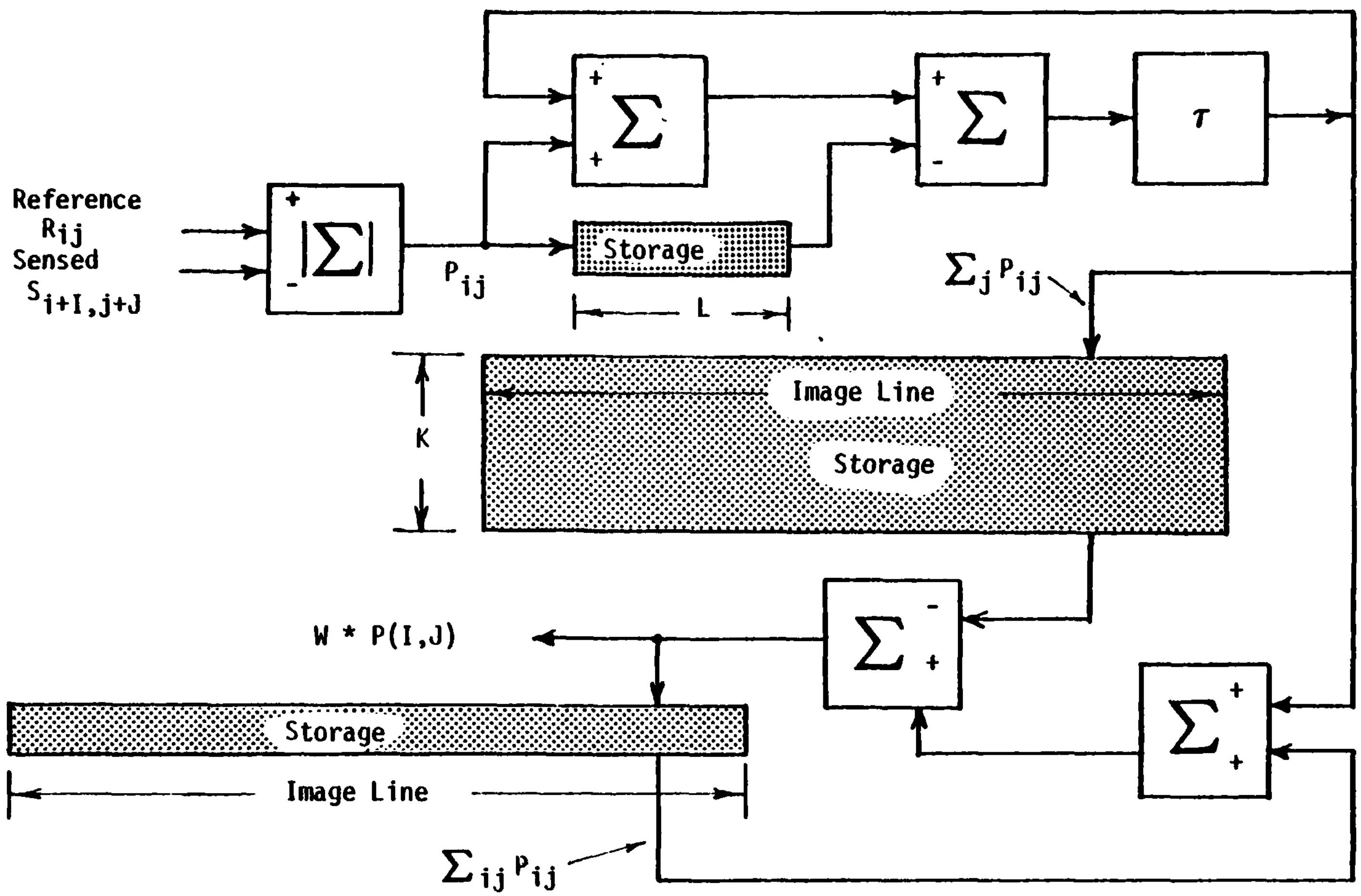


Fig. 22. Hardware mechanization for individual pipe which implements the MAD correlation algorithm for a given relative shift vector (I,J) between sensed and reference images. The convolutional mask W is assumed to be an L x K array of unit elements.

3.4.4 Depth/Range Calculations

Knowledge of the image motion vector field defined over the entire frame (or at least at points of interest) permits depth and range determination for the corresponding object points (cf. Eq. (6)). A schematic hardware mechanization of the depth calculation is shown in Fig. 23. This mechanization may consume excessive power due to the three multiplier cascade. It is hoped that an iterative architecture can be developed to alleviate this difficulty, if it should pose a potential problem. The solution would depend upon required update rates and allowable computational compromises in the depth calculation.

Depth calculations are necessarily noisy due to uncertainties in platform motion, sensor signal-to-noise, image resolution and statistics, and inherent errors in the correlation and other algorithms. The previous software validation of the motion stereo concept employed simple temporal filtering of a sequence of depth determinations to statistically refine the depth estimates. These filters involved a simple running average (cf. Eq. (6)), corresponding to filter coefficients equal to one, and a ratio of two weighted averages (cf. Eq. (9)).

Hardware must be designed to implement the selected filter for smoothing the depth distribution d(i,j) over several frames. Again, memory considerations dictate that the number of frames of storage be kept to a minimum, so a filtering scheme which requires only the present frame and a cumulative average of all prior results would be desirable. Clearly, the simple running average could be implemented in such a way, as could a wide class of simple recursive techniques such as the exponential filter shown in Fig. 24. More advanced smoothing schemes would be conceptually straightforward to implement, but memory storage requirements as well as computation time would be increased.

Results of calculations carried out with the presently available data bases have shown that only a small fraction of the available image points are suitable for reliable depth determination by the methods described in this proposal. Factors such as image gradient statistics, resolution

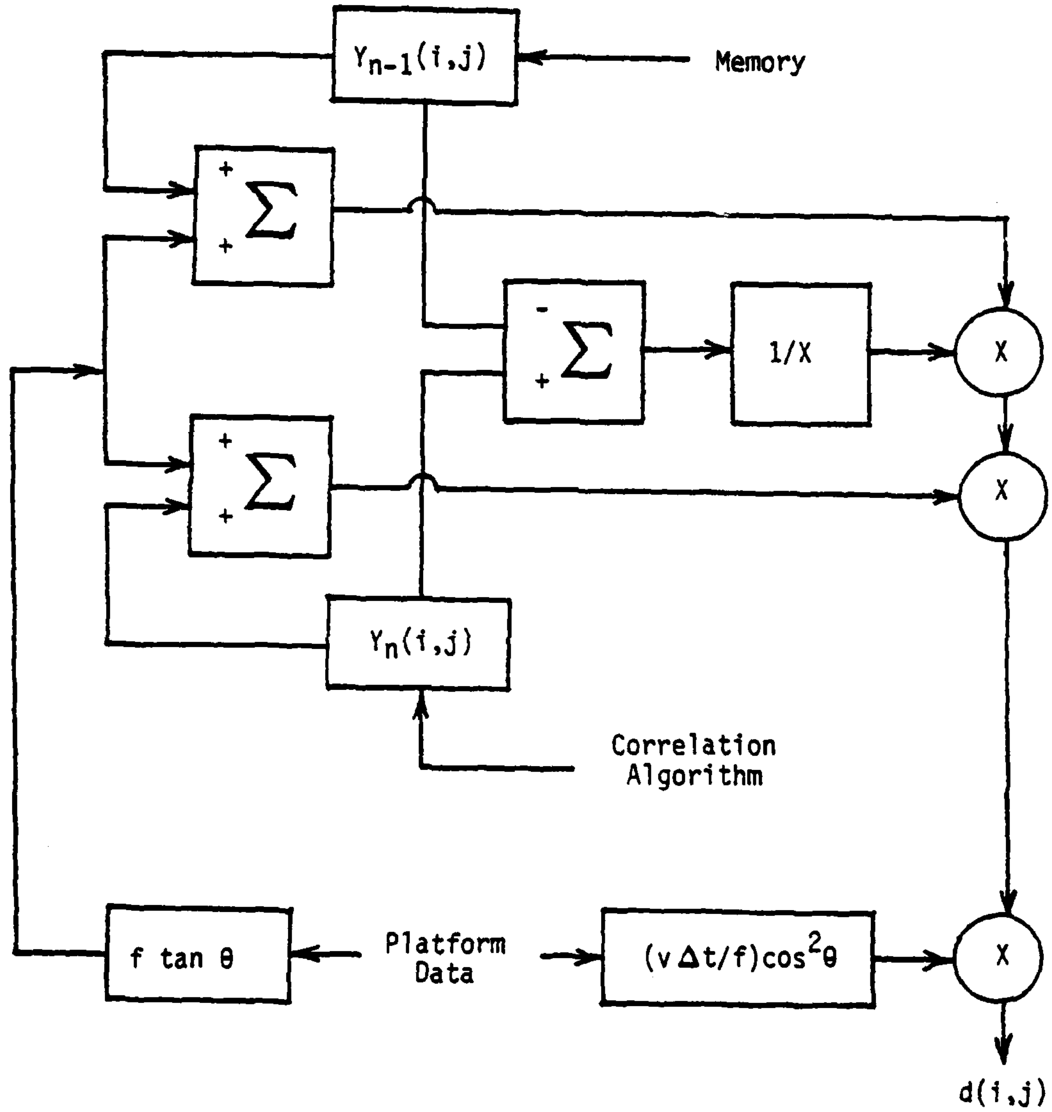


Fig. 23. Pipeline hardware mechanization of depth calculations.

and quantization, and interframe change of geometric perspective can defeat the correlation tracking algorithms presently available. In addition to seeking refinements of the basic algorithms which will reduce these problems, it will be necessary to develop post-processing algorithms for removal of unsuitable points. It is anticipated that some form of clustering and noise removal will be required, but the nature or complexity of the spatial filtering remains to be determined.

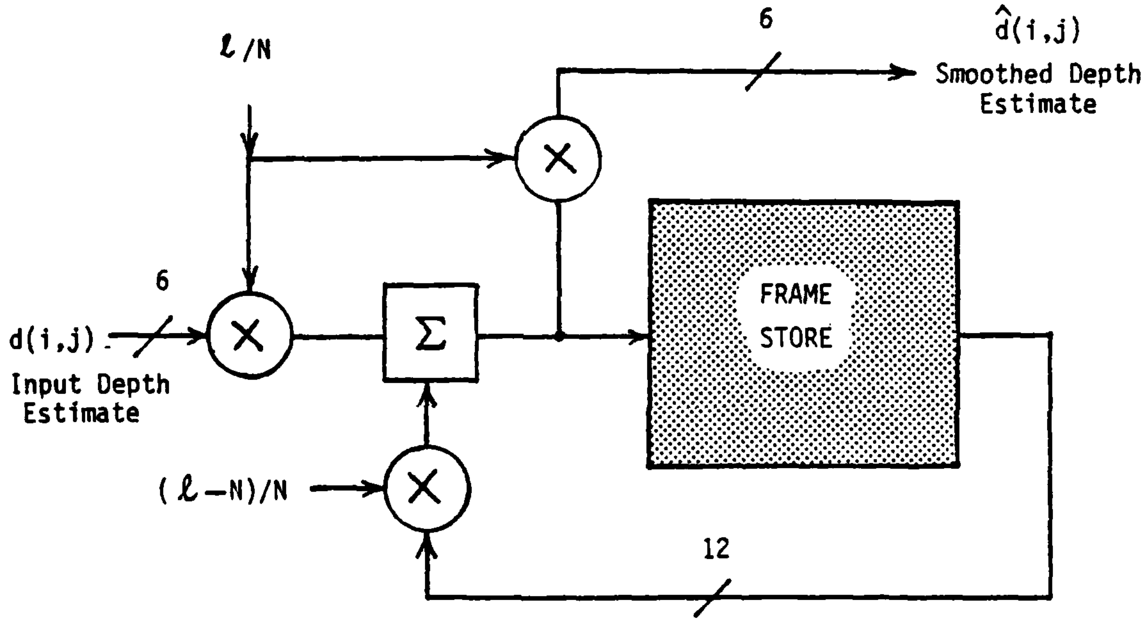


Fig. 24. Exponential filtering of input depth estimate.

4.0 CONCLUSIONS

The capability for passive determination of the three-dimensional form of an object scene by exploitation of motion stereo analysis of dynamic imagery acquired by a moving sensor platform is an attractive concept. Results obtained from a variety of data bases have now validated the concept of passive range and/or depth determination of a target scene, and use of such data could either supplement or provide an alternative to autonomous target acquisition techniques. The present schemes, based upon image processing algorithms for onboard comparison of a sensed scene with a stored replica of a predesignated target area, are currently directed toward image intensity matching for aimpoint determination. Image intensity matching algorithms have been shown to perform more reliably when sensed and reference data is augmented with range imagery. Alternatively, instead of correlating image intensity distribution over the two-dimensional image coordinates, elevation values defined over two-dimensional ground coordinates could be used for match point determination.

The development of practical depth-aided target acquisition techniques requires computationally efficient algorithms for computation of the interframe changes in image point addresses over an entire frame. Motion stereo processing then proceeds by inversion of the known transformation between the image plane and the object scene (which is provided by the camera model) to extract the depth and range information that is implicitly contained in the sequence of dynamic imagery. In addition to algorithms for computation of the image motion vector field, preprocessing and postprocessing of the video imagery may be required, and techniques for spatial and temporal filtering of derived depth and range data must be developed.

For example, real imagery may need to be corrected to reduce problems associated with uncompensated frame motion originating from an unstabilized sensor platform. Furthermore, techniques such as image segmentation or threshold gradient extraction may be required to restrict depth determination to points of high cultural context such as edges or corners.

Use of spatial filters may be useful for enhancing the accuracy of the motion vector field computations and the resulting determinations of depth, and may reduce the correlation area required for frame-to-frame tracking. The necessity for clustering and noise reduction filtering, both spatial and temporal, is anticipated for smoothing the results and obtaining the best statistical estimate of the range and depth distribution. Simple approaches to temporal filtering were illustrated by the running average of a sequence of depth determinations, and to spatial filtering by the image segmentation and averaging described for the Lockheed data base. Clearly, more sophisticated techniques will be required for future optimization of the algorithms.

The validation of the motion stereo concept has been accomplished by computation of the motion vector shifts for discrete points by means of frame-to-frame block tracking. Preliminary investigations indicate that it may be feasible to implement this straightforward approach in a highly-parallel computational architecture that will meet constraints for real-time hardware imposed by the current ATHP application. However, in formulating the preliminary study, several assumptions were made about platform stabilization, vehicle velocity and height, and characteristics of the sensor, imaging, and inertial navigation systems. In addition, there are many fundamental questions which remain to be answered concerning the inherent errors to be expected in the tracking algorithm itself, and the extent to which the present techniques are scene-dependent.

In the immediate future, efforts should continue to refine the motion vector algorithms (correlation tracking, or other approaches), video preprocessing and postprocessing algorithms, and spatial and temporal techniques for filtering of derived data to obtain statistical best estimates of range and depth distributions. In addition, due to the intense scheduling pressure to Jevelop operational real-time hardware, a parallel effort in processor design and emulation should be undertaken to implement the block tracking formulation.

DISTRIBUTION

Final Technical Report --

Commander US Army Missile Command ATTN: DRSMI-RAS/Mr. Hagood Redstone Arsenal, Alabama 35809	1 copy
Director Defense Advanced Research Projects Agency ATTN: STO/Cmdr Thomas Wiener 1400 Wilson Blvd. Arlington, Virginia 22209	1 copy

Special Technical and Final Technical Reports --

Commander US Army Missile Command ATTN: DRSMI-RAS/Mr. Hagood Redstone Arsenal, Alabama 35809	1 copy
Director Defense Advanced Research Projects Agency ATTN: Program Management (MIS Division) 1400 Wilson Blvd. Arlington, Virginia 22209	2 copies
Hughes Research Laboratory 3011 Malibu Canyon Road Malibu, California 90276	1 copy
Boeing Aerospace Co. ATTN: Dr. Richard Nuno P.O. Box 3999, MS 2101 Seattle, Washington 98124	1 copy
ESL ATTN: R. Venti 495 Java Drive Sunnyvale, California 94086	1 copy
RAND Corporation ATTN: Hy Shuman 1700 Main Street Santa Monica, Caifornia 90406	1 copy

DISTRIBUTION

TASC Six Jacob Way Reading, Mass. 01867 ATTN: Dr. Charles Ormsby	1 copy
Lockheed Missiles and Space Co. ATTN: Dr. James Pearson 3251 Hanover Street Palo Alto, California 94304	1 copy
Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2 copies
SRI International ATTN: Dr. Walter Chestnut 333 Ravenswood Avenue Menlo Park, California 94205	1 copy
Grumman Aerospace Co. ATTN: Abe Reich Bethpage, N. Y. 11714	1 copy
Systems Control, Inc. ATTN: Dr. Fred Smith 1801 Page Mill Road Palo Alto, California 94304	1 copy
ERIM P.O. Box 8618 ATTN: Max Blair Ann Arbor, Michigan 48107	1 copy
Rockwell International Missile Systems Division ATTN: Mr. Robert Kin P.O. Box 4182 Anaheim, California 92803	1 copy