SRI International

HIERARCHICAL WARP STEREO

Technical Note No. 402

December 11, 1986

By: Lynn H. Quam

Artificial Intelligence Center
Computer Science and Technology Division

This paper also appeared in the Image Understanding Workshop Proceedings (1984).

333 Ravenswood Ave. • Menlo Park, CA 94025
(415) 326-6200 • TWX: 910-373-2046 • Telex: 334-486

Hierarchical Warp Stereo

Lynn H. Quam
SRI International
333 Ravenswood Avenue
Menlo Park, California 94025

September 10, 1984

Abstract

This paper describes a new technique for use in the automatic production of digital terrain models from stereo pairs of aerial images. This technique employs a coarse-to-fine hierarchical control structure both for global constraint propagation and for efficiency. By the use of disparity estimates from coarser levels of the hierarchy, one of the images is geometrically warped to improve the performance of the cross-correlation-based matching operator. A newly developed surface interpolation algorithm is used to fill holes wherever the matching operator fails. Experimental results for the Phoenix Mountain Park data set are presented and compared with those obtained by ETL.

1 Introduction

The primary objective of this research was to explore new approaches to automated stereo compilation for producing digital terrain models from stereo pairs of aerial images. This paper presents an overview of the hierarchical warp stereo (HWS) approach , and shows experimental results when it is applied to the ETL Phoenix Mountain Park data set.

The stereo images are assumed to be typical aerial-mapping pairs, such as those used by USGS and DMA. Such pairs of images are different perspective views of a 3-D surface acquired at approximately the same time and illumination angles. Normally these views are taken with the camera looking straight downward. The major effect of non verticality is to increase the incidence of occlusion, which increases the difficulty of point correspondence.

We shall call one of these images the "reference image," and the other the "target image." We will be searching in the target image for the point that best matches a specified point in the reference image.

It is also assumed that the epipolar model for the stereo pair is known, which means that for any given point in one image we can determine a line segment in the other image that must contain the point, unless it is occluded from view by other points on the 3-D surface. This is certainly a reasonable assumption, since an approximation to the epipolar model can be derived from a relatively small number of point correspondences if the parameters of the imaging platform are not known a priori.

The primary goal is to automatically determine correspondences between points in the two images, subject to the following criteria:

- Minimize the rms difference between the disparity measurements and "ground truth." Without ground truth, we cannot measure this.
- Maximize the sensitivity of the disparity measurements to small-scale terrain features, while minimizing the effects of noise.
- Minimize the frequency of false matches.
- Minimize the frequency of match failures.

These criteria are mutually exclusive. Under ideal conditions, increasing the size of the match operator decreases the effects of noise on the disparity measurement, but it also diminishes sensitivity to small terrain features. Similarly, tightening the match acceptance criteria reduces the frequency of false matches, but results in more frequent match failures.

One of the goals of this system is to minimize the number of parameters that must be adjusted individually for each stereo pair to get optimum performance.

2 Approach

This section briefly explains the HWS approach, which consists of three major components:

- Coarse-to-fine hierarchical control structure for global constraint propagation as well as for efficiency.
- Disparity surface interpolation to fill holes wherever the matching operator fails.
- Geometric warping of the target image by using disparity estimates from coarser levels of the hierarchy to improve the performance of the cross-correlation-based matching operator.

2.1 The Use of Hierarchy and Surface Interpolation to Propagate Global Constraints

The goal of stereo correspondence is to find the point in the target image that corresponds to the same 3-D surface point as a given point in the reference image. It is often impossible to select the correct match point with only the image information that is local to the given point in the reference image in combination with the image nformation along the epipolar line segment in the target image. When the 3-D surface contains a replicated pattern, there is the likelihood of match point ambiguity. Let us consider, for example, a stereo pair that contains a parking lot

This research was supported by the Defense Advanced Research Projects Agency under Contract No. MDA 903-83-C-0027.

with repetitive markings delimiting the parking spaces. Around the edges of the lot there are image points that can be matched unambiguously. Within the parking lot, ambiguity is likely, depending on the orientation of the repetititive patterns with the epipolar line. A successful stereo correspondence system must be able to use global match information to resolve local match-point ambiguity.

HWS approaches this problem in two ways. First, global constraints on matches are propagated by the coarse-to-fine progression of the matching process. Disparities computed at lower resolution are employed to constrain the search in the target image to a small region of the epipolar line, which also greatly reduces the probability of selecting the wrong point when ambiguity is present. Second, whenever the match process fails to find a suitable match or detects a possible match ambiguity, a disparity estimate is inserted that is based on a surface interpolation algorithm, which uses information from a neighborhood around the disparity "hole," with the size of the neighborhood depending on the number of neighboring "holes."

2.2 The Use of Image Warping to Improve Correlation Operator Performance

One of the greatest problems in the use of area correlation for match point determination is the distortion that occurs because of disparity changes within the correlation window. Since area-based correlation matches areas, rather than individual points, the disparity it calculates is influenced by the disparities of all of the points in the window, not just the point at the center. When there are high disparity gradients or disparity discontinuities, the correlation calculated for the correct disparity can actually be so poor that some other disparity will have a higher correlation score.

The effect of correlation window distortion can be greatly mitigated in a hierarchical system by using the disparity estimates from the previous level of matching to warp the target image geometrically at its current resolution level into closer correspondence with the reference image.

2.3 Related Work

Norvelle [1] implemented a semi automatic stereo compilation system at the U.S. Army Engineer Topographic Laboratories (ETL) that operates in a single pass through the images. It uses disparity surface extrapolation both to predict the region of the epipolar segment for matching and to estimate tbe local surface orientation so as to warp the correlation window. He found that these techniques improved the performance of the system significantly, but that considerable manual intervention was needed when the surface extrapolator made bad predictions, or when the image contained areas with no information for matching, with ambiguities, or with occlusions.

3 Sequence Of Operations In Hierarchical Warp Stereo

Figure 1 illustrates the hierarchical control structure of the system.

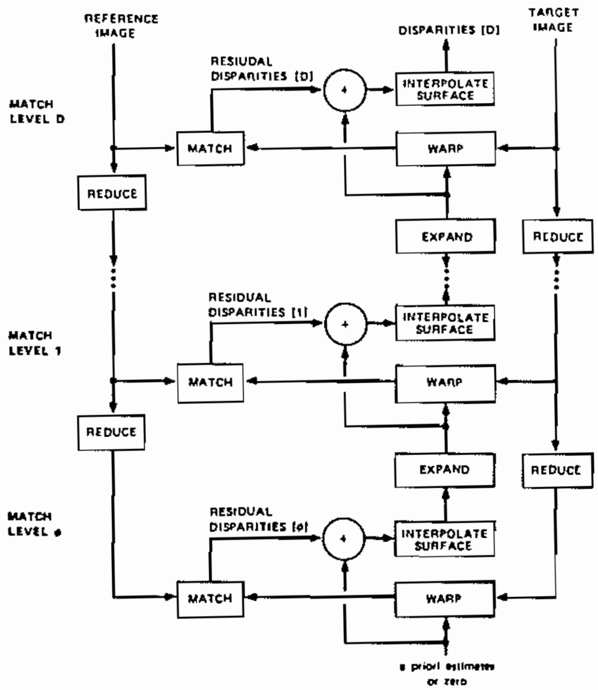


FIGURE 1

Block Diagram of Hierarchical Control Structure

1. Initialize:
   - Start with a stereo pair of images (assumed to be of the same dimensions).
   - Call one of these images the "reference image," the other the "target image."
   - Construct Gaussian pyramids (Burt [2]) $reference_i$ and $target_i$ for each image. The images at level i in these pyramids correspond to reductions of the original images by a factor of 2^i.
   - Set $disp_{-1}$ to either the a priori disparity estimates or all zeros.
   - Start the iteration at level $i = 0$.
   - Choose the pyramid depth D so that:

   $$D = ceiling(log2(uncertainty)) - 1.$$

   where $uncertainty$ is an estimate of the maximum difference between $disp_{-1}$ and the "true" disparities. Tbis guarantees the "true" disparities will be within the range (-2 : +2) at level 0 of the matching.

2. Warp: Use the disparity estimates $2 * disp_{i-1}$ to warp $target_{D-i}$ geometrically into approximate alignment with $reference_{D-i}$. Note that the factor of two is equal to the ratio of image scales between level i and level $i-1$ of the hierarchy.

3. Match: Using the matching operator, compute the residual disparities $\Delta disp_i$ between the warped target and the reference images at level i.

4. Refine: Compute the refined disparity estimates:

$$disp_i = 2 * disp_{i-1} + \Delta disp_i.$$

5. Fill: Use the surface interpolation algorithm to fill in disparities estimates at positions where matching operator fails because of no image contrast, ambiguity, etc.

6. Increase resolution: If $i = D$, quit; otherwise let $i = i + 1$ and go to Step 2.

4 Disparity Estimation

Disparity estimation consists of three parts:

- Computing match operator scores for disparities along an epipolar segment.
- Accepting or rejecting the collection of scores according to a model for the shape of the correlation peak.
- Estimating the subpixel disparities at acceptable peaks.

4.1 Match Score Operator

The HWS approach presented here can be implemented with a variety of match operators. All results reported here were obtained with an operator that closely approximates Gaussian-weighted normalized cross correlation. The values of the Gaussian weights decrease with Euclidean distance from the center of a square correlation window. In the examples shown here, the window dimension is 13×13 pixels with a standard deviation of approximately 2 pixels in the Gaussian weights. Preliminary results indicate that the Gaussian-weighted correlation operator is better than uniformly weighted correlation operators at locating changes in disparity while maintaining a given level of disparity precision.

4.2 Evaluation of Correlation Surface Shape

The match operator reports a failure if any of the following conditions exist:

- Disparity out of range: The maximum match score is found at either extreme of the epi-polar segment.
- Multiple peaks: The best and next best match scores is found at disparities that differ by more than one pixel.

There are other models for the expected shape of the correlation surface that can be based on the autocorrelation surface shape of the windows in the reference and target images. Further investigation is needed to evaluate the utility of such models for both surface shape evaluation and disparity estimation.

4.3 Subpixel Disparity Estimation

The subpixel location of the correlation surface peak is estimated by parabolic interpolation of both the x and y directions of disparity. For each direction, three adjacent match scores – s_{i-1}, s_i, and s_{i+1}, where s_i is the maximum score – are used to compute the peak as follows:

$$.5 * \frac{s_{i+1} - s_{i-1}}{2 * s_i - s_{i+1} - s_{i-1}}$$

More complicated approaches to peak estimation, such as two-dimensional least-squares fitting of the correlation surface, might yield better estimates, but at a higher computational cost.

5 Surface Interpolation Algorithm

The goal of the surface interpolation algorithm is to estimate values for the disparity surface at points where the match operator reported failure; such points will be called "holes." The approach to filling a hole at location x, y is to model the surface by employing the disparity measurements over the set of nonholes $\overline{H}$ in the $n \times n$ pixel neighborhood centered at x, y. The set H contains the indices of all holes in the neighborhood.

This surface interpolation algorithm is based on the solution to the hyperbolic multiquadric equations described in Smith [3]. The surface is known at the set of points x_i, y_i, z_i where $i \in \overline{H}$, and can be estimated at other points $h \in H$ by the formula

$$z(x_h, y_h) = \sum_{i \in \overline{H}} c_i * g(x_h - x_i, y_h - y_i),$$

where g is the basis function for the surface respresentation, and coefficients c_i are the solutions to the set of linear equations:

$$z(x_j, y_j) = \sum_{i \in \overline{H}} c_i * g(x_i - x_j, y_i - y_j) \quad for\ all\ j \in \overline{H}$$

Clearly, this irregular grid solution could be used to compute the surface values at the holes in the disparity data, but this involves solving for the coefficients c_j for each different configuration of holes and nonholes in the $n \times n$ neighborhoods of the disparity surface.

An alternative approach, which is used here, is to convert the quasi-regular grid problem into a regular grid problem in which each c_i at a hole is forced to be zero, and the corresponding z_i remains as an unknown. This results in the same solution that would have been obtained from the irregular grid formulation and produces the following system of linear equations:

$$\sum_{i \in H} A^{-1}_{h,i} * z_i = - \sum_{j \in \overline{H}} A^{-1}_{h,j} * z_j \quad for\ all\ h \in H, \qquad (1)$$

where A^{-1} is the inverse of the matrix $A_{i,j} = g(x_i - x_j, y_i - y_j)$ for $i, j \in H \cup \overline{H}$. This system of equations must be solved for each z_i for $i \in H$. Thus, we have reduced the size of the linear system of equations that must be solved from the number of elements in $\overline{H}$ to the number of elements in H. Of course, the matrix A must be computed and inverted once.

Areas on the disparity surface that contain large clusters of holes cause problems. The previous surface interpolation algorithm degenerates to a surface extrapolation algorithm when the nonholes in the neighborhood are not more or less isotropically distributed over the entire neighborhood. The problem can be overcome by increasing the size of the neighborhood until some spatial-distribution criterion is met, but this would require solving extremely large linear systems.

Large holes are filled by means of the following hierarchical approach:

Procedure Surface-Interpolate($surface_i$)

1. If $surface_i$ contains large holes then

(a) Compute filled-surface$_{i+1}$ = *expand*(surface-interpolate(*reduce*($surface_i$))), where *reduce* computes a Gaussian convolution reduction by a factor of two, surface-interpolate is a recursion call to this interpolation algorithm, and *expand* computes expansion by a factor of two, using bilinear interpolation.

(b) For each hole in $surface_i$ that is completely surrounded by other holes, fill the hole with the value from the filled-surface$_{i+1}$.

2. For each hole in $surface_i$ fill the hole by solving the system of linear equations (1) for the $n \times n$ pixel neighborbood centered at the hole (n = 7 in the examples).

3. Return the filled $surface_i$.

6 Examples

This section describes the experimental results achieved when the HWS technique was applied to areas of the ETL Phoenix Mountain Park data set, and compares these results to those obtained from the semiautomatic system developed by Norvelle [1].

The following components of the Phoenix Mountain Park data set were used:

- Left image: 2048 x 2048 pixels, 8 bits per pixel
- Right image: 2048 x 2048 pixels, 8 bits per pixel
- x-correspondence array: 400 x 400 points , floating point.

The left and right images had been scanned such that the epipolar lines were almostly exactly horizontal. The ETL x-correspondence array was converted to an x-disparity image to enable comparison between ETL and HWS results.

Results are shown for two different areas of the Phoenix data set. All disparity measurements are indicated in terms of pixel distances in the 2048×2048 Phoenix stereo pair, rather than the resolution of the selected windows.

- Area A is defined by two approximately aligned 150×150-pixel windows of the Phoenix pairs which were reduced by a factor of four (the windows thus corresponding to the 600×600-pixel windows of the originals). The measured disparities for area A range from -40 to +16 pixels.
- Area B is defined by two approximately aligned 125×125-pixel windows of the Phoenix pairs which were reduced by a factor of two (the windows thus corresponding to the 250×250-pixel windows of the originals). The measured disparities for area B range from -40 to -34 pixels.

Figures 2 and 3 show the inputs and outputs of three levels of the hierarchy for areas A and B, respectively. Columns 1 and 2 are the reference and target images at each level. Column 3 is a binary image that indicates the positions of match failures. Column 4 shows the resulting disparity image of each level after the match failures have been replaced by surface-interpolated disparity values.

Figures 4 and 5 contain a comparison of the HWS results with those obtained at ETL by Norvelle for areas A and B respectively. The bottom-left images of figures 4 and 5 show the pixel-by-pixel differences, after contrast enhancement, between the HWS and ETL disparities. The graphs to the right of these difference images depict the histograms of these differences.

The mean and standard-deviation values shown with the histograms provide a useful quantative comparision between the HWS and ETL results. They show that the average disparity differences were .082 and .025 pixels, and that the standard deviations of the disparity differences were .67 and .34 pixels for the A and B window pairs, respectively, in terms of pixel distances in the 2048×2048 Phoenix pairs. These standard deviations become .17 and .17 pixels when expressed relative to the scales of A and B windows, respectively.

Similar results have been achieved for other examples that include both higher resolution and larger windows.

7 Problems

HWS is still very experimental. Some of the parameters that affect the system, such as the range of disparities to compute at each level of hierarchy and the size of the correlation operator, are still specified manually.

There are problems in estimating the range of disparities to be computed at each level of the hierarchy. If the estimate is too low, there will be frequent out-of-range match failures. If, on the other hand, the estimate is too high, computation time will increase and there will be more potential for match point ambiguity.

HWS has difficulty dealing with steep terrain features that have small image projections, but large disparities. At low resolutions in the matching hierarchy, the disparities of the terrain surrounding the feature dominate those of the feature itself, resulting in a disparity estimate that is usually intermediate between that of the feature and that of the surround. At higher resolutions in matching, the disparity of the steep feature may be outside the permissible disparity range.

HWS has even greater problems with oblique stereo pairs containing many occlusions. At low matching resolution, the disparities of foreground and background in the same neighborhoods cannot be distinguished. As the matching resolution increases, foreground and background features are discernible as separate objects, but their disparities are out of range for the matcher.

Most of the difficulties caused by sudden changes in disparity might be solved by preceding the disparity surface interpolation step with an algorithm that attempts to match still unmatched regions in the reference image with regions in the target image that likewise have not yet been matched. We thus attempt to match holes with holes.

8 Conclusions

HWS produces very good results for vertical stereo pairs of rolling terrain. With the incluson of a hole-to-hole matching step, HWS should be capable of comparable performance for terrain characterized by steep slopes and frequent occlusions.

Bibliography

[1] Norvelle, F. Raye, *Interactive Digital Correlation Techniques For Automatic Compilation of Elevation Data*, U.S. Army Engineer Topographic Laboratories, Fort Belvoir, VA 22060, Report Number ETL-0272, Oct. 1981.

[2] Burt, Peter J., *Fast Filter Transforms for Image Processing*, CGIP, 16, 20-51. 1981.

[3] Smith, Grahame .B., *A Fast Surface Interpolation Technique*, Technical Note 333, Artificial Intelligence Center, SRI International, Menlo Park, California, August 1984. (Also in these proceedings).

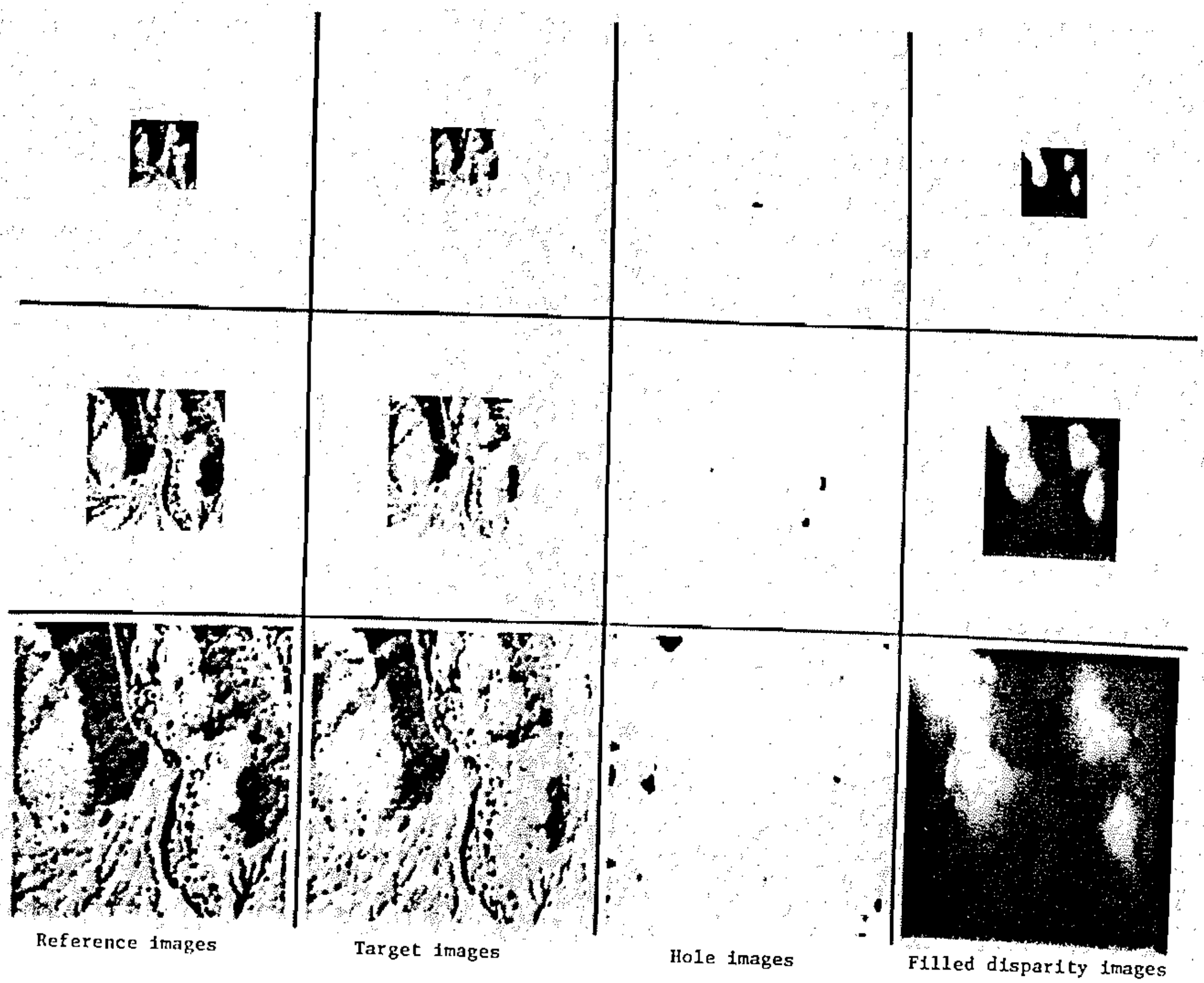


FIGURE 2 HWS results for area A

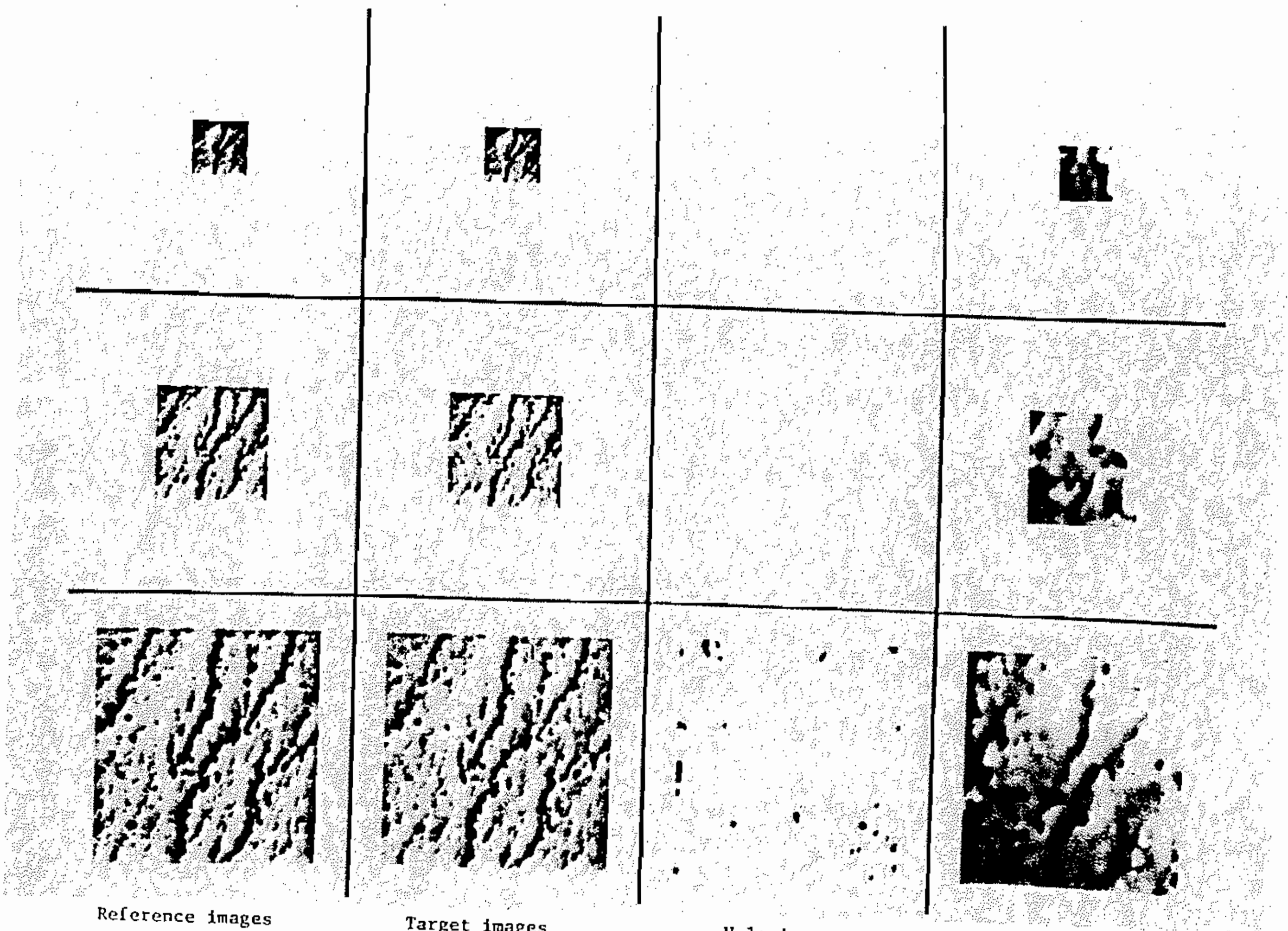
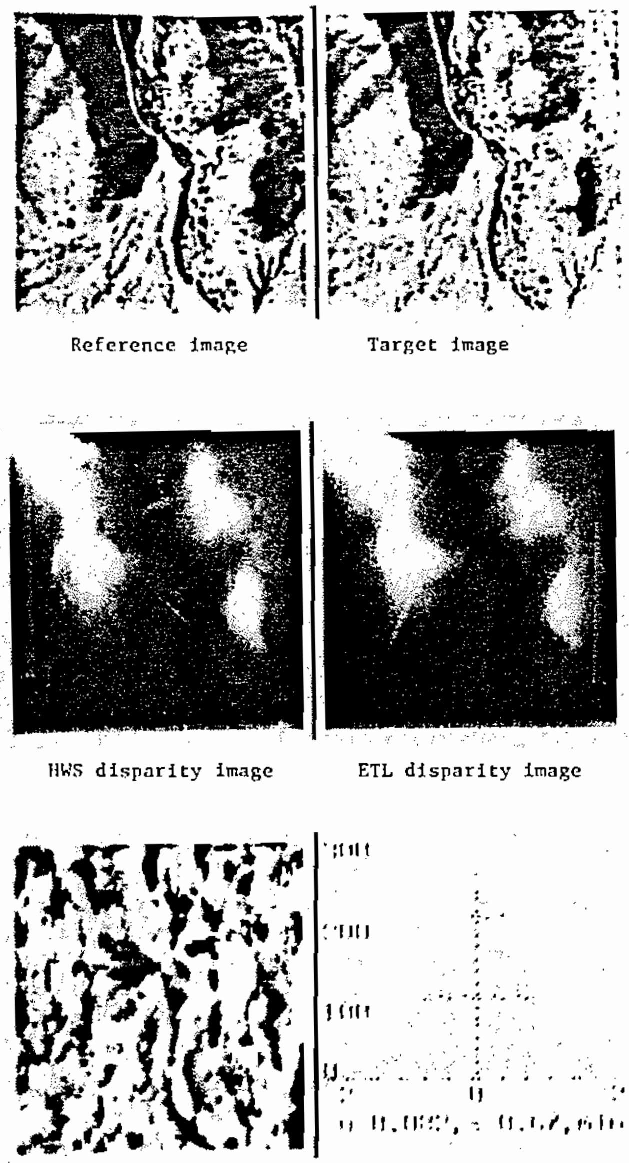


FIGURE 3 HWS results for area B

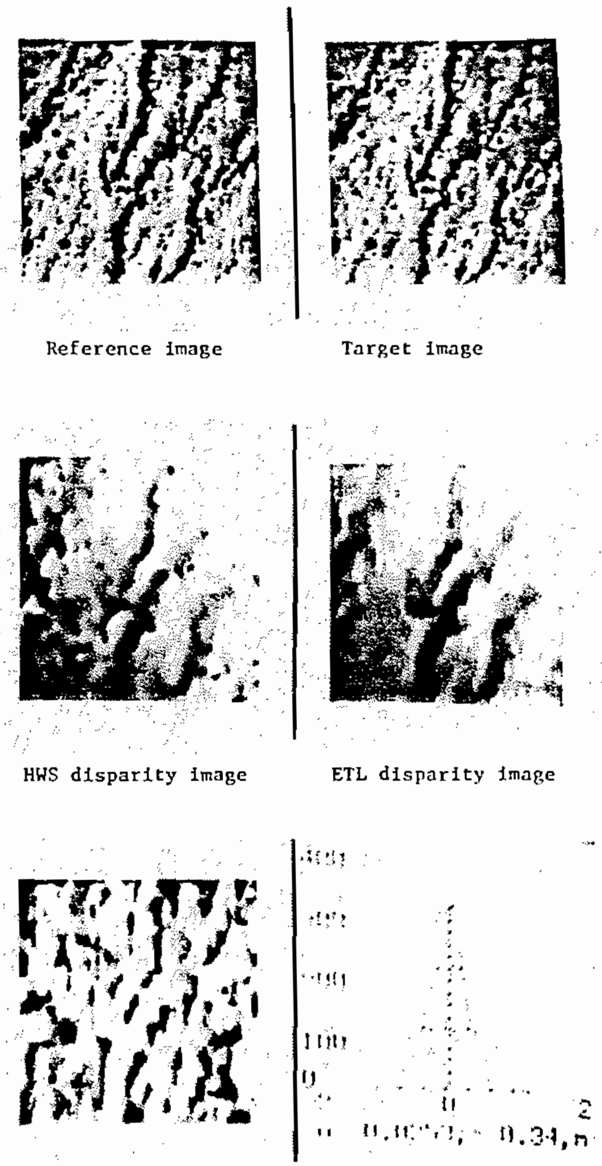


Reference image — Target image

HWS disparity image — ETL disparity image

HWS - ETL difference — HWS - ETL histogram

FIGURE 4 HWS vs. ETL results for area A



Reference image — Target image

HWS disparity image — ETL disparity image

HWS - ETL difference — HWS - ETL histogram

FIGURE 5 HWS vs. ETL results for area B