EXPERIENCES IN AUTOMATED GENERATION OF DTMs FROM AERIAL PHOTOS

Mr. Bon A. Dewitt, Graduate Student
Dr. Paul R. Wolf, Professor
University of Wisconsin, Madison, Wisconsin
USA
Commission 4

ABSTRACT

An automated analytical photogrammetric system for generating DTM data from measured image densities of aerial photos is presented. Phases of the system which are described include (a) image coordinate refinement based on automatic location of calibration marks, (b) digital image correlation, (c) relative orientation, (d) absolute orientation from automatically located ground control panels, and (e) computation of DTM points. Results achieved on a recent test project are discussed.

INTRODUCTION

Over the past twenty years, research has been conducted on methods of extracting three dimensional digital terrain models (DTM's) from pairs of overlapping, digitized photos. Recently, with improvements in computer hardware such as dedicated parallel image processors, the possibility of efficiently realizing this goal has been enhanced. Algorithms which operate on the massive amounts of data that must be manipulated in this problem are becoming feasible through use of this new hardware. Consequently the systems which were conceptualized in the past, are now in the process of becoming practical.

At the University of Wisconsin, a prototype image processing software system has been under development to handle the complexities which arise in the compilation of DTM's from digital photos. This paper is presented as an update on the status of the current implementation of this software. A discussion of the difficulties encountered in the development of the system is presented, followed by a general description of methods of implementation. Attributes of the system discussed are the basic data structure used, a method of photo coordinate refinement, relative and absolute orientation of stereopairs, and extraction of the three dimensional digital terrain models. Results of two tests indicating accuracies which can be obtained for coordinates of points in both photo and ground space are presented as well. Results of the first test indicate that coordinates of discrete images in digital photos can be extracted and refined to accuracies of approximately one-fourth the pixel size. In the second test an aerotriangulation adjustment was performed, and results obtained indicated that the accuracy of the digital method approaches that of conventional monocomparator measured photo coordinates.

PROBLEMS ASSOCIATED WITH SYSTEM DESIGN

There are many difficulties inherent in the task of automated mapping using digital photography. The major difficulty which becomes immediately apparent is the massive data storage requirements of digitized photographs. On a typical 9x9 inch photo, scanned at 50 micrometer pixel size, there are roughly 20 million image densities which must be stored, each requiring 6 to 8 bits to encode. When considering stereopairs, this number is doubled. Storing image arrays of this size in fast access core memory is impossible for all but the very largest mainframe computers. Thus secondary memory units, such as disk or tape files must be used to store the image arrays. The significance of this in terms of system efficiency, is that much of the time required to process photos will be spent in I/O. In order to reduce this processing time, images should be scanned at as large a pixel size as possible, while still allowing sufficient accuracy to be obtained in the end product.

Another problem which is basic to the system, is the need to recognize and correlate conjugate imagery between photos in a stereopair so that their coordinates can be obtained and used in analytic equations to determine three dimensional ground locations. A well known statistic used in image matching, the normalized cross correlation coefficient, while being a very robust measure of similarity, is extremely inefficient when applied in a "brute force" approach. Therefore some sort of artificial intelligence must be imparted into the program to minimize the number of computations of this coefficient. Search regions must be narrowed down to a reasonable size in order for cross correlation to work efficiently. In addition, other measures of image likeness, in conjunction with certain types of array transformations, can be substituted for cross correlation when narrowing down these search windows.

Under carefully controlled laboratory conditions it is possible to orient the exposures such that parallactic displacements of points between photos will occur along lines parallel with the scan lines in the digitized photos. However, with aerial photography this is possible only to an approximation. This results in conjugate image arrays being mis-registered not only in terms of x and y translation, but also in rotation. Rotational mis-registrations of this type can result in difficulties for the image matching algorithm, due to apparent differences in density sub-arrays for conjugate points. Re-registering of the density arrays of the photos such that image displacements occur along scan lines will facilitate subsequent image matching. In addition, it will allow use of one-dimensional pattern matching techniques, since it is known that all image matches will fall along the resampled lines.

Another difficulty, which is magnified when using brute force approaches, is the problem of false matches being obtained from the image correlator. In large areas of homogeneous high frequency texture, it is possible for matches to be incorrectly made for a pair of points due to their similarity. For example, in overlapping photography of an apple orchard, two trees A and B may be falsly matched because their neighborhood arrays appear identical. This stems from the fact that trees in an orchard have a highly consistent shape. The end result of a false match such as this will be a sharp discontinuity in the subsequent DTM. Avoidance of false image matches is an essential attribute of the DTM generation algorithm.

To achieve a high order of map accuracy, a dense pattern of DTM points must be matched, and their 3-dimensional coordinates computed, to enable accurate contours to be plotted. This dictates that a tremendous number of image comparisons must be made to derive the model. Therefore it is important that the matching algorithm be as efficient as possible in order to keep execution time at a practical level.

One last difficulty which arises is the need to automatically recognize standard image types such as fiducial marks and ground control panels. Fiducial marks must be located in the scanned images to allow for transformation of row/column scanner coordinates into the photo x,y system, as well as provide for coordinate refinement control (Wolf and Dewitt, 1984). Ground control panels must also be located to enable the derived DTM to be transformed into meaningful coordinates, such as state plane coordinates and elevations. Automatic recognition of these image types requires that a standard template be generated, which can then be matched to the actual mark as it appears in the digitized photo.

In the development of an automated mapping system it is necessary to address these problems so that efficient algorithms and data structures will be implemented.

DATA STRUCTURE

In selecting an appropriate data structure, an attempt was made to choose one which would be compatible with the process a human employs when matching conjugate images (as in a stereoplotter). When humans go through the process of relative orientation of a stereoplotter, they do not immediately attempt to match points in a high resolution mode. Instead, large features such as farm fields, lakes, major highways, etc. are roughly matched in order to bring the model into an approximate orientation. Then refinements are made in which more precisely defined points are used to clear y parallax. It is this low to high resolution procedure which should be modelled in the computer program. Therefore a structure which allows matches to be performed in a multi-resolutional mode is desired.

One data structure which has been proposed that has this high to low resolution feature is the image pyramid. (Burt 1980) In this structure, multiple layers of the image at successively lower resolutions are generated and "stacked" on top of the original image array. In each lower resolution layer, a square region with a width of two times the original pixel size is replaced by one representative value, resulting in an image array which is one half the size in height and width. The stack which results from this has smaller and smaller layers as it goes up, hence the name, image pyramid. A schematic representation of this structure is shown in figure 1. This figure shows a sequence of three layers with the lowest resolution layer at the top. The figure is intended to be a conceptual aid only, and the appearance that four pixels are registered exactly with one pixel in the higher level, is not the actual structure. Instead, pixel centers in a higher level line up directly above every second pixel center in the level below. Searching for images in the upper levels of the pyramid corresponds to the low resolution matching phase performed by a human. Going down the pyramid for matches allows refinement of image correlations at a higher resolution.

In building the pyramid, a window operator is used to suppress high frequency information from being passed to the upper levels. The Gaussian operator is used for this task. It extracts a weighted density average of a small neighborhood (eg. 5x5 array) at evenly spaced points in the image array. Extracting averages at such a registration results in a half-resolution copy of the image which will be "stacked" on top of the original. Recursive application of this procedure to every new level generates the image pyramid. Special consideration must be provided for in the algorithm to handle the cases which occur when the window overlaps the boundary of the digital image.

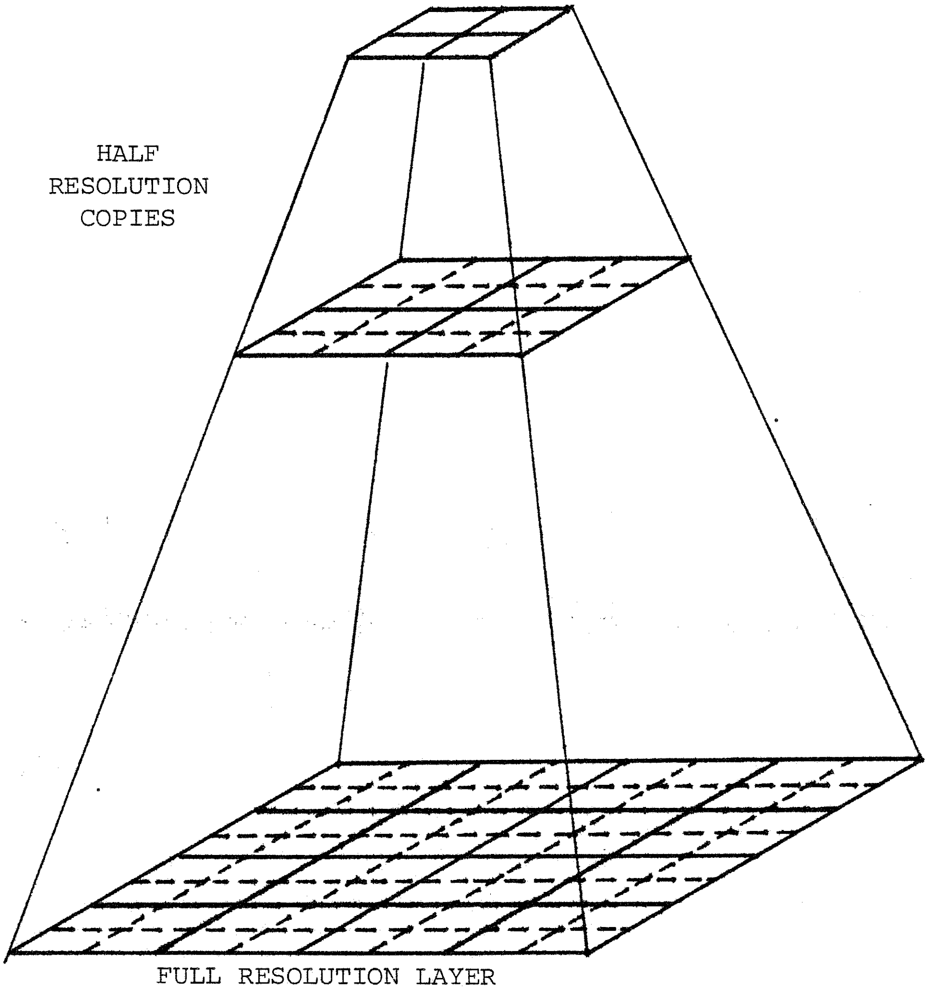

Figure 1.
Image Pyramid

As previously mentioned, this pyramid structure allows multi-resolution searching to be performed in correlating conjugate points. The procedure for doing this is as follows: Selecting a candidate target point from the left photo, and starting at the top level in that photo pyramid, follow a directed path through the levels, to arrive at the desired, full resolution location for this potential match point. In the process the subarrays in the neighborhoods of the pixels encountered in the path are saved. Next, starting at the top of the right photo pyramid, follow a downward path dictated by correlations with the stored neighborhood arrays from the left photo. Finding a matching neighborhood at a particular level, gives the location of the point to within one pixel in the next lower level. Therefore only small sub-arrays need be searched at any given level to find the match at that resolution.

The number of correlations required for this "top-down" search is on the order of the logarithm (base two) of the original image array width. This is analagous to a binary search procedure in a sorted array to find a particular element. This results in acceleration of an order of magnitude when compared to brute force searching within the original array. In addition to the increased efficiency, there is another desirable by-product of multi-resolution searching. False point matches are reduced by virtue of the accurate prediction of the position of the point at the next lower level. Returning to the example of the trees in the orchard, the possible false match of tree A to tree B is bypassed at the upper pyramid levels. This is achieved by matching low resolution information such as the orchard boundaries, to accurately predict the location of the tree at a lower pyramid (higher resolution) level.

SYSTEM OVERVIEW

Having described the problems associated with system design, and the basic data structure to be used, the general system algorithm is presented. First, however a brief description of the necessary input information is given.

The most obvious input required by the program is the digital representation of the photos in the stereopair. This may be obtained by scanning film diapositives in a microdensitometer to generate digital arrays. Alternatively, data which is recorded in a digital form by the sensor (eg. linear array camera) could be used. Other input data includes: (1) Approximate flight parameters (rough values for camera exterior orientation elements), (2) Calibrated fiducial and/or reseau grid coordinates, and (3) Ground coordinates of paneled control points.

The first step in the processing is to build the Gaussian image pyramid. In this process, the computer tape containing the digital photo is downloaded into a disk file (one for each photo). Then, file records are loaded into a buffer, the size of which depends on the size of the window for the Gaussian operator (5x5 window for this system). As the operator is passed through the values in the buffer, the averaged low resolution values are extracted and written to the file in the position corresponding to the pyramid level being generated. When a full row has been generated, the next two rows from the input level are loaded into the buffer, overwriting the two rows that fall below the current window range. This operation progresses until the full image is converted. Then this procedure is applied recursively to the resulting layers, up to the top level. Since each higher level has only one-fourth the number of pixels of the previous layer, the total storage required is $1 + 1/4 + 1/16 + 1/64 + \ldots = 1.333$ times the size of the original image.

Next, a search for the fiducial and/or reseau marks is initiated. Location of the first two marks is most time consuming because only general locations of the marks are known. After two marks have been found, a two-dimensional conformal transformation is computed which enables accurate prediction of the locations of the remaining marks. Due to the accuracy of the predicted locations, these remaing marks can be searched for in the full resolution image by computing correlations at only a few window positions. Once all available marks have been located, a polynomial transformation is applied which corrects for the systematic errors introduced by the scanner and other sources (film shrinkage, etc.). These transformation parameters are subsequently used to convert scanner coordinates of matched images into photo coordinates.

The next phase is analytical relative orientation by an iterative process. As mentioned in the section on system input, approximate flight parameters are known, giving an initial relative orientation. In the iterative approach, conjugate image points are found one at a time and their coordinates used to refine relative orientation. Each refinement allows more accurate prediction of the location of conjugate images for subsequent matching. A large number of redundant points is desired to enable rejection of possible falsely matched points by casting out those with large residuals, thus giving a stronger solution while retaining a satisfactory number of degrees of freedom. Once a strong relative orientation has been obtained, future searches for conjugate images are then constrained along epipolar lines.

Having achieved relative orientation, paneled ground control points are located and absolute orientation computed. This process is also iterative. As each control panel is located, absolute orientation is recomputed, adding strength to the solution with every additional point. When a sufficient number of points have been located to enable elimination of the approximate flight parameter constraints, the remaing control locations are accurately predicted by direct solution of the collinearity equations. The process of finding the initial control panels in the scanned image requires large search areas to be probed in the full resolution image. Correlation in this mode gives panel coordinates to the nearest whole pixel. This is followed by a refinement process involving iterative resampling and correlating of the standard image template with the photo image, until the maximum correlation peak is obtained. The resampling is done by making fractional pixel translations (along with rotation and scaling) until the correlation coefficient stabilizes at its peak value. This is essentially the same process used for finding fiducials.

Once absolute orientation has been computed, the next step is correlation of DTM points in a grid pattern covering the stereomodel. Although this could be accomplished with the same two-dimensional correlation technique used in the relative orientation phase, efficiency can be improved by constraining the search along epipolar lines. To accomplish this, the image densities are resampled along epipolar lines in each photo, thus enabling the correlations to be done in a one-dimensional mode. Since it is known that conjugate images lie along these lines, the resampled signals can be matched in much the same way as done in analog correlation devices such as the Gestalt Photomapper. Performing the matching in this one-dimensional mode reduces the time complexity of correlation from order (M^2xN^2) to order (MxN), where M is the width of the search window, and N is the width of the target. There is a negative attribute in this approach however. Since the correlations are computed using fewer pixel values (N pixels rather than N^2) the chances for a false match are increased by virtue of the reduced number of degrees of freedom. The trade-off between two- and one-dimensional correlation depends on the amount of rotational mis-registration (relative kappa between photos). This rotation appears to affect two- but not one-dimensional correlation. More investigation concerning this trade-off should be performed.

Because of the likelihood of generating false matches in the correlation process, a data smoothing operation follows the computation of DTM coordinates. By applying thresholds on the allowable slopes, and changes of slope in the model, falsely matched points (which normally stand out as sharp peaks in the DTM) can be effectively removed (Hannah 1981). After this filtering operation is performed, the DTM grid can be used as input to a contour mapping routine. The computer drawn contour maps are the final products of this automated mapping system. Adding planimetric detail to the map by use of edge detectors could also be incorporated into the system, but this is beyond the scope of the current investigation.

RESULTS

At present, most of this research is aimed at improving accuracies obtained in the absolute orientation phase. Tests were performed to evaluate the accuracy to which coordinates of discrete points, in both image and ground space, can be derived from digital photographs. Results of two tests are presented. The first test consisted of refining scanner-derived coordinates for five photos. The results of the first test were applied in the second, where the extracted coordinates of ground control panels were used in an aerotriangulation adjustment of a test model. Black and white photography, taken with a Zeiss RMK 15/23 camera from an altitude of 3000 feet and scanned at a fifty micrometer pixel size was used for this test. A number of paneled control points were imaged in the photos, which could be used to control the pair and provide the necessary checks.

The first step in the usual approach to photo coordinate refinement consists of correcting for film distortion errors by transforming arbitrary comparator coordinates to the fiducial axis system, often done with an affine transformation. When performing this refinement with scanned images however, because of geometrical distortions introduced in the digitizing process it is desirable to have additional calibration marks available, such as a reseau grid. For the purposes of this research, a reseau grid was simulated by drilling a pattern of 32 holes into the photographs prior to scanning. The coordinates of the holes were precisely measured by monocomparator to provide the distortion control. The fiducials and drill holes were located in the scanned image and their coordinates extracted. This is analagous to the usual monocomparator measurement phase, except that the coordinates were in terms of scanner rows and columns. Polynomial equations were used to model the distortions introduced by the scanner. For this phase, monocomparator measured drill hole coordinates were used as control, and row/column coordinates were transformed into the photo x,y system. The results of the polynomial adjustments, performed on a set of five photographs, indicated that the scanner derived coordinates could be corrected to an RMS positional accuracy of approximately one-fourth the pixel size (Wolf and Dewitt 1984).

To test the accuracy with which ground coordinates could be derived from digital imagery, an absolute orientation was performed. It was based on photo coordinates of panel locations extracted from the scanned image through correlation with standard templates. Prior to the adjustment, the scanner coordinates were refined by polynomial modelling as described above. In the test, four panels were used as ground control for a bundle adjustment, and four additional panels served as checkpoints and were allowed to adjust without ground control constraint. The adjustment gave RMS discrepancies of 0.53, 0.16 and 0.24 feet, respectively, in X, Y and Z. For comparison, an adjustment was performed using comparator measured coordinate values of the same panels. This adjustment yielded RMS checkpoint discrepancies of 0.35, 0.19 and 0.16 feet, respectively in X, Y and Z.

CONCLUSIONS

The foregoing sumarizes research to date at the University of Wisconsin - Madison in automated generation of DTM's from digitized photos. Results obtained from experimental photography processed by the described software are encouraging. Accuracies of the extracted photographic coordinates to within one-fourth of the pixel size can rival those obtained conventionally by monocomparator, given a sufficient reduction of pixel size. With such a refinement of the photo coordinates, the positional accuracy of the three dimensional model coordinates can be on the same order as that obtained by conventional methods.

Contemplated areas of further research include: (1) Investigation of system performance with photography of varying scales and pixel sizes, (2) Use of multi-spectral photography, (3) Extraction of planimetric detail through use of edge detectors, and (4) Ultimate implementation on parallel processing hardware to enhance speed.

REFERENCES

1) Burt, P.J., "Fast Hierarchical Correlations with Gaussian-Like Kernels," Proceedings of the Seventh International Joint Conference on Artificial Intelligence, 1980, IEEE.

2) Burt, P.J., et.al., "Multi-Resolutional Flow-Through Motion Analysis," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1983, IEEE.

3) Glazer, F., et.al., "Scene Matching by Hierarchical Correlation," ibid, 1983, IEEE.

4) Hannah, M.J., "Error Detection and Correlation in Digital Terrain Models," Photogrammetric Engineering and Remote Sensing, v 47, n 1, Jan., 1981.

5) Lucas, B.D. and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," Proceedings of the Fifth International Conference of Pattern Recognition, 1980, IEEE.

6) Rosenfeld, A., "Quadtrees and Pyramids for Pattern Recognition and Image Processing," Proceedings of the Seventh International Joint Conference on Artificial Intelligence, 1980, IEEE.

7) Wolf, P.R. and B.A. Dewitt, "Correction for Systematic Errors of a Drum-Type Scanning Microdensitometer," Archives of the International Society for Photogrammetry and Remote Sensing, Commission III, Rio de Janeiro, Brazil, 1984.