STRATEGIES FOR AUTOMATIC TIE-POINT DETECTION[1]

Frank Stolle, Gary Whitten, and Allan Hanson
Department of Computer Science, University of Massachusetts at Amherst

Commission III Working Group 1

KEY WORDS: Automatic Tie-Point Extraction, Matching

ABSTRACT

Recently the Computer Science and Forestry and Wildlife Management Departments at the University of Massachusetts, Amherst were awarded a grant to investigate automatic classification of tree species in forests from aerial data. Faced with the problem of registering immense numbers of small-scale aerial photographs and video data, it quickly became apparent that automation of tie-point selection and bundle adjustment would greatly improve productivity. It was decided to investigate the possibility of building such a system. The need for automatic processing also originated from another project, which concentrates on building site models of urban areas from aerial data. Both projects rely heavily on aerial views, but the type and quality of the imagery is different. While the former uses a combination of small-scale aerial photography and video imagery, the latter relies on classical large-scale aerial photography.
As different operators and strategies show varying behavior under different contexts, it was felt that the selection of interest operator and strategy was important enough to explore several possibilities. Specifically, two methodologies were explored. The first approach to finding tie-points that was investigated is based on interest points generated by an operator proposed by Förstner (Förstner, 1994), the second is adapted from a research effort on a model-based automated target recognition system by Whitten (Whitten and Rosenfeld, 1996).

1 INTRODUCTION

Research by Ackerman (Ackerman and Tsingas, 1994) and Krzystek (Krzystek, 1998), Tang (Tang and Heipke, 1994) and Schenk (Schenk and Toth, 1993) has made significant progress towards robust automated point selection and matching for use with a photogrammetric bundle adjustment procedure. Based on this research, several extraction and matching systems have been built and have had success, some in the commercial market. PHODIS-AT (Tang and Heipke, 1994), MATCH-AT (Ackerman and Tsingas, 1994), and Schenk's (Schenk and Toth, 1993) system use low-level image features (such as relative edge orientation) as tie-points along with a second stage refinement technique, usually using least-squares matching. Some systems integrate block adjustment into the process of tie point selection and matching (MATCH-AT) whereas others require a separate block adjustment package (PHODIS-AT) to eliminate error points. Because neither of these systems provided the architecture we required, an effort was launched to create our own.

2 TIE POINTS FROM LOCAL CURVATURE

The first approach to finding tie-points that we investigated is fairly traditional. It is based on interest points generated by the operator proposed by Förstner (Förstner, 1994). The operator is well-known and we will not go into great detail to describe it here. This operator evaluates local curvature in a window. It takes as parameters the window size, a weight threshold w, and a curvature threshold q. In our experiments, w is set to 0.95 and q to 0.75. These numbers were determined empirically. The operator returns a list of interest points for each image.

Since one of the immediate needs was to recover the epipolar geometry (relative orientation) to facilitate DEM generation via a stereo system, initial experiments were conducted on pairs of images. The control strategy proceeds as follows. In an iterative fashion, the interest operator is applied to each of a pair of images across a range of scales (window sizes). Points are ordered by the variance in the operator window.

The next step is to calculate the cross-correlation between windows in the first image, centered around the top 50% of the first interest point list, and windows in the second image centered around all points of the second list. This step is again carried out iteratively across a range of scales. It requires that approximate relative orientation information is provided. A correlation threshold is used, and all correspondences that fall below are rejected. In our experiment, the threshold was empirically set to 0.8 The window size for correlation is a variable parameter, updated during each iteration. Pairs of points are ordered first by the number of scales that are found to be above the correlation threshold, and within that ordering by the highest correlation coefficient.

One of the goals is to remove erroneous correspondences in the final list. The search space is generally very large, so restrictions are imposed on the search by initial estimates of the camera orientation and simple topological constraints. At present, only one topological test is employed. It is based on separating the plane into half-planes. The procedure starts with an empty set *S*. Two pairs of corresponding points (A,A`), (B,B`) are found and inserted into the set. In each of the two images, a line is generated between consecutive points in each of the two images (A-B) and (A'-B'). The ordering of points implies a line direction. A new point pair can be joined into the set only if the points in both images are on the same side of the line.

In case the initial correspondences are wrong, almost none of the remaining points will pass the test. We use a search procedure to backtrack and use a starting point lower in the list if this occurs.

[1] Sponsored by Grants from the Defense Advanced Projects Research Agency (DACA76-97-K-0005), the National Science Foundation (EIA-9726401) and the National Fish and Wildlife Foundation (98-089)

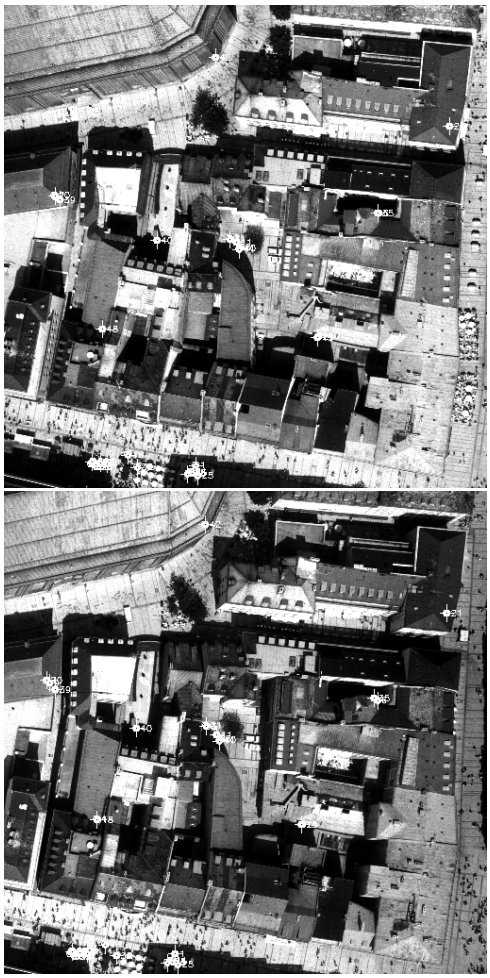

Figure 1: Munich image pair

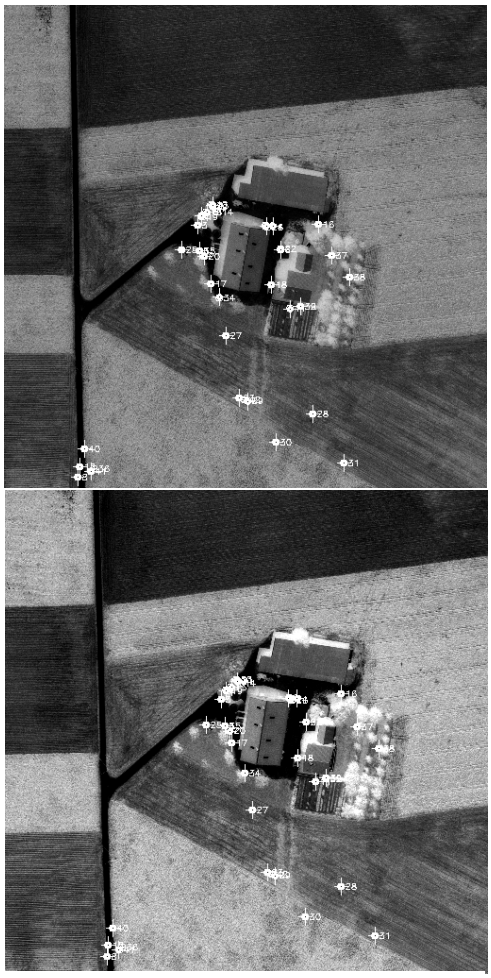

Figure 2: Echallens image pair

Figure 3 illustrates the constraint. When more than two pairs of points are in the set, the line constraint is tested on all consecutive pairs of points. The procedure works well in eliminating outlier pairs, but may also remove good pairs. This occurs because some points fall very close to a line, and due to local surface and viewing geometry could appear in one image on one side of the directed line, but on the other side in the other image. More sophisticated topological tests are the focus of ongoing research. Specifically, we are planning to employ other orientation and inclusion tests.

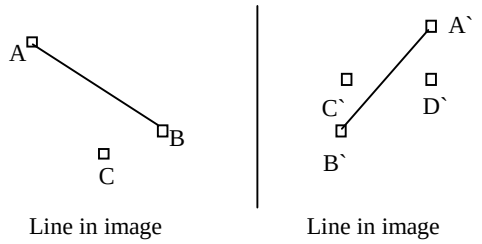

Figure 3: Topology constraint. A-A` and B-B` are already part of the set. C-C` could be included if other conditions are met, but C-D` does not pass the test

2.1 Data sets and results of first approach

For our initial experiments, we selected images from two of the ISPRS tie-point test data sets (Munich, Echallens), and a data set taken with our own sensor package (Winrock). The former were taken using a photogrammetric camera, the latter using a small-scale commercial camera. The data are presented with correspondences found, in Figures 1,2 and 4, respectively.

Parameters were set to detect a large number of possible correspondence candidates, and to significantly reduce that number using the correlation and topology tests. Parameters are given in Table 1and results are summarized in Table 2.

In each of the image pairs, sufficiently many points were detected for recovery of relative orientation. The ratio of outliers is significantly below 50%, which allows a standard robust estimator to be used for bundle adjustment. The results were obtained with standard parameters across the data sets. No specific parameter tuning was performed.

While the results are acceptable, we found some problems with this approach. The distribution of points in the images is not optimal. Points tend to be clustered, and some obvious outlier points are not removed. As an example, in Figure 2, most of the points are concentrated around the buildings.

Figure 4: Winrock image pair

At this points we turned our attention to a system already at hand, that, while designed for another purpose, promised to be useful in our problem domain. This is described in greater detail below.

3 VERTEX SPACE SYSTEM

The second system explored here utilizes model-based recognition techniques developed and implemented by Whitten (Whitten and Rosenfeld, 1996). It uses a localized interest operator based on curvature to generate points in the two images. The points are then matched using a special Hough-like 2D feature representation. The system's original problem domain of target recognition is very similar to that of tie point extraction on urban aerial imagery. It involves detection of rigid polyhedral structures, mostly in the form of land vehicles. Buildings and cultural objects found in urban scenes possess many of the same characteristics of those vehicles (i.e. clear faces and edges that form vertex junctions).

In the original problem domain, recognition is performed using a transformation invariant representation of features taken from the image data to index into a database. The database contains multiple representations of objects, each corresponding to a different view of one of the target models. In the case of tie-point selection, one of the two images (the reference image) is considered to be the target view. Low-level feature extraction is done on both images (described in detail below). Features from each of these extractions are mapped into respective vertex space (VS) representations. VS is a 2D map of the extracted features based on two key feature parameters: angular size and orientation. It is insensitive to 3D transformations with the exception of any rotation about the optical axis, which will be determined through a search process.

3.1 Curvature features

The curvature operator measures the orientation of each pixel to determine contour edges. Convolution of Gaussian derivatives is used to find pixel orientation and is subject to a threshold to impose strength requirements on edge contrast. Given an orientation image, the operator examines each pixel for curvature. A circle of a user-defined radius is placed around the pixel of interest. Each pixel that lies on the circle around that central reference pixel is examined for direction consistency. A pixel on the circle has directional consistency with the central pixel if a line extended from the pixel in the direction of its orientation passes through or within some maximum distance from the central pixel. If the pixel on the circle points towards the central pixel it is taken to be on the same contour as the central pixel.

Figure 5 illustrates the operator. If two pixels or clusters of pixels are found to have directional consistency with the central pixel, this indicates that there exists a contour that passes through the two perimeter pixels and the central pixel.

Further support for existence of a vertex is found by taking the pixels that are directionally consistent with the central pixel and stepping outwards in the direction determined by that pixel's orientation. If the pixels found when stepping outward are directionally consistent with the previous pixel then there is greater support for a vertex located at the central pixel.

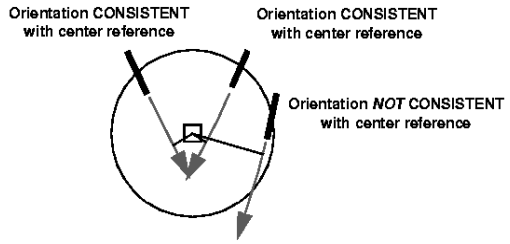

Figure 5: Directional consistency constraint of curvature operator

The curvature operator is likely to produce clusters of candidate vertex pixels in close proximity. To ensure uniqueness, these clusters are merged using morphological region operators. Some cluster regions can be eliminated based on a connected region size threshold (number of pixels) and then nearby regions are combined using merging and splitting. Finally, the resultant regions are collapsed down to a single vertex at the position of the region centroid. Vertex locations are stored along with angular size, orientation, and coordinates for the two endpoints of the vertex at a distance equal to the radius of the inner circle of the curvature operator.

The features that will be stable and consistently extracted over a broad range of viewpoints in aerial images will generally be

larger features, such as vertices found at the corners of buildings or in ground patterns (i.e. roads and sidewalks). These larger cultural features are more likely to be visible from a broad range of viewing angles (off the vector normal to the ground plane) because they are less likely to be obscured. It was found experimentally that larger features corresponding to vertices found along contours that are several meters long provided good candidates for use as tie points. Consequently, the interest neighborhood of the curvature operator was set to only extract vertices with such longer contours. A summary of the parameters is given in Table 3.

Aliasing problems in some of the data were counteracted by smoothing the input image using uniformly weighted averaging.

3.2 Vertex space

Extracted features can be directly mapped into VS using the size and orientation attributes of the features (see Figure 6). The motivation behind using VS is that the search space is 2D thus keeping complexity down (versus six dimensions of search for full 3D representation), and VS is relatively insensitive to 3D translation. However, VS representation is susceptible to rotation about the optical axis, which would manifest as a constant offset between the vertices of two different VS maps. VS is sensitive to general 3D rotation to the extent that perspective distortion due to rotation affects the vertex angular size. VS is insensitive to scale as it is based on the vertex features which are determined only by a localized neighborhood of interest.

For illustration, in Figure 6, an example extraction is given of an image containing two rectangles. The VS map shows that all vertices have the same size (lying in the same size bin) of 90° with a constant shift on the orientation axis. The size of this

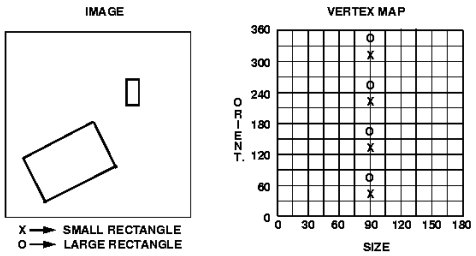

Figure 6: Constant orientation offset shown in VS map

constant shift (between corresponding X's and O's) gives the orientation difference between both objects. If these two objects were to be matched based on VS information alone, the only discrepancy in the correspondence would be this orientation shift. In this trivial example, corresponding corner vertices have exactly the same shift and all support the same hypothesis for orientation offset. In real data, however, due to complexity and poor image quality the candidate orientation offset may not have such strong support. The mechanisms for determining offset are described in detail below.

3.3 Matching

In its original domain, the matching process in the target recognition system determines the correspondence between features in the image data and those in the model data. The appropriate target model is identified from these correspondences, as well as estimates for the relative pose transformation including the 3D orientation of the target and the rotation within the image plane. Relative scale information is also provided from the verification process.

For generation of tie points from stereo image pairs, issues of target model and viewpoint indexing are irrelevant. No search is necessary to find the proper view as it is given as the second image. Accurate knowledge of viewing geometry (besides rotation around the optical axis) is not needed as long as the difference in viewpoints does not cause perspective distortion to affect the angular size of the extracted vertices significantly. If this assumption does not hold for the input image data then a strategy for finding the appropriate difference in viewing perspective must be examined.

The reference image VS map for the hypothesized model view is used to index back into the second image VS map to determine orientation offset hypotheses. Each vertex correspondence between the VS maps defines an orientation offset. Offsets from each of the correspondences are collected in a histogram where peaks indicate hypotheses for the actual relative orientation difference about the optical axis for the two images. The peaks of the offset histogram determine the order in which hypothesis offsets are later verified.

The final stage in the matching process involves verifying the hypotheses produced by indexing. The orientation offset selected through indexing is a hypothesis based only on information available from VS maps, namely vertex size and orientation. To verify these hypotheses only the coordinates of the vertex features in the images are used (size and orientation are ignored in this step). The position of the features can be utilized to examine the relative spatial relationship of the extracted features. Comparison of overall spatial structure of the features between the two images will serve to verify the hypothesis. If correspondence is correctly hypothesized, corresponding objects in both images should yield vertices with a relatively similar spatial arrangements. Orientation offset is also verified by comparing the relative orientation of the spatial structure of the features in both images.

Consistency of the hypothesized orientation offset is verified by first taking pairs of correspondences from each image and determining if they are geometrically consistent with the offset hypothesis. For every pair of features in the reference image, the orientation of the line connecting the two points is calculated and compared with the orientation of the line connecting the corresponding features in the second image.

If the angular difference between orientations of these two lines falls into the same orientation bin as the hypothesized orientation offset, then the pair of correspondences is said to be pair-wise consistent. This consistency checking is done for every pair of correspondences between the images. To further constrain the verification of the geometric relationship between features, each pair-wise consistency found above is searched to find consistencies that have a mutual vertex in common (to ensure that consistency is maintained across two or more vertices). Scale information is then calculated as the ratio of the distances between two correspondences.

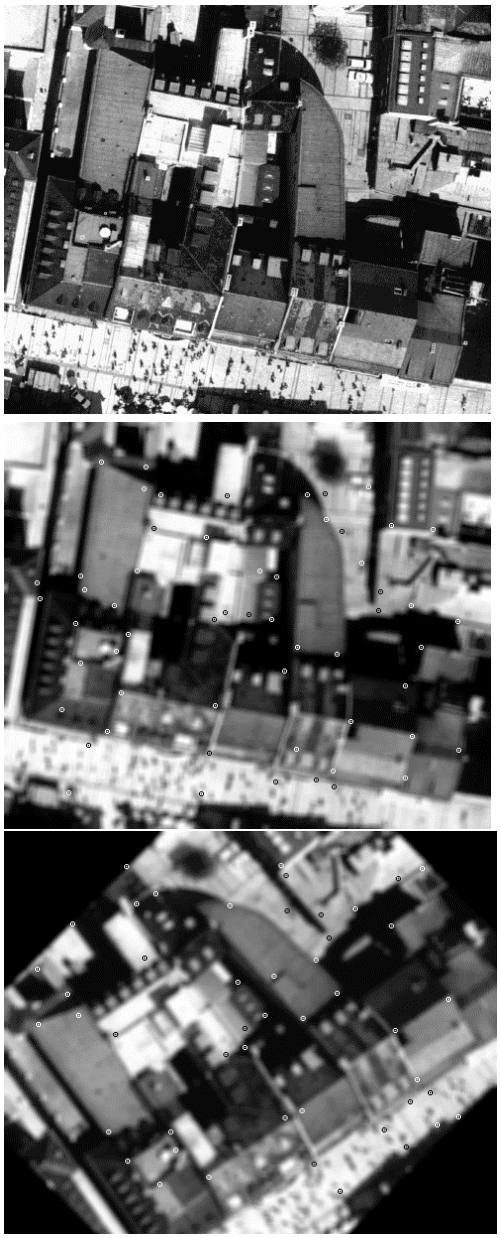

Figure 7: Munich images with extracted vertices (original, smoothed, rotated smoothed)

3.4 Results of vertex space system

The first experiment used the Munich data since it provided typical conditions for aerial urban images. The image, shown in Figure 7, is of an urban area densely packed with building structures (mostly connected) and a few other cultural and natural objects (tree, cars, and people).

To evaluate the matching system, especially the indexing and orientation offset hypothesis generation, a pair of input images was created by taking the original image from Figure 7 as one of the two, and then rotating a copy of that image by some preset amount to create the other image. This manual rotation of the second image is useful since the true orientation offset is known exactly and therefore the hypothesis generated through indexing can be verified.

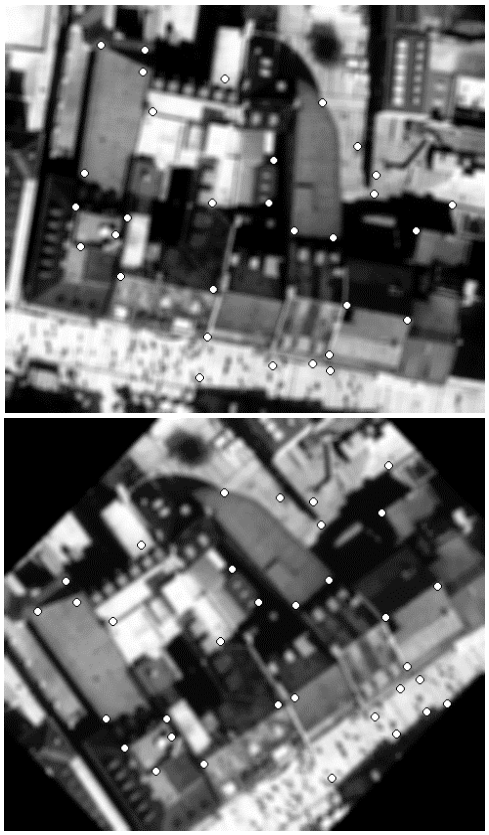

Figure 8: Matched vertices in Munich images

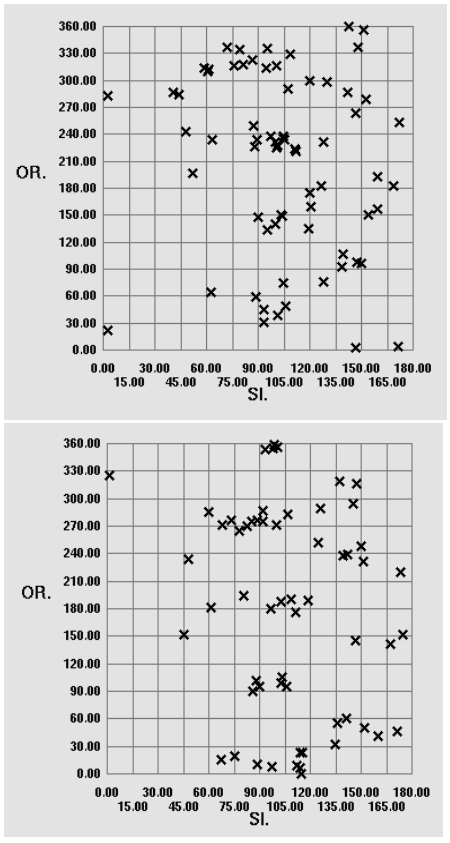

Figure 9: VS maps for original and rotated images in Munich scene

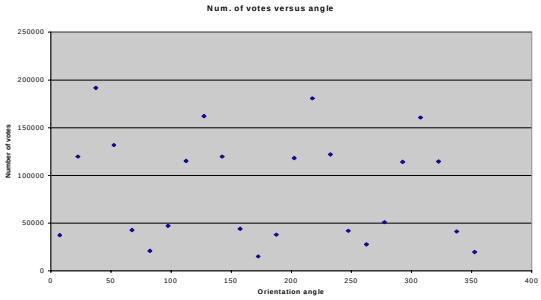

Figure 10: Orientation histogram for Munich images

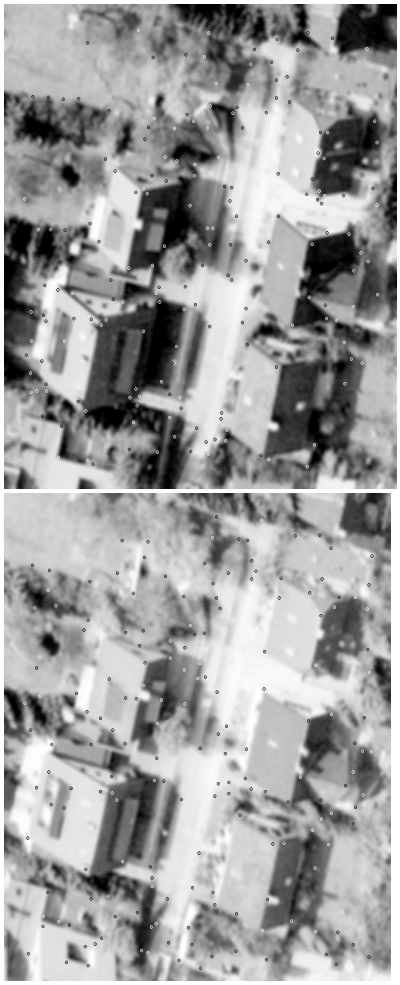

Figure 11: Extracted vertices in Kapellen images

For this experiment, both the original image from Figure 7, and a copy of it rotated 45° were smoothed. This pair of images was passed into the system, with the rotated image coming second in order. 60 vertices were extracted. On operator inspection of the extracted vertices in both images, at least 17 vertices should intuitively match as they are extracted from well-defined corners of buildings. Later it will be shown that all of these candidate matches were found.

As was described in Section 3.1, the feature extraction process begins by calculating the gradient direction at each pixel. The gradient mask is applied in the horizontal and vertical directions

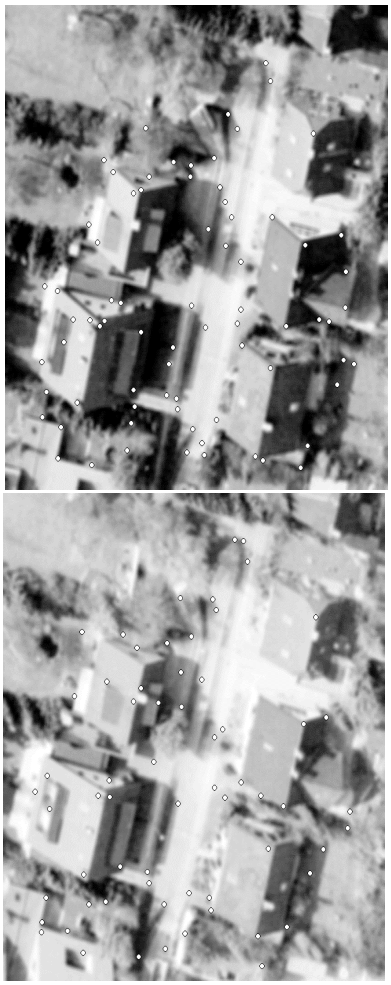

Figure 12: Matched vertices in Kapellen images

from which, by taking each directional response to be a component of a 2D vector, the absolute direction of the pixel (as belonging to a local edge) can be calculated. As a result of this calculation an orientation field is generated through assignment of a local orientation to each pixel (determined by a neighborhood corresponding to the gradient kernel size).

In the post-processing stage, regions created by the curvature operator are manipulated to combine adjacent vertices and to eliminate vertices which do not have enough curvature support. After this manipulation is complete, a single vertex is extracted from each connected region.

We can see from Figure 7 the effect of requiring larger features. It precludes vertices that lie on shorter but strong edges from being extracted. Also, poor contrast on some edges weakened the gradient signature thus making many edge pixels fall below the gradient magnitude threshold.

The first step in the matching process is to index vertices from one vertex space map to the other to generate a histogram of possible relative orientation offsets. Vertex space maps for each of the input images are given in Figure 9. It should be stated that the bin boundaries shown on the vertex space maps in those figures are not accurate with respect to the actual system. Actual bin size for the vertex space representation used for indexing is 10° on both the size and orientation axes. As each

reference is made between two vertices (which lie in the same size bin) from the two different images, a vote is cast towards a certain orientation offset based on the difference between the orientations of the two vertices. The votes are tallied in a histogram and the offset bins are verified in order based on the number of votes. The histogram, shown in Figure 10, correctly indicates that 45° is the most likely offset between the two images as that respective bin received the greatest number of votes.

After formulating the offset histogram, the hypotheses can be verified using the geometric positional constraints described in Section 3.3. Each offset hypothesis is verified by applying these constraints. 36 vertices were matched automatically by the system. After operator inspection, it was found that 6 of these matches were incorrect. The final matched vertices are shown in Figure 8.

The system was also run on images that are a subset of the Kapellen data set of the ISPRS tie-point data set collection. Figure 11 shows the input aerial images (with extracted vertices) which give two slightly different perspectives of the scene. These images were not smoothed before being run through the extraction and matching system. Due to the size and number of natural features (foliage) in the images, a larger number of vertices were extracted. 219 vertices were extracted from both images. Only 84 of the 219 vertices matched because the features extracted from foliage regions were vastly different due to the difference in visible shadows. Figure 12 shows the matched vertices. On operator inspection it was found that the system had matched 44 vertices correctly. However, the correct relative orientation offset was found through indexing to be 0°.

4 CONCLUSIONS AND FUTURE IMPROVEMENTS

This paper shows initial findings from a work in progress. We will continue to analyze the results and make further improvements, with the goal to build a system that may make use of the different approaches under different contexts. We are in the process of integrating the tie-point algorithms with a bundle-adjustment procedure. This will allow us to implement better control strategies. Such better control structure could use orientation and error information output from the bundle adjustment to direct subsequent extractions of features, thus iteratively minimizing residual error in extracted feature locations.

The first approach investigated performed as we would expect from the literature. We are currently adding more topological constraints to the system. The possibility of developing a test bed to use such constraints was a major factor in pursuing this avenue in the first place.

The second approach appears to be quite applicable to the task, especially with respect to urban scenes. It provides for a mechanism to recover relative orientation and scale with no initial estimate, which the first one does not.

Clearly a more comprehensive study of performance and error analysis using bundle adjustment are essential and will be performed. We are in the process of collecting more data with our own sensor package, which also records video data.

Further efforts will be made to extend the techniques to handle not only image pairs, but strips and blocks as well.

5 ACKNOWLEDGEMENTS

We wish to thank Chris Holmes and Adam Stachelek for their help with programming and data processing.

REFERENCES

Ackerman, F. and Tsingas, V., 1994. Automatic Digital Aerial Triangulation, ASPRS/ACSM Annual Convention, Reno, pp. 1-12.

Förstner, W., 1994. A Framework for Low Level Feature Extraction, European Conference on Computer Vision, pp. 383-394.

Krzystek, P., 1998. On The Use Of Matching Techniques For Automatic Aerial Triangulation. International Archives of Photogrammetry and Remote Sensing, 32: 58-67.

Schenk, T. and Toth, C., 1993. Towards an Automated Aerial Triangulation System, ASPRS/ACSM Annual Convention.

Tang, L. and Heipke, C., 1994. An Approach for Automatic Relative Orientation, Optical 3-D Measurement Techniques II. Wichman Verlag, pp. 347-354.

Whitten, G. and Rosenfeld, A., 1996. Vertex Space Analysis for Model-Based Target Recognition. Final Report for ONR program #N00014-95-1-0859.

Parameter	Value
Minimum interest operator window size	5x5
Maximum interest operator window size	22x22
Minimum correlation window size	5x5
Maximum correlation window size	22x22
Correlation threshold	0.8

Table 1: Parameter values for first approach

Data set	Total # of points detected	# of points after topology test	# of correct points after topology test	% outlier points
Munich	5244	42	33	21.4
Echallens	41870	42	38	9.5
Winrock	2235	14	12	14.3

Table 2: Results of first approach

Parameter	Value
Inner Radius of Curvature Operator	10
Outer Radius of Curvature Operator	30
Gradient Kernel Size	9x9
Gradient Smoothing Kernel Size	9x9
Gradient Magnitude Threshold	10
Image Smoothing Kernel Size	7x7
Curvature Region Size Threshold	60

Table 3: Parameter values for second approach

Data set	Total # of vertices	# of vertices matched	# of correct vertices matched	% outlier points
Munich	60	36	30	16.7
Kapellen	219	84	44	47.6

Table 4: Results of second approach