Scene Registration in Aerial Image Analysis

Frederic P. Perlant
David M. McKeown

March 27, 1989

CMU-CS-89-127$_2$

Digital Mapping Laboratory
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

This report is a revised and extended version of a paper presented
at the *SPIE Conference on Reconnaissance, Astronomy, Remote Sensing, and Photogrammetry*
Los Angeles, California. January, 18-19, 1989.

This research was primarily sponsored by the U.S. Army Engineering Topographic Laboratories under Contract DACA72-87-C-0001 and partially supported by the Defense Advanced Research Projects Agency, DoD, through DARPA order 4976, and monitored by the Air Force Avionics Laboratory Under Contract F33615-87-C-1499. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Engineering Topographic Laboratories, or the Defense Advanced Research Projects Agency, or of the United States Government.

Table of Contents

List of Figures

List of Tables

Abstract

In this paper we discuss the importance of scene registration for several tasks in the automated interpretation of aerial imagery. These tasks are structure matching, stereo matching, and stereo visualization. While the processes of registration and matching have traditionally been treated as separate problems, particularly in the case of stereo matching, we describe techniques that may unify these processes. We also demonstrate the automatic generation and matching of control points in complex aerial imagery and show that the resulting registration is comparable to that achieved using manual control point selection. Finally, methods for the generation and visualization of stereo disparity images and stereo ground truth scene segmentations are described.

1. Introduction

Scene registration is a fundamental requirement for a number of image analysis tasks such as stereo matching, multi-image matching for temporal changes, and image sequence or motion analysis. As a result there exists a rich variety of techniques to perform scene registration. For example scene registration can be accomplished by identification of image points to a common frame-of-reference via control points whose three-dimensional location is accurately known. Registration can also be accomplished in a relative manner by identifying corresponding points between one or more images, i.e., the establishment of image-to-image control points. The position of these points need not to be known in the three-dimensional world. For some applications registration can be accomplished with respect to a cartographic map, a photomosaic, or an orthophoto that has been warped in order to remove position distortions due to terrain relief.

Depending on the type of registration there arise many issues in accuracy. Accuracy is generally evaluated within the context of a particular task requirement. Often techniques with some inherent inaccuracy can suffice in many task situations. If we look at traditional photogrammetric techniques for recovering the position of an image we require detailed information of the camera, its focal length and lens characteristics, the platform coordinates in terms of height above the ground and its three-dimensional position (pitch and yaw). Inaccuracies arise when we are unable to know any of these parameters precisely. In the case of digital imagery, the image formation process can introduce additional errors. For example, the sampling process, in terms of geometric precision and radiometric accuracy, is rarely modeled for digital imagery generated by digitizing photographic film. There are a variety of digitization apparatus including rotating drum scanners, flatbed optical scanners with a single optical element, scanners composed of linear arrays of elements or a full three-dimensional element array. Each has its own inherent inaccuracy. The kinds of errors introduced include those due to sampling a continuous tone photograph into a discrete intensity range over an arbitrary sampling window and limitations in the sensors' dynamic response. The geometric accuracy with which the scanner is positioned and moved over the film varies greatly between scanning methods. Even once we have accurate image, sensor, and platform information our ability to locate ground control points accurately in digital imagery is independent of the inherent accuracy of those control points.

This paper raises issues in how scene registration can be achieved in digital imagery and illustrates the importance of accurate registration for three analysis tasks. In Section 2 we discuss some general issues in registration in computational vision. These issues include use of a spatial database to provide coarse ground control information, the selection of manual control points, and the automatic determination of control points. In Section 3 we discuss the importance of registration to three particular tasks in the interpretation of aerial imagery. These tasks are the correlation (or fusion) of monocular analysis from partially overlapping views, computational stereo matching techniques, and the visualization of stereo

matching results. We also discuss the interrelationship between registration and matching. In Section 4 we describe the results of a fully automatic scene registration from initial coarsely registered stereo pair to a final three-dimensional interpretation. Finally, in Section 5, we present conclusions and discuss future directions for our research.

2. Scene Registration

In this section we describe some basic principles of stereo photogrammetry, a part of computational photogrammetry [25]. The primary goal of stereo photogrammetry is to determine the three-dimensional position of any object point that is located in the overlap area of two images taken from two different camera positions. The determination of the orientation of each camera at the moment of exposure and the relationship between the cameras is a necessary step in the photogrammetric process. In Section 2.1 we discuss the problem of camera orientation that determines the relationship between the image points and ground points in the scene. In Section 2.2 we describe the classical epipolar geometry for stereo imagery. When two images are registered in the epipolar geometry the spatial relationship between corresponding points in the left and right images is greatly simplified.

2.1. Camera orientation problem

The solution to the general camera orientation problem has four components: the interior orientation, the exterior orientation, the relative orientation, and the absolute orientation.

The *interior orientation* refers to the perspective geometry of the camera. The parameters of the camera are generally known *a priori* and can be determined by precise calibration. This includes the focal length, the position of the principal point in the image plane of the camera and the geometric distortion characteristics of the lens system. These parameters are intrinsic to the camera and are generally detailed on a standard camera certificate.

The *exterior orientation* characterizes the orientation of the camera during the image event. It is defined by the geographic position of the optical center in a three-dimensional rectangular coordinate system and the direction of the optical axis. Therefore, the exterior orientation determines the projective relationship that exists between the image coordinates of the image points and the ground coordinates of the corresponding object points in the scene. In the context of stereo photogrammetry, the exterior orientation can be decomposed into the relative orientation and the absolute orientation.

The *relative orientation* determines the relative three-dimensional position of the two images in the stereo pair with respect to each other. As shown in Figure 2-1 three lines characterize the exterior orientation of the two cameras. Two lines are the rays emanating from an object point P and passing through each of the optical centers, C_L and C_R, of the two cameras. These lines are P to P_L (passing through C_L) and P to P_R (passing through C_R). The third line is C_L to C_R passing through each of the optical centers. This is called the baseline of the stereo model. These three lines are represented in a three-dimensional rectangular coordinate system and must be coplanar. This coplanarity relationship gives an equation with twelve parameters that defines the exterior orientation of the two cameras. Of the twelve parameters, just five are necessary to define the relative orientation of the two cameras. In order to determine the five parameters, we need five pairs of corresponding points (P_L,P_R) in the left and right image. Each left/right pair defines one coplanarity equation.

After the relative orientation is accomplished, the stereo model must be scaled, translated, and leveled

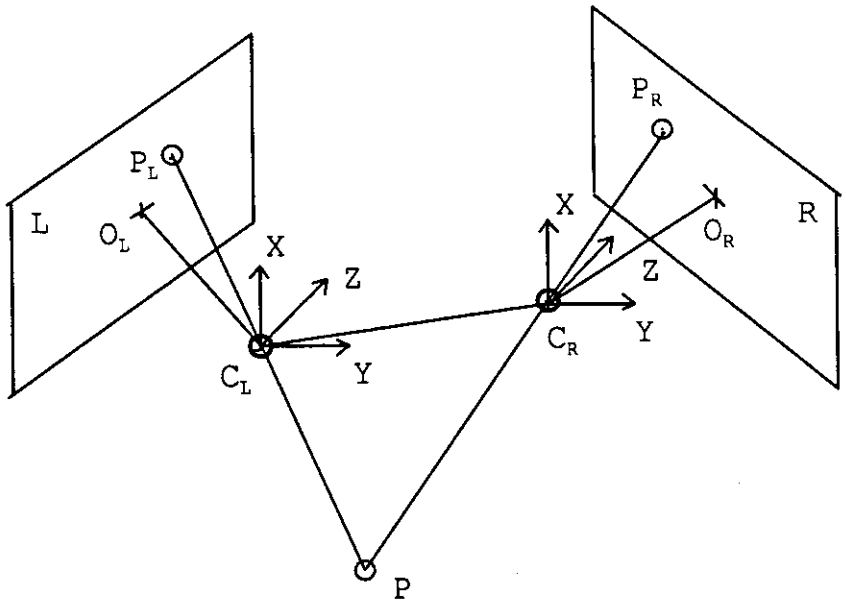

Figure 2-1: Relative orientation of stereo imagery

with respect to a ground reference system. The process of orienting a stereo model into an absolute reference system is called the *absolute orientation.* It relates the absolute coordinates of an object point in the ground reference system to its coordinates in the model coordinate system of the camera. Each control point, for which the ground coordinates are known, gives rise to three projective transformation equations when its model coordinates are measured. Three control points are necessary to define all the parameters of the absolute orientation.

In our work we consider that the interior orientation or calibration has been already performed since we assume an ideal pinhole camera. Therefore, the calculation of the relative orientation is the registration problem [12]. Knowing the relative position and attitude of the two images in the stereoscopic pair with respect to each other defines the relationship between the two images of the scene. All of the results presented in this paper will be relative measurements. However, as we have discussed, these relative measurements could be used to calculate absolute metrics, such as height, length, and area by using three-dimensional ground control points to establish the absolute orientation.

2.2. Epipolar geometry

After we have established the relative orientation of the two images, it is possible to reformulate this relationship into the epipolar geometry. The *epipolar geometry* defines a constraint based on the geometric relationship that the plane containing the two optical centers of the cameras, C_L and C_R, and the ground point P, intersect the two image planes on two lines. These lines are called the conjugated epipolar lines. As shown in Figure 2-2 these lines, E_L to P_L and E_R to P_R, contain the two image points P_L and P_R. These lines emanate from a common epipolar center, E_L and E_R. This epipolar center corresponds to the intersection of the image plane and the stereo baseline C_L to C_R. Thus, the points in the left and right image (P_L and P_R) correspond to a single three-dimensional scene point (P) and are on the same conjugate epipolar lines E_L to P_L and E_R to P_R. After this relationship has been established, it is common to register the two images so that the conjugated epipolar lines become corresponding scanlines in the left and right image. Therefore the corresponding points are on the same scanline in each image and the displacement between the points, or disparity, corresponds to the relative height of

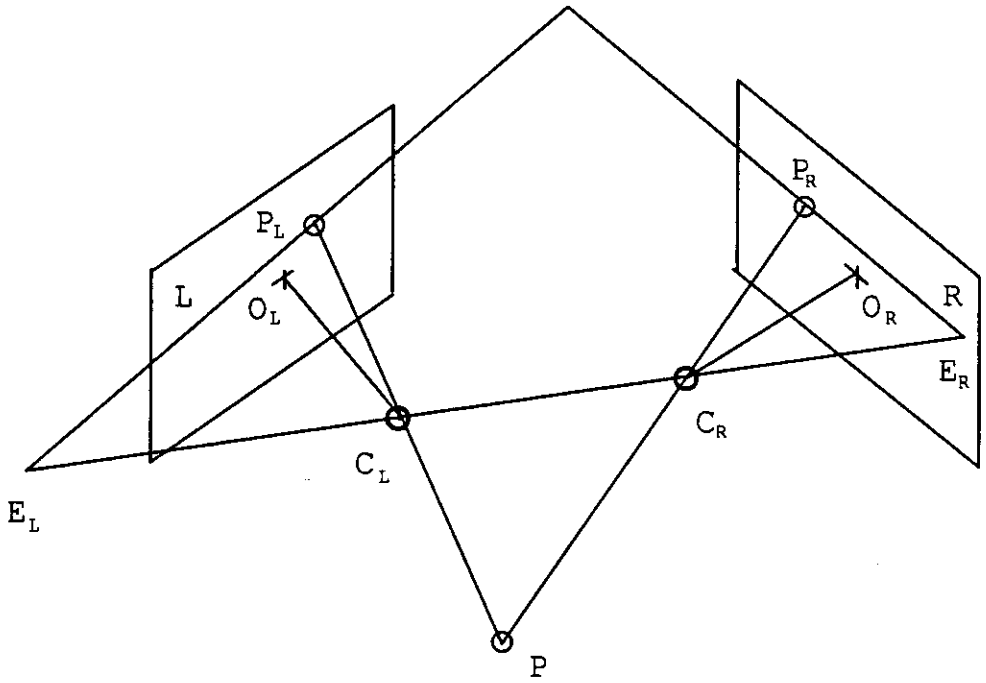

Figure 2-2: The epipolar geometry

the three-dimensional scene point.

This epipolar geometry is a common framework for most stereo matching algorithms: [1], [5], [6], [22], [24]. These stereo matching techniques assume that the registration is ideal and that the epipolar constraint is completely satisfied. Some researchers have attempted to explicitly account for the inaccuracy of the image registration and have attempted to improve it by preprocessing the imagery before beginning the matching process [8], [9], [11], [29]. Modeling inaccuracy in image registration has most often been studied within the context of the robotic applications where it is common to have a good deal of control over the cameras [10] and where a detailed preregistration and calibration step is possible. However, for many applications in aerial image analysis one is often simply given overlapping images or partial image areas where the epipolar geometry must be derived.

In the following section we present two methods for scene registration given overlapping stereo imagery. The first method performs a coarse registration using landmarks from a spatial database. The second method uses pairs of corresponding points in the two images to perform a relative orientation. As we will see, many of the techniques used in computer vision to establish scene registration are approximations to the photogrammetric ideal. These approximations cause the scene registration to be inaccurate. The effect and implications of these inaccuracies will be explored within the context of two matching tasks.

2.3. Coarse registration using a spatial database

The most common method to establish the relative orientation between two images is to select pairs of corresponding points in the two images. One alternative method is to independently tie each image to a common frame of reference. A cartographic coordinate system such as <latitude,longitude,elevation> is one possible frame of reference. Thus, the two images are related to a ground coordinate system, or map. The use of landmarks with known <latitude,longitude,elevation> is a common method to orient

each image. The overall accuracy of the registration is dependent on the accuracy of the three-dimensional position of the landmark and the accuracy with which we can recover the image position of the landmark. We use the landmark database component of CONCEPTMAP, a spatial database system that integrates imagery, terrain, and map data to provide landmark descriptions [17, 18]. Each landmark description in the database has a reference image fragment, and a ground position definition which contains the <latitude,longitude,elevation> information, its position in the reference image fragment, and a brief textual description of the landmark for the user. Each image in the CONCEPTMAP database is put into correspondence using manual selection of landmarks.

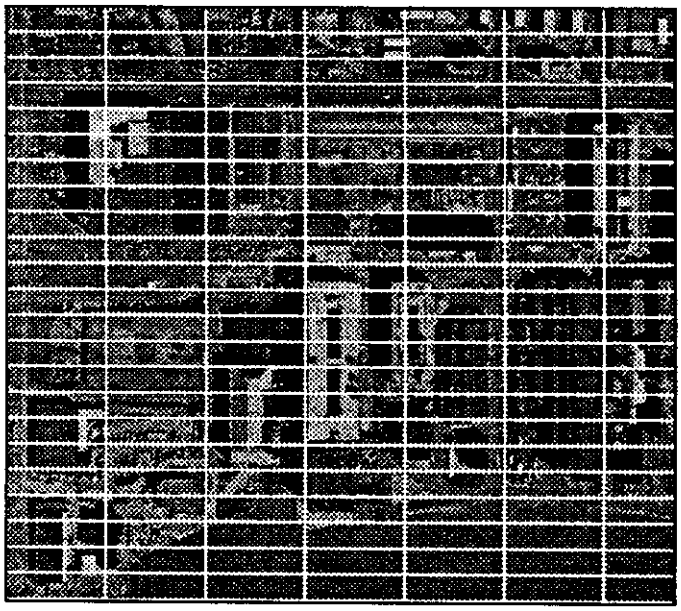

Figure 2-3: Left image DC38008 with CONCEPTMAP database registration

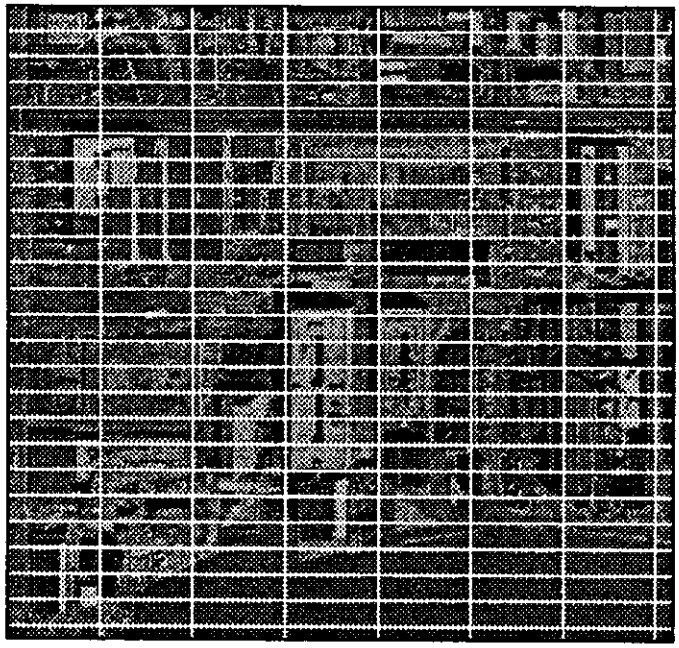

Figure 2-4: Right image DC38007 with CONCEPTMAP database registration

Figure 2-3 and 2-4 show a stereo image pair of an industrial area taken from the CONCEPTMAP database. These images were digitized from standard nine inch format mapping photography taken at the altitude of 2000 meters by a 153 millimeters camera. One pixel corresponds to 1.3 meters on the ground. The left image is a 512 x 512 sub-area selected from 2300 x 2300 image. The right image sub-area was generated by calculating the <latitude,longitude> for the corner points of the left image and projecting those points onto the complete right image. This projection is then used to extract the image sub-area from the complete right image. We have superimposed a set of gridlines on both images in order to make it easier to see the actual misregistration. Typically CONCEPTMAP provides a registration accuracy of between ten to thirty meters for imagery digitized to a 1.3 meter ground sample distance.

2.4. Fine registration using image control points

As we have seen, the computation of the relative orientation can be accomplished by selecting pairs of corresponding points in the two images. After the relative orientation is calculated, two images can be transformed so that they satisfy the epipolar constraint. We begin the fine registration with the coarse registration described in the previous section.

We make several assumptions that simplify the relative orientation model. We assume that the cameras are metric and have the same interior orientation. We also assume that their optical axes are parallel and are, in fact, vertical. Because we are using aerial imagery taken by the same camera along the same flightline, these assumptions are not unreasonable. The largest source of error is whether the camera platforms were at precisely the same altitude and orientation at each imaging event. Given these assumptions, the transformation between the left and right image is only a translation and a rotation, because the image planes are the same. Therefore, of the five parameters describing the general relative orientation, only two remain to be solved. The absolute distances in the two images are preserved and the epipolar lines are already parallel. After the transformation the epipolar lines correspond to the scanlines.

Problems with the accuracy of point selection lead us to use more points to determine the transformation between the two images. On some images this model was not flexible enough to account for the variation in the ground elevation and the variations to our ideal sensor model. As a result we developed a polynomial transformation adjusted by least squares to fit the selected corresponding points. In the following sections we discuss the manual and automatic selection of image control points.

2.4.1. Manual selection of common points

The classical method to select corresponding points in order to perform interior orientation is by the manual identification of landmark points in stereo imagery. Typically, man-made features such as road intersections, boundary corners of fields or parking lots, or markings such as road centerlines are used because of the ease with which they can be found in the imagery. We chose to manually select shadow corners since these points were the focus of our experimentation in automatic landmark detection. Given that we are working in an urban environment, shadow corners have the advantage that they are generally on the ground and therefore in the same plane, assuming only small changes in terrain elevation. Although the shadow position changes as the sun moves, if we have imagery taken at nearly the same time, as is common in aerial mapping photography, the shadow corners will fall on the same point in the three-dimensional scene. Such corners also tend to be uniformly distributed in scenes containing large numbers of buildings. The manually selected shadow corners give us a baseline against which we could measure the accuracy of the automatic landmark selection process. Figures 2-5 and 2-6 show the manually selected shadow corners in the left and right images respectively.

2.4.2. Automatic selection using shadow corners

Clearly, one requirement for automated registration is the automatic selection of corresponding points in the stereo pair images. There are actually two problems that must be solved. First we must automatically detect potential landmarks in each image, and then we must determine those landmarks that have been found in both images. General landmark matching is an unsolved problem and most automatic registration techniques relie on the matching of characteristic points [21] that often have no physical significance or reference with respect to landmarks.

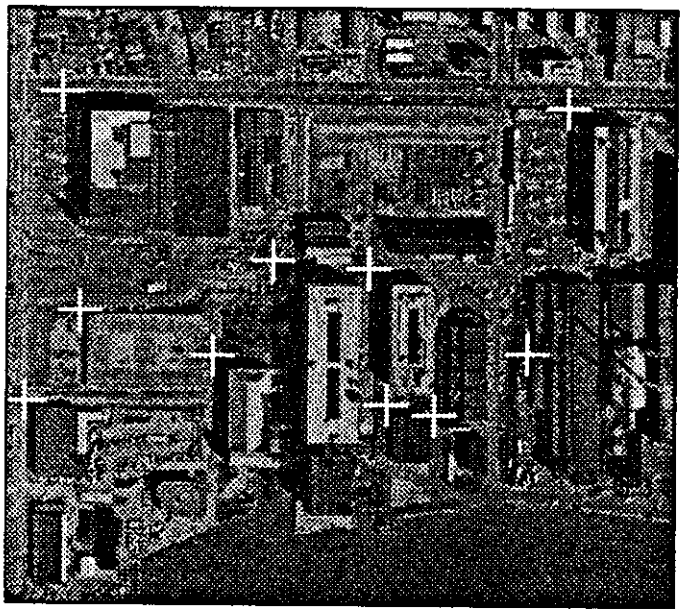

Figure 2-5: Manual selection of points [left image] with coarse registration

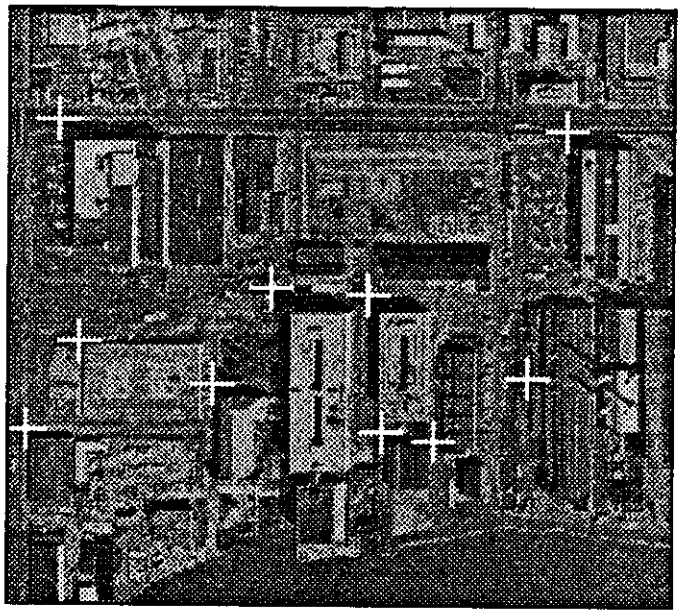

Figure 2-6: Manual selection of points [right image] with coarse registration

Figure 2-7: Automatic selection of points [left image] with coarse registration

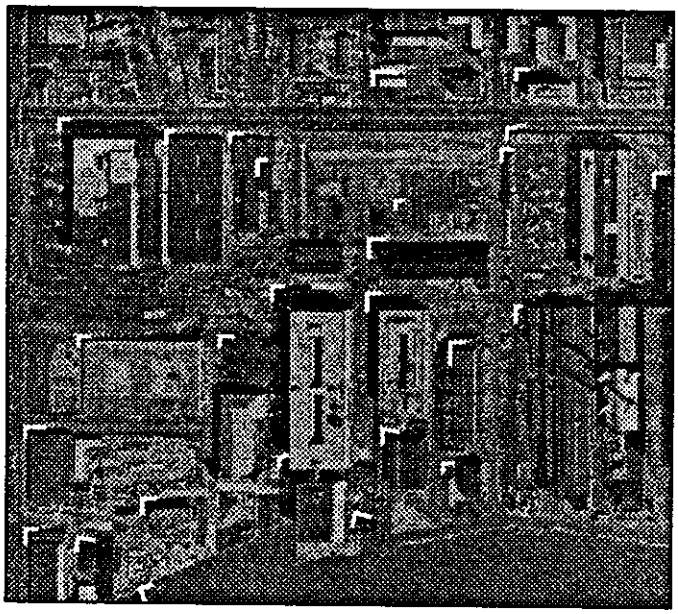

Figure 2-8: Automatic selection of points [right image] with coarse registration

For this experiment, we assume that a coarse registration of the two images, such as described in Section 2.3 has already been performed. Using this coarse correspondence, we are able to limit the search to find corresponding features in the images. Most of the remaining error is translational rather than rotational which simplifies the determination of corresponding points.

As described in the previous section, shadow corners are good candidates for automatic detection and

correspondence as well as for manual selection. We use a monocular detection of shadow regions and determine the boundary line between the shadow and the building [14]. This boundary is used to determine the position of the shadow corner in the left and the right images [2]. After removing corners that were inconsistent with shape and orientation constraints imposed by the sun angle, we selected sets of shadow corners that were detected in both images. Figures 2-7 and 2-8 show these corresponding shadow corners on the two images. Note that the corners selected differ from those selected manually.

Figures 2-9 and 2-10 show the results of the fine registration using shadow points selected automatically. This registration is obviously better than the coarse registration using the CONCEPTMAP database shown previously in Figures 2-3 and 2-4. In the following section we attempt to quantify the registration quality.

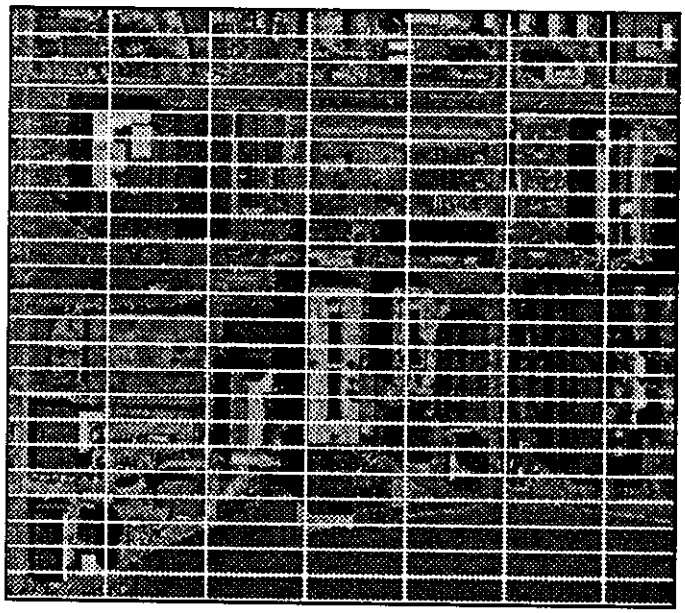

Figure 2-9: Left image of the fine registration

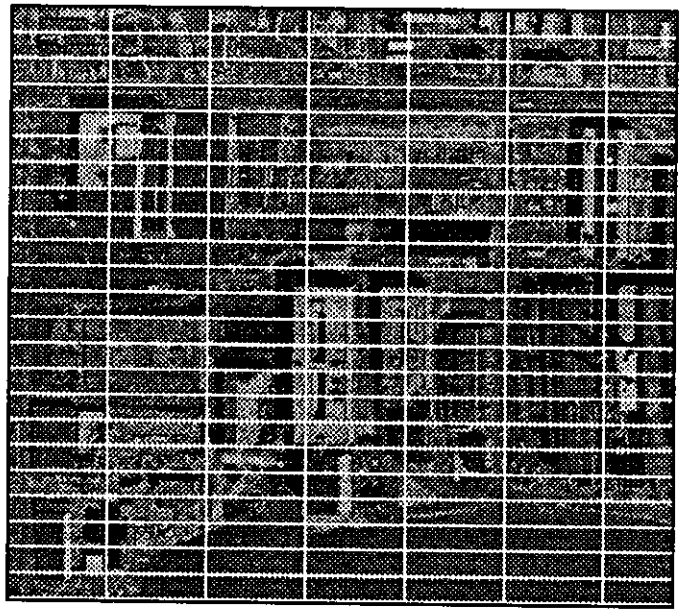

Figure 2-10: Right image of the fine registration

2.4.3. Quality of registration

Tables 2-1 and 2-2 show the local accuracy of the different scene registrations performed on DC38008 and LAX stereo image pairs.[1] The first three rows of each table characterize the quality of the CONCEPTMAP registration using three set of control points: the points selected manually, the points generated by automatic detection of shadow corners, and the points derived from structure matching. In the case of DC38008 11 conjugate point pairs were manually selected, 26 shadow corner pairs were automatically extracted, and 16 point pairs were found using structure matching. In the case of the LAX stereo pairs, these numbers are 14, 13, and 16, respectively. Because the CONCEPTMAP coarse registration is derived by a polynomial fit for the entire scene (2300x2300 for DC38008 and 2000x4000 for LAX), it is interesting to evaluate the quality of the local fit for the 512x512 image sub-areas using each set of independently derived control points. The CONCEPTMAP registration produced a translation to within approximately 12 pixels (16 meters) of the the true registration for both images. This registration was quite consistent across all three sets of test points.

[1]We will introduce some matching results for the LAX airport scene in Section 3.1.2.

For the other two types of scene registration, isometrical and polynomial, we evaluated the quality of registration with respect to the manually derived control points. That is, the solutions for manual, corner, and structure matching were validated using the manual points. In all cases both registrations achieved significantly better results than the CONCEPTMAP coarse registration. In several cases the registrations achieved by matching shadow corners and structures is quite comparable to the manual registration. However, the manual registration is in all cases as good as any of the automatic control point experiments. In all cases the manual selection of corresponding points produced a registration of less than one meter, or subpixel accuracy. In some cases similar subpixel results were achieved using the automatic point selection. Finally, the polynomial approach led to better results although the simpler isometrical model gave comparable results.

Statistics on the quality of the different registrations for DC38008						
Type of registration	Number of points	Avg. row offset	Std. row offset	Min/Max row off.	Avg. col offset	Std. col offset
Coarse manual	11	-12.4	1.6	-15/-8	905.5	1.2
Coarse corner	26	-13.2	1.6	-18/-10	905.7	1.7
Coarse structure	16	-12.1	1.6	-15/-8	909.3	3.4
ISO manual	11	0.1	0.7	-1/2	1.3	1.3
ISO corner	11	1.7	1.7	-1/7	5.1	1.2
ISO structure	11	0.5	0.6	0/2	-3.4	1.4
POLY manual	11	0.1	0.3	-1/1	0.1	0.5
POLY corner	11	-0.2	1.8	-3/4	0.0	1.6
POLY structure	11	0.1	0.5	-1/1	-3.3	1.5

Table 2-1: Statistics for different registrations on DC38008 stereo pair

Statistics on the quality of the different registrations for LAX						
Type of registration	Number of points	Avg. row offset	Std. row offset	Min/Max row off.	Avg. col offset	Std. col offset
Coarse manual	14	10.6	0.9	9/13	1866.6	0.7
Coarse corner	13	10.9	0.7	9/12	1866.8	1.4
Coarse structure	16	10.9	0.4	10/12	1869.3	1.7
ISO manual	14	-0.4	0.9	-2/2	0.6	0.7
ISO corner	14	-0.4	0.9	-2/2	1.6	0.7
ISO structure	14	-0.4	0.9	-2/2	-2.4	0.7
POLY manual	14	0.0	0.1	-1/1	0.1	0.7
POLY corner	14	1.3	1.0	-1/3	1.5	0.9
POLY structure	14	-0.3	0.7	-1/2	-2.9	0.9

Table 2-2: Statistics for different registrations on LAX stereo pair

One additional issue is how well our local solution performs as a global registration in other areas of the complete stereo pair. In Table 2-3 we show the results of using our local fine registration for both the isometrical and polynomial methods in four quadrants of the complete stereo pair. In each of the four quadrants we manually selected 12 control points and used the manual solutions for DC38008 to calculate residual errors. Because of the large variation in the row and column offsets, it is clear that the

local model can not be treated as a global model even though the row residuals stay within reasonable bounds. However, it is the case that the fine solution should be a better global solution than the CONCEPTMAP coarse registration.

Quality of the registrations for the complete image DC38008						
Type of registration	Region in image	Avg. row offset	Std. row offset	Min/Max row off.	Avg. col offset	Std. col offset
ISO manual	North	1.7	1.5	0/4	4.5	0.8
	West	-1.3	0.4	-2/-1	1.4	1.2
	East	1.3	0.6	0/2	-2.4	1.5
	South	-2.5	0.8	-5/-1	3.1	1.1
POLY manual	North	-1.6	1.2	-4/0	-70.5	16.6
	West	-1.3	0.7	-2/0	-2.4	3.0
	East	-0.1	0.3	-1/1	2.7	3.7
	South	0.8	0.7	-2/2	-51.5	14.3

Table 2-3: Statistics for different registrations on DC38008 stereo pair

Although traditional error analysis can give us an idea of relative accuracy for each of these approaches, this does not necessarily translate into the effectiveness of the registration. That is, for many tasks in scene analysis a coarse-grain registration to within 10 to 30 meters is quite adequate, especially considering that the imagery covers several square kilometers. For instance, tasks that require assembling a collection of image subareas taken over time for change detection and analysis can be supported using this level of accuracy. However, for other tasks, such as matching and stereo analysis, the effect of mis-registration may become more critical. In Section 3 we will see how such tasks are affected by coarse and fine registration.

3. Tasks Requiring Accurate Scene Registration

In this Section we describe three scene analysis tasks that require or support scene registration. These tasks are matching structures derived by monocular analysis of overlapping imagery, traditional stereo matching using area-based and feature-based matching techniques, and the construction of a three-dimensional image to present matching results to a human using a stereo display monitor.

In the case of matching of monocular structures, we can acquire additional information about the actual structure of the objects, including their height, as a result of the matching process, and also generate new automatic control points to refine the registration. The goal is to match high-level structures in two overlapping images, where the actual detection and delineation of the structures is likely to contain significant errors, and matching is complicated due to a large numbers of false alarms produced by the structure generation process.

In stereo analysis the goal is to automatically match points in the left and right images of the stereopair in order to establish a *disparity* between these points. This disparity can be used, along with the camera model, to calculate the actual height of the matched point in the three-dimensional scene.

Finally, it is becoming increasingly important for researchers to be able to visualize the three-dimensional models that their analysis programs are generating. Such a visualization tool allows us to directly compare these results to three-dimensional ground truth models for performance evaluation.

3.1. Correlation of monocular analysis

There are many situations where overlapping coverage imagery is available but may not be suitable for stereo matching due to sensor acquisition parameters, temporal or seasonal changes, or image scale. The issue then becomes one of how to relate the results of independent monocular analysis. One of the first examples in the literature was symbolic change detection [26, 27] and the matching of coarse regions such as lakes, fields, and forests based upon relationships that were largely invariant over small rotations in the image plane (< 45 degrees) and relatively large scale changes (factor of 10 resolution). These techniques have been generalized to the matching of semantic network descriptions generated by separate monocular analysis or from a baseline cartographic description [28].

Our interest in matching of monocular interpretations arises from our desire to relate structure descriptions generated from a building hypothesis system. The BABE Built-up Area Building Extraction system [4], performs monocular analysis on an image by extracting lines and corners and generating structure hypotheses. This work is similar to Huertas and Nevatia [13], but differs in that a large number of hypotheses are purposely generated such that buildings are rarely missed. These structures are then evaluated by a number of techniques such as shadow verification, shadow prediction, and shadow grouping [14]. The processes of verification, prediction, and grouping are used to rank order or prune the large number of BABE structure hypotheses.

3.1.1. Structure Verification

First of all monocular matching can be viewed as another form of structure verification. That is, sets of independently derived hypotheses from different images are matched using the scene registration model to relate absolute ground position in the two images. The results of this matching provide information that is not available in a single image, including an estimate of structure height and the reliability of each hypothesis. For example, because matching allows multiple hypotheses in one image to correspond with a single hypothesis in the second image, we can use this fact to guide a re-examination of the structure delineation in the first image. The fragmentation of structures is a common source of error in computer vision, and understanding fragmentation requires some external process to predict its occurrence or to identify situations where it has occurred. Even in cases where there is a good one-to-one match between structures, different viewing angles, accidental alignments of objects in the scene, or differences in imaging conditions will produce differences in the segmentation which can be recognized as cues for further interpretation.

The use of high-level image cues such as aligned or oriented structures composed of lines, corners, and surfaces for perceptual grouping and stereo matching has recently seen some research activity [7, 23]. In our examples we have focused on the verification and grouping of hypotheses in order to improve monocular analysis. The determination of height is only one component of the matching process, rather than the primary result. We also used this matching as another way to select control points automatically and to perform scene registration.

3.1.2. A Matching Experiment

In this section we describe our matching results on a portion of a stereo pair of Los Angeles International Airport (LAX) used by Huertas and Nevatia in their building extraction research [13]. Figures 3-1 and 3-2 show the results of BABE hypothesis generation on the left and right image, respectively. The BABE results have been pruned automatically using shadow analysis [4]. As is evident in these results, BABE generates hypotheses for most of the buildings in the scene. However, it is apparent that there are differences in the quality of the delineation and in detection errors between the

two images. In the ideal case, we should have the same number of building hypotheses in each image. Further, the roof delineation of each building should be quite similar but not identical, because of the displacement due to height.

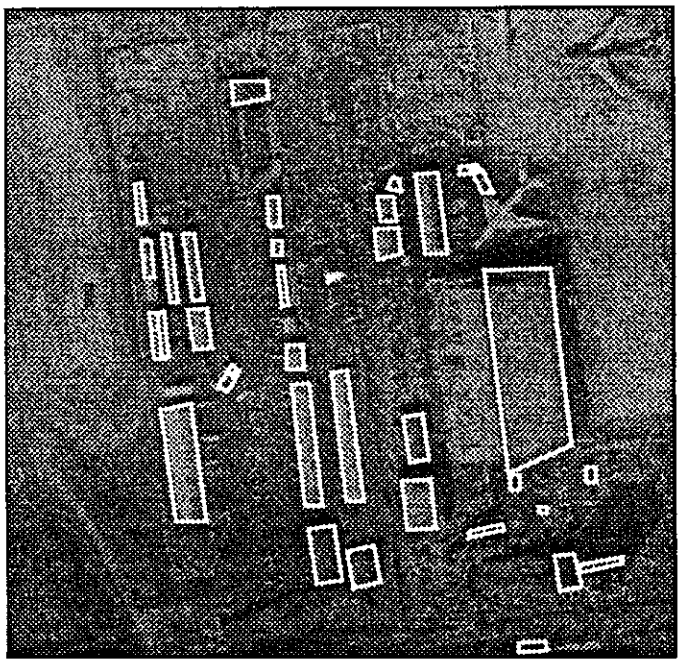

Figure 3-1: BABE building hypotheses for LAX [left image]

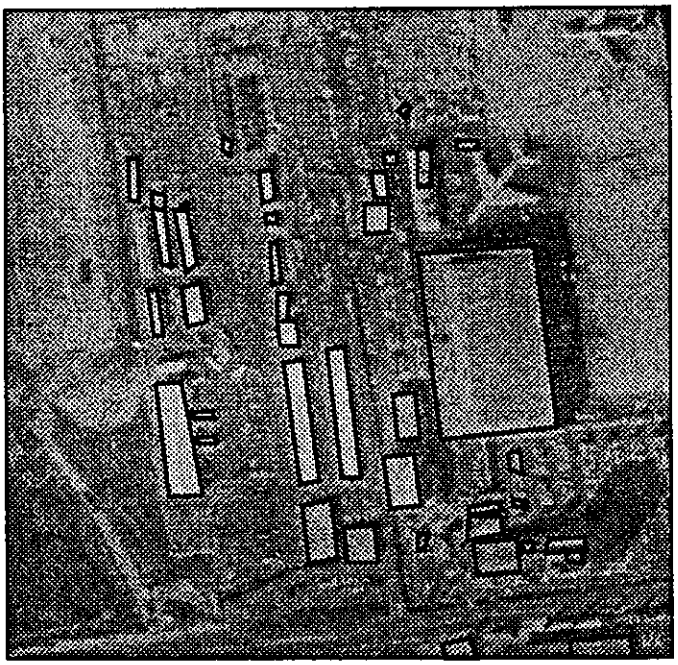

Figure 3-2: BABE building hypotheses for LAX [right image]

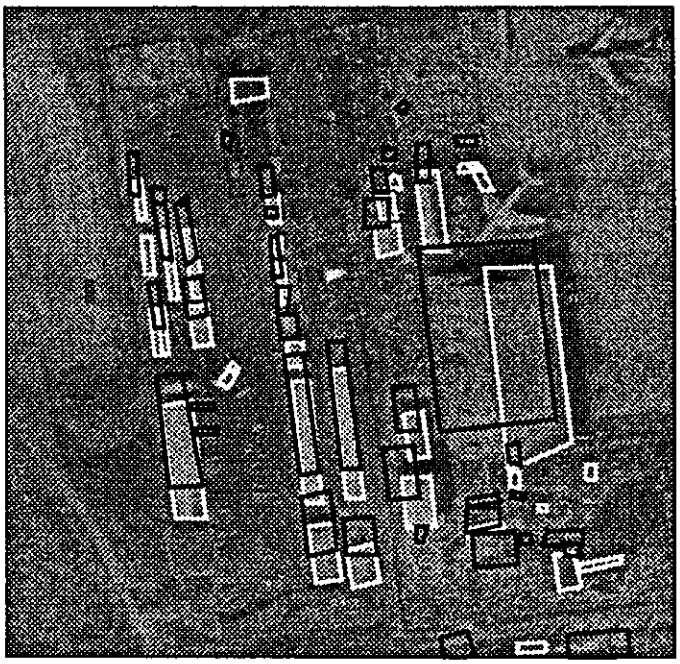

Figure 3-3: BABE results superimposed on the left image using coarse registration

Figure 3-4: BABE results superimposed on the left image using fine registration

Figure 3-4 shows the superimposition of BABE results on the LAX left image using the fine registration technique described in Section 2.4 while Figure 3-3 shows the accuracy of the coarse registration described in Section 2.3. This superimposition allows us to see directly the structure correspondence as well as the differences between the two monocular analysis results in Figures 3-1 and 3-2. The horizontal displacement between the hypotheses for the fine registration can be related to the relative height of the structures because of the epipolar constraint. Coincident structure hypotheses give very strong support for hypotheses of buildings. This is due to the fact that the feature extraction process rarely fails in exactly the same way in each of the images.

To automate matching between the hypotheses in the left and right images we utilize geometric constraints. We take each BABE hypothesis from the left image and find the best corresponding hypothesis on the right image. To evaluate the effects of the registration, we performed this matching on both the coarse registration using the CONCEPTMAP scene model, and the fine registration described above. Figure 3-5 shows the matching results using the CONCEPTMAP registration while Figure 3-6 shows the matching results of the fine scene registration. In both cases we have chosen a small area from the left-center of the LAX scene to illustrate the details of matching. The light (dotted) structures are hypotheses from the left image while the dark (solid) structures are from the right image.

The matching process is a global search between two sets of boxes according to local limitations in the search area. The epipolar geometry of the fine registered images can be used to constrain the search area to a range of scanlines in the images. Then we use a simple criteria to select potential matches: the position of the hypothesis center-of-mass projected into a rectangular search space and amount of overlap between the pairwise structures.

This simple matching process allows us to consider arbitrarily complex polygonal structures because we are not performing discrete vertex or structure matching to establish a stereo correspondence such as in Mohan and Nevatia [23]. This is required given the relatively complex imagery and imprecise segmentation delineation provided by BABE. In many cases detailed high-level structure matching (as in [23]) will be defeated by errors in monocular feature detection due to occlusion, texture, and accidental alignments of objects and background. These are precisely the errors that cause area-based and feature-based matching to fail, although propagated to a high-level matching process.

Because our matching criteria are not very selective, we must disambiguate among many plausible matches. However, even if we devised a more specific set of match criteria, it is unclear whether we could account for situations in built-up urban areas where the buildings are very close, and have very similar shapes, orientations, and heights simply by using a set of local optimal matches. There are several examples of the alignment of similar buildings even within the LAX imagery. Thus, there is a virtue in the application of weak constraints because they do not require detailed high-level knowledge about the mis-registration. Instead of trying to disambiguate the matches locally we use global considerations based on the plausibility of the matched sets of structures. We define four different situations that occur depending on whether several structure hypotheses share the same correspondence with a hypothesis structure in the other image.

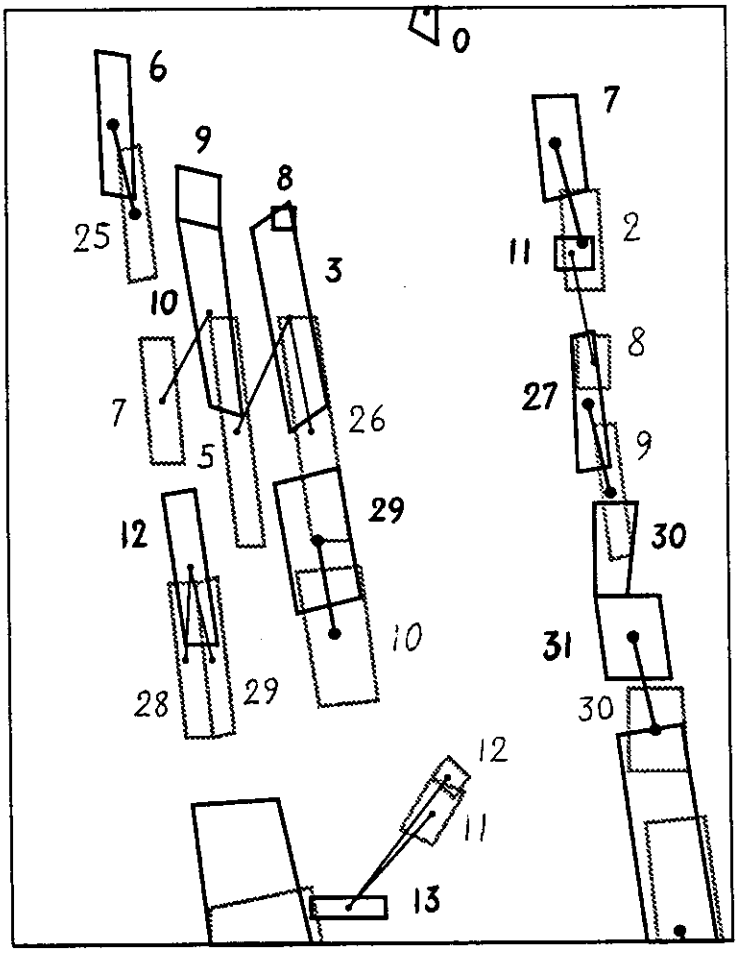

Figure 3-5: Matching of buildings using a coarse registration

Results of the matching of boxes for the coarse registration					
Type of corres.	Left box	Right box ———	Corres. score	rel. height estimate	rel. line offset
1	2	7	0.92	4.0	18.2
	8	11	0.73	3.5	21.0
	9	27	0.85	4.4	16.4
	10	29	0.94	3.7	18.3
	25	6	0.83	3.8	17.4
	30	31	0.93	4.2	18.2
2	7	10	0.73	-9.3	16.9
	28	12	0.89	-0.1	17.5
	29	12	0.80	4.4	17.1
3	5	3	0.75	-10.0	22.0
	26	3	0.87	4.5	21.5
4	11	13	-44.5	19.6	-24.9
	12	13	-34.5	16.0	-18.5

Table 3-1: Matching Results For LAX Building Hypotheses with coarse registration

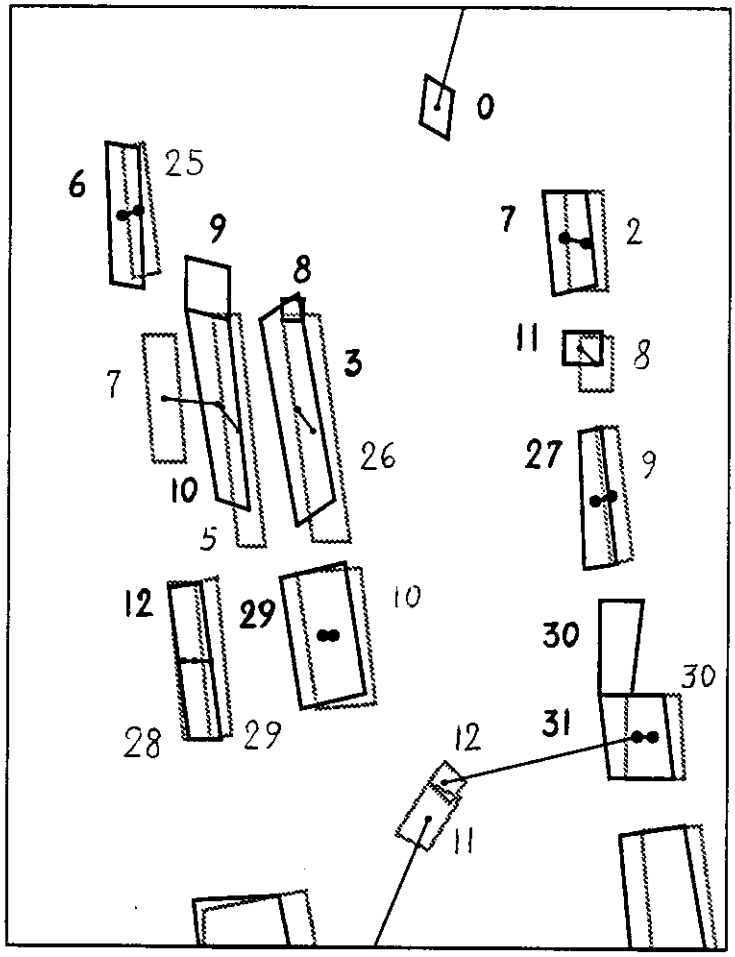

Figure 3-6: Matching of buildings using a fine registration

Results of the matching of boxes for the fine registration					
Type of corres.	Left box	Right box _____	Corres. score	rel. height estimate	rel. line offset
1	2	7	0.92	3.1	0.2
	8	11	0.73	2.5	3.0
	9	27	0.86	3.4	-1.6
	10	29	0.94	2.7	0.3
	25	6	0.83	2.8	-0.6
	30	31	0.93	3.2	0.2
2	5	10	0.75	3.7	4.9
	28	12	0.89	-1.1	-0.5
	29	12	0.81	3.4	-0.9
3	26	3	0.87	3.5	3.5
4	7	10	-11.4	-10.3	-1.1
	11	31	-44.6	-35.5	-1.6
	12	13	-51.5	15.0	-36.5

Table 3-2: Matching Results For LAX Building Hypotheses with fine registration

Tables 3-1 and 3-2 show the results of the monocular matching using the coarse and fine scene registration. In this experiment we search for the best match for each of the building hypotheses produced by BABE in the left image. A similar analysis could be performed on the structures generated from the right image. Four different situations can occur during matching that correspond to the application of local and global properties:

- **Type 1** This case corresponds to the "ideal" situation where we have a unique correspondence between a single hypothesis in the left and right image. The score of the correspondence gives us an estimate of the quality of the match between the two structures. A match score greater than 0.9 indicates that the two structures are quite similar. Pair (10,29) is an example of such a good match. A lower match score, as in the case of Pair (8,11), indicates that while there is a correspondence between structures their BABE delineation is not completely consistent between the left and right image.
- **Type 2** This case occurs when a structure in the left image shares a right structure correspondence with other structures in the left image. This correspondence is therefore ambiguous. In this case the match score is not sufficiently different to disambiguate between the multiple choices. However, knowledge of a reasonable height range for these structures could be used to select the correct correspondence. For example, Pair (28,12) and Pair (29,12) have a significant difference of 4.5 meters in their height estimate. However, neither height is sufficiently unusual to prefer one interpretation over another without some external information. However, in the case Pair (7,10) for the coarse registration, this match could be discounted due to a height interpretation that is below the local terrain.
- **Type 3** This case occurs when several matches were possible with different structures in the right image, for example Pair (26,3). The correspondence selected has the highest confidence match but other correspondences are possible, i.e., Pair (26,10). Once again, knowledge about the reasonable height range for structures could be used to select the appropriate correspondence.
- **Type 4** This case occurs when structures in the left image do not have any reasonable correspondence in the right image such as in Pair (11,13). Nevertheless the best correspondence is given. The utility of such a match is to provide another analysis process with a context to search for new structures in the right image or to eliminate this structure from further analysis.

As we have seen, the ability to perform scene registration allows for the efficient correlation of monocular analysis. There appears to be no major difference between the matching results using the coarse or fine scene registration. In the case of good matches, i.e., Type 1, having a high score, the results are identical for the coarse and the fine registration. Even if, according to the registration, different types of matching are possible for the ambiguous matches, this technique still seems to be quite robust. However, different scene clues can be derived by the analysis of the absolute match score, the match type, the estimated height, and the relative row offset for the different registrations. For example, in Type 2, matching a choice between two competing interpretations must be made. A verification process could be invoked to locate a better matching and delineation of the structures. The estimate of structure height in the fine registration can be compared to the results of specific stereo matching algorithms. We can also characterize the quality of the scene registration (epipolar geometry in particular), by the relative row offset results for the matching of a given set of accurate structure delineations. For example, given a reference manual segmentation of buildings in the left and right images we can use the local translations in rows and columns between the matched structures in the coarse registration (i.e. the estimated height and the relative row offset) as estimates of the

misregistration. Actually a simple analysis of the translations for the good matches in the coarse registration is used to set a spatial constraint to reduce the search window while refining the matching process. The relative row offset in the coarse registration is approximately 18 pixels. This can be derived by averaging the row offset of the good matches and can be verified by subtracting the relative row offset of the fine registration from that of the coarse registration.

Thus, the matching of high-level structures extracted from monocular analysis can be used for evaluation of the quality of the structure delineation, height estimation, or provide a refined estimate of the scene registration. In the following section we discuss in more detail the use of structure matching to refine the registration.

3.1.3. Structure Matching for Registration

As we have seen, starting from the coarse registration, we are able to perform structure matching that provides an estimate of the local offsets in rows and columns for the center-of-mass of the structures generated by BABE in the left and right images. We can consider these corresponding points as control points selected automatically and then perform a registration of the stereopair exactly as with the shadow corners.

This observation can be generalized, and we can consider the automated registration more globally. As we have seen, one requirement for automated registration is the automatic selection of corresponding points in the stereo pair images. Classically these points are physical features of the images like shadow corners, road intersections or specific marks, but we can also think of "virtual geometric points" defined by a geometrical relationship to real features in the images. The center-of-mass of the structures generated by BABE belong to this category.

With such a definition of control points, the automated registration can be performed using many different features such as isolated points, edges, boundary contour, regions, and structures. The problem is, as before, to perform accurate matching in order to end up with control points whose position is accurately known. Our ability to match various features is the characteristic that tends to limit our choice of techniques.

The traditional "characteristic points" approach is often accurate in point position, but because it is highly local in nature it is very difficult to find an accurate one-to-one correspondence. The more inherent complexity in the feature selected as a match point, the more constrained the matching becomes. For example, single intensity points in the images are very accurate in position but the matching process to find the correspondences is very difficult. On the other hand, very complex structures can be very easily disambiguated to find good correspondences. The problem is that as we match high-level structures our ability to determine match point positions becomes less accurate. Because complex features such as road intersections, building outlines, unique terrain points require a detailed analysis, their detection and delineation may be inexact. From these inexact corresponding structures we must generate "virtual" control points whose position in the feature is less sensitive to these errors.

The Figure 3-7 shows the "virtual" control points we selected in the images using the matching of structures generated by BABE. The structures generated by BABE in the left image are shown in white, the right image structures are shown in black. The automatic control points selected correspond to the good structure matches. Given that BABE does a fairly good job in structure delineation and generates consistent hypotheses in both images, the set of control points considered is quite reliable. In fact, the

residuals of the registration shown in Table 2-2 are comparable to the registrations using a manual set of control points.

Figure 3-8 represents the BABE results superimposed on the left image using the isometric registration derived using these control points. The results are nearly identical to those derived using manual registration and show a good superimposition of the building structures.

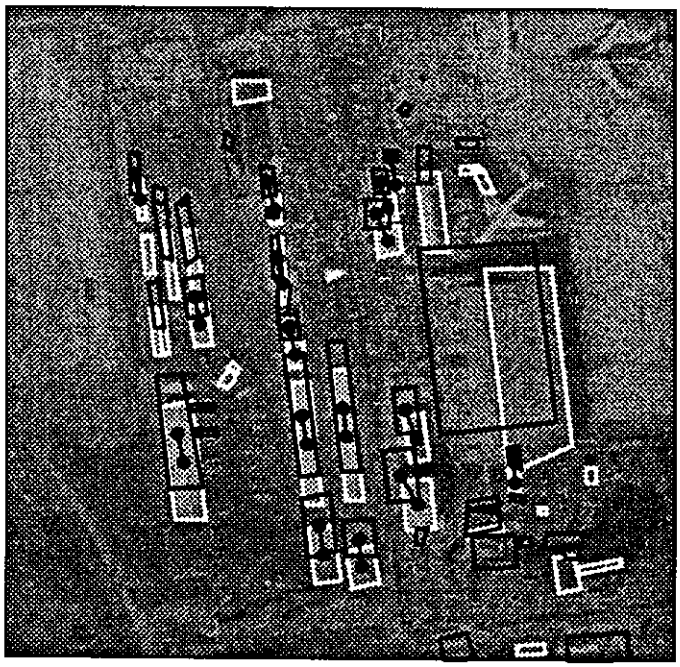

Figure 3-7: Structure matching using coarse scene registration

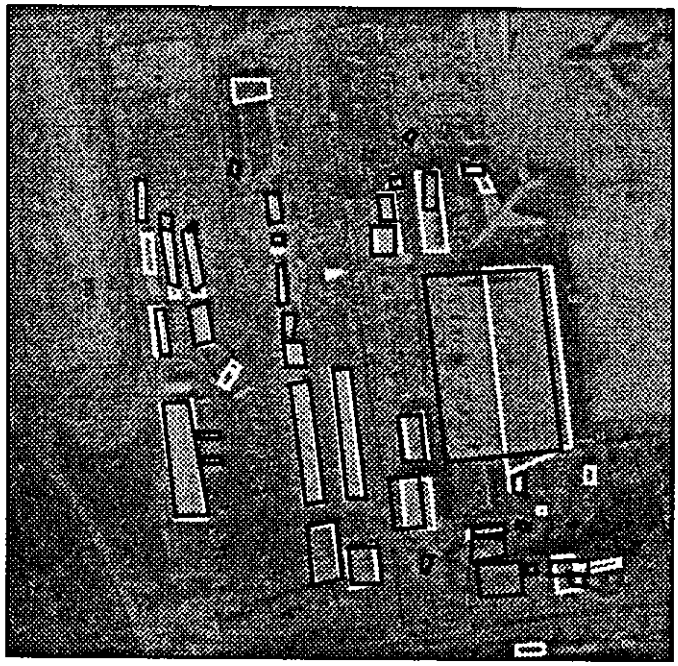

Figure 3-8: Superposition of structures using structure matching registration

3.2. Registration for stereo matching

The central issue in computational stereo is the solution of the correspondence between features visible in two overlapping images. The correspondence of a point feature visible in the left and right image of a stereo pair can be used to generate the three-dimensional description of that point in the scene. Points need not be the only feature matched. As we have seen, the result of matching structures generated by monocular analysis can be considered as stereo matching and yields a relative height estimate.

In Section 2.2 we discussed how the epipolar constraint is used to simplify stereo matching by reducing it to a one-dimensional problem. This is because the epipolar lines in the imagery are registered to be corresponding scanlines in the left and right image. The assumption that the scene registration is ideal and that the epipolar constraint is totally satisfied is rarely warranted in imagery digitized from aerial photography. This is due to all of the orientation problems previously discussed as well as inaccuracies inherent to the digitization process primarily due to rotation of the image. In the following section we discuss the effect of coarse and fine scene registration on two stereo matching algorithms.

3.2.1. Two stereo correspondence algorithms

Algorithms for stereo correspondence can be grouped into two major categories: area-based and feature-based matching. Examples of area-based matching include correlation techniques for matching image intensity patches using various evaluation functions including normalized cross-correlation, mean-square-difference, or surface fitting residual error. Feature-based techniques match image features derived from edge, line, or boundary detection. Area-based techniques provide a dense disparity map with an estimate generated at every point in the image. Feature-based approaches provide depth information only at points where the features are generated, often points of intensity discontinuity that may correspond to discontinuities in depth.

Both classes of techniques, area-based and feature-based, have advantages and drawbacks that primarily depend on the task domain and the three-dimensional accuracy required. For complex urban scenes, feature-based techniques appear to provide more accurate information in terms of locating depth discontinuities and in estimating height. However, area-based approaches tend to be more robust in scenes containing a mix of buildings and open terrain. For this reason we have developed two stereo matching algorithms. S1 is an area-based algorithm and uses the method of differences matching technique developed by Lucas [15, 20]. S2 is feature-based using a scanline matching method that treats each epipolar scanline as an intensity waveform. The technique matches peaks and troughs in the left and right waveform. Both are hierarchical and use a coarse-to-fine matching approach. Each is quite general as the only constraint imposed is the order constraint for the feature-based approach. The order constraint should generally be satisfied in our aerial imagery except in cases of hollowed structures.

Both matching algorithms assume the epipolar geometry but have different sensitivity to its accuracy. The S1 area-based approach uses a hierarchical set of reduced resolution images to perform a coarse-to-fine matching of small windows in the two images. At each level the size of the windows for the matching process depends on the resolution of the reduced image. An initial disparity map is generated at the first level. Subsequent matching results computed at successively finer levels of detail are used to refine the disparity estimate at each level. Therefore the amount of error in the scene registration that can be tolerated by this matching algorithm depends on the size of the matching windows. However, since there is a relationship between the matching window size and the level of accuracy, simply using larger matching windows may not be desirable.

The S2 feature-based approach matches epipolar lines in the left and right image. It uses a hierarchical approximation of the intensity waveforms to match peaks and valleys at different levels of resolution. To avoid mismatches it uses inter-scanline consistency that enforces a linear ordering of matches without order reversals. It also applies an intra-scanline consistency that considers the matches in adjacent scanlines. Application of intra-scanline constraint is used to increase the confidence of matches found to be consistent across multiple scanlines and to delete improbable matches.

Figure 3-9 is a complex industrial area scene and was the focus of our discussion on coarse and fine scene registration in Sections 2.3 and 2.4. This scene contains many of the difficulties found in stereo matching, including occlusion, complicated textures, large depth discontinuities, and complicated three-dimensional objects. The Figure 3-10 shows the results of the matching for the CONCEPTMAP coarse registration using the area-based algorithm S1. In all of the disparity match results presented in this paper, brighter regions are closer to the camera and have greater height. Darker regions are at or below the relative terrain ground plane. The results using the coarse registration are quite errorful. We can barely discern the general shape of the taller buildings in the middle and the upper left areas of the scene. The S2 algorithm is completely unable to use a coarse registration since the scanline matching

assumes that the epipolar constraint is satisfied.

Figure 3-9: Intensity image of the scene

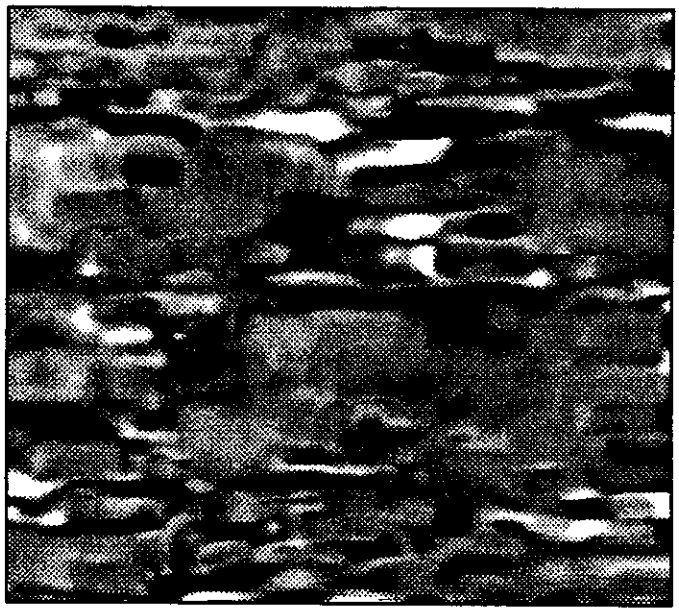

Figure 3-10: S1 disparity map using coarse registration

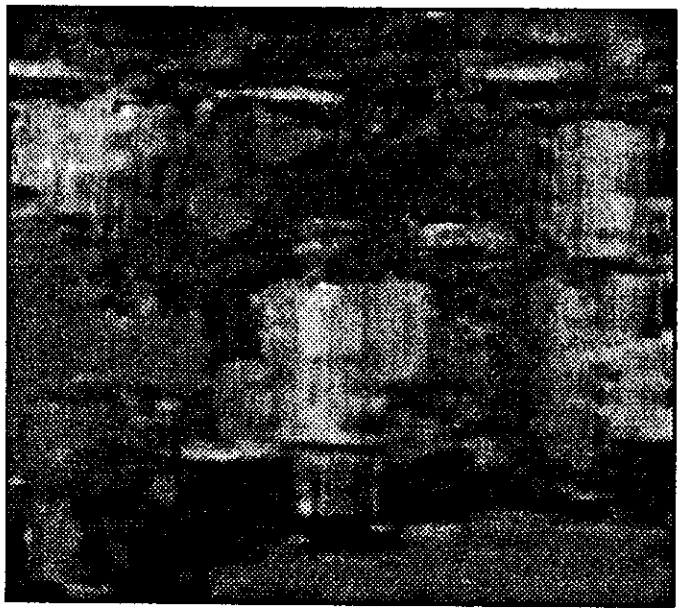

Figure 3-11: S1 disparity map using fine registration

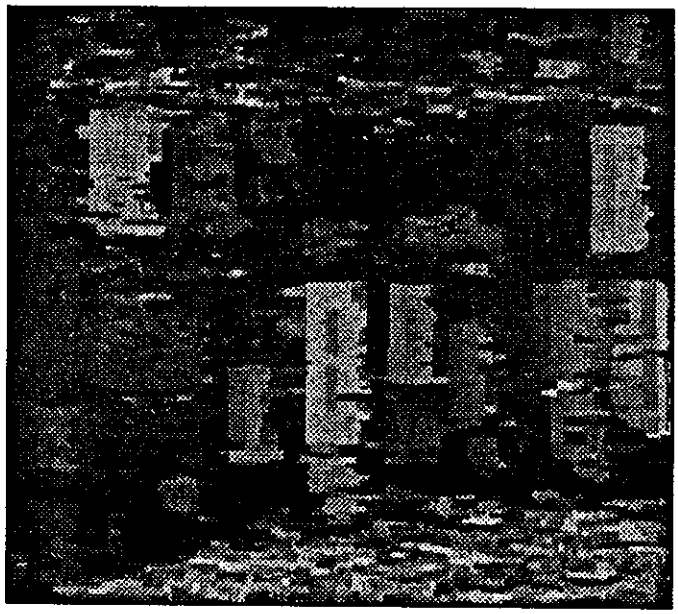

Figure 3-12: S2 disparity map using fine registration

Figure 3-11 shows the results of the matching using the S1 algorithm with the fine registration produced by the manual selection of shadow corner points. The matching results are significantly better with the bright areas again representing the highest regions and corresponding to most of the buildings in the scene. Although the delineation is not crisp there are few major mismatches, and we have an adequate impression of the range of heights in the scene. The S1 algorithm has many of the advantages

and drawbacks of any area-based technique. As we can see in Figure 3-11 most of the errors are due to abrupt changes in height due to man-made structures.

Figure 3-12 shows the results of the matching using the S2 algorithm with the same fine scene registration in Figure 3-11. This technique performs very well on the disparity discontinuities caused by man-made structures, and therefore we have a much better delineation of the buildings than in Figure 3-11. Nevertheless, despite post processing of the disparity results, the resulting disparity image is noisy. As expected, the S2 algorithm can not provide robust matches in areas of uniform intensity or in highly textured areas.

The results of the two stereo matching algorithms are quite complimentary, and we believe that it is possible to take advantage of the different failure modalities in order to form a composite disparity map that gives a more accurate three-dimensional representation of the scene. It is also clear that stereo matching relies on a more accurate scene registration than is provided by the coarse registration described in Section 2.3. Even when the matching window size for an area-based stereo algorithm is larger than the inherent mis-registration, it may be difficult for the matching algorithm to recover from mis-matches due to poor scene registration. This is in contrast to the results in structure matching presented in Section 3.1 that appear to be much less sensitive to a coarse scene registration.

3.3. Registration for visualization

A practical means for three-dimensional visualization is needed in order to understand the quality of the stereo results. This can be achieved by the construction of a left and right stereo image from the original unregistered overlapping imagery. The goal is to reconstruct the scene as the human eyes will naturally interpret it: the left eye sees the left image of the scene and the right eye simultaneously sees the right image. Thus, we need a stereo pair already registered and satisfying the epipolar constraint. To see the relief we look at the scene with our vision baseline parallel to the scanlines (epipolar lines), and we accommodate at the infinite.

There are two common techniques to present stereo images to a user. The simplest is anaglyph stereo where the left/right images are presented in blue/red channels of the display and the user wears glasses with blue/red filters. This is the technique commonly used for viewing 3-D movies in commercial movie theatres. A superior but more costly approach is to use polarized glasses to separate the left and right images. Several manufactures use a liquid crystal shutter to display the two stereo fields alternating at 1/60th or 1/120th of a second. The detail and quality of the polarized stereo is much crisper and easier for viewers to fuse than using the anaglyph technique. For our experiments, we have used the anaglyph method, however the construction of a synthetic stereo image is identical regardless of viewing method.

3.3.1. Construction of the right image from the left image

Given a left and right image registered into the epipolar geometry, as are the intensity images after fine registration, we can display them using different stereo viewing methods as described above. This is the most common way to achieve stereo visualization. However, a more interesting approach is to use this representation to generate a stereo pair (left and right image) from the three-dimensional information we have computed by stereo matching in order to directly visualize the matching results as a three-dimensional scene. We call the generation of a stereo pair from three-dimensional information derived from stereo analysis *synthetic stereo reconstruction*. Synthetic reconstruction can be used to visualize and compare the results of stereo matching by direct visualization.

Figure 3-13: S2 stereo matching result [left]

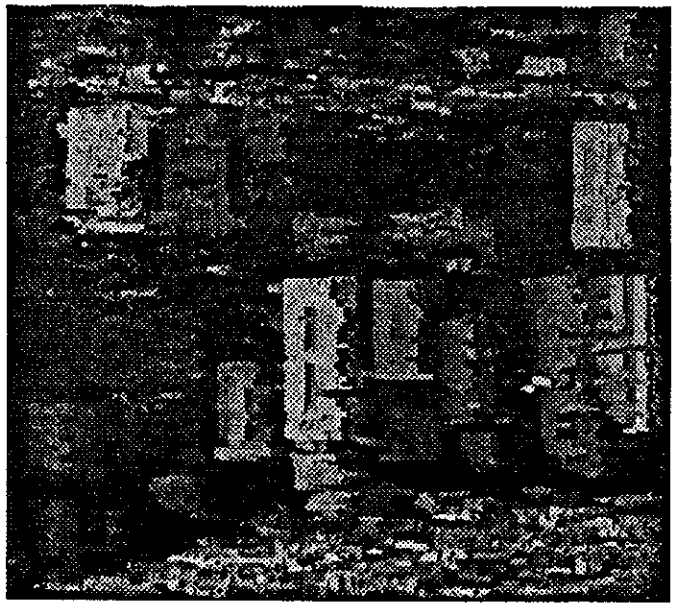

Figure 3-14: S2 stereo matching result [right]

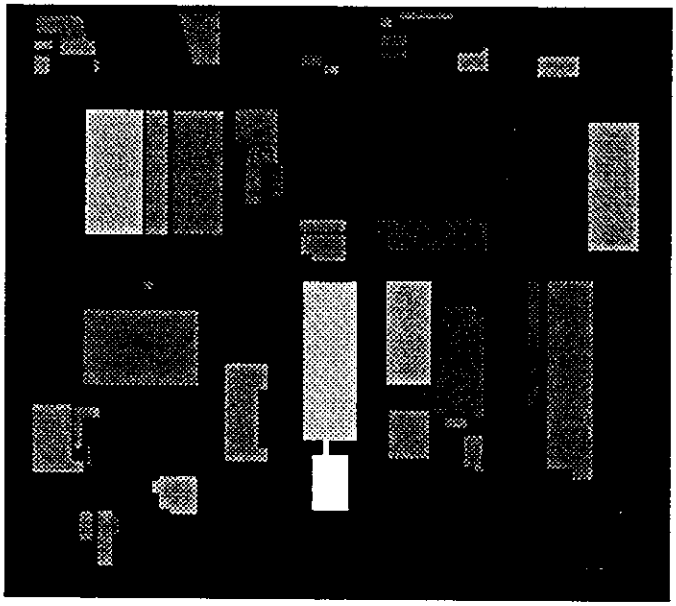

Figure 3-15: Ground truth left image for DC38008 scene

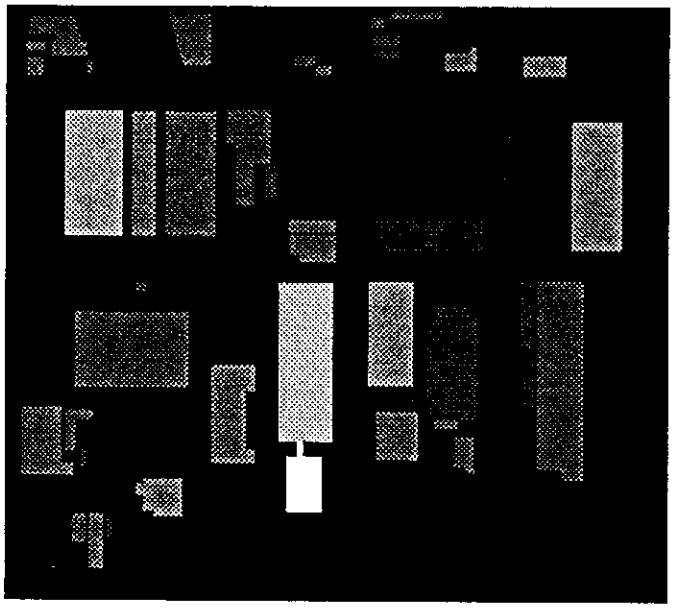

Figure 3-16: Ground truth right image for DC38008 scene

A relative height computation, or disparity, is the result of most stereo matching algorithms. The disparity is usually encoded in a representation related to the geometry of the left image. In this disparity map, the values of each point in the map correspond to the relative height of that point in the left image. In order to generate a synthetic reconstruction containing the information extracted by the matching process we must generate a new right image. Each point in the right image is the corresponding point in the left image displaced by the relative height estimate in the disparity map. This process is exactly the opposite of that used to generate the disparity map. However, we do not have to

solve the matching problem. The computed right image is, by definition, perfectly registered since there are only local horizontal shifts between the left and the right image. Thus, we satisfy the epipolar constraint. Figure 3-14 and Figure 3-13 shows reconstructed synthetic stereo images for the stereo matching results produced by S2 in Figure 3-12.

3.3.2. 3D segmentation for ground truth determination

The visualization of scenes and stereo matching results is a powerful tool for the qualitative comparison of different scene interpretation techniques. One technique is the side-by-side comparison of the original stereo scene and the automated reconstruction. Such a comparison allows us to quickly see those buildings that are missing or have errors in height or ground position.

However, a quantitative approach is also possible and is potentially more useful. Using the anaglyph display techniques we can generate a three-dimensional segmentation that allows us to store the structure of each building in the scene. The form of the data is simply a segmentation description file containing collections of left image points and their relative height. From that representation we can infer a partial three-dimensional representation of the buildings guessing the shape of the invisible parts, much as is done with simple wireframe models. We can also use this representation as a baseline reference representation for buildings in order to compare and contrast the various processing results.

Figures 3-15 and 3-16 show how this technique can be used to construct a simple three-dimensional ground truth segmentation that can be visualized as a stereo scene. We simply represent the building roofs as horizontal surfaces and displace these surfaces proportionally to the actual height of the buildings that we got from the three-dimensional segmentation of the roofs. Such a stereo pair is very useful in order to compare the results of automatic scene matching.

4. Automatic scene registration

As we have seen, structure matching, stereo matching, and visualization all rely on the quality of the stereo registration. The registration and the matching process are therefore interdependent. Structure matching appears to be a task that we can reliably perform even with a coarse scene registration. Further, the results of structure matching provides a method to automatically refine the initial coarse scene registration. In this Section we demonstrate a complete end-to-end scenario of automatic structure matching, fine registration, and stereo analysis. Thus, we can automatically generate a three-dimensional representation of the scene starting from the CONCEPTMAP image database.

We began with the DC38008 test area corresponding to Figures 2-3 and 2-4, previously shown in Section 2.3. We then utilize the BABE structure results and perform structure matching to select reliable control points. The structures generated by BABE are fragmented and are not as consistent as those generated for the LAX stereo pair. Nevertheless, we are able to find a number of good matches, well distributed across the image, as shown in Figure 4-1. Subjectively, the registration quality is good, as seen in Figure 4-2, where many of the building fragments are now aligned. The overall registration quality is detailed in Table 2-1 (ISO structure) in Section 2.4. While it is not as accurate the registration derived by manual ground control selection, it is clearly comparable.

Finally using this automatically registered stereo pair we performed stereo matching to get a dense disparity map of the scene. Figures 4-3 and 4-4 show the results for the S1 and the S2 matchers. The results are comparable to those in Figures 3-11 and 3-12 achieved using manual selection of control points. Thus we have shown the feasibility of end-to-end processing to establish precise local

registration using automatic ground control point estimation.

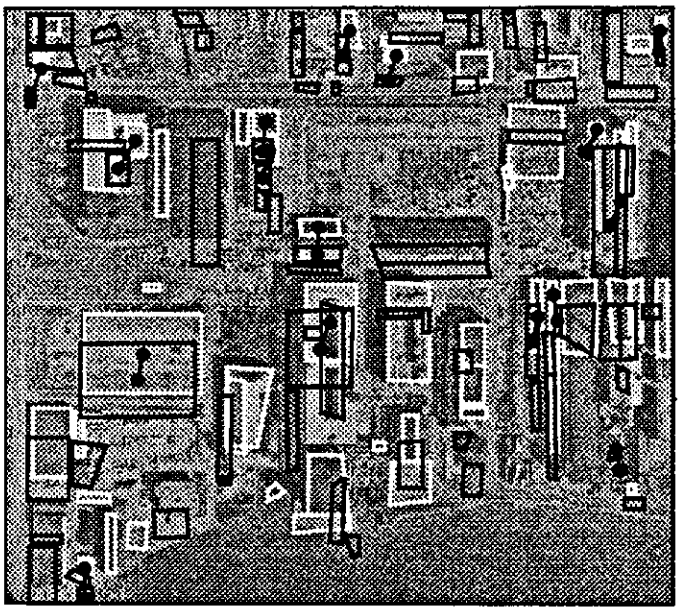

Figure 4-1: Automatic control points using structure matching

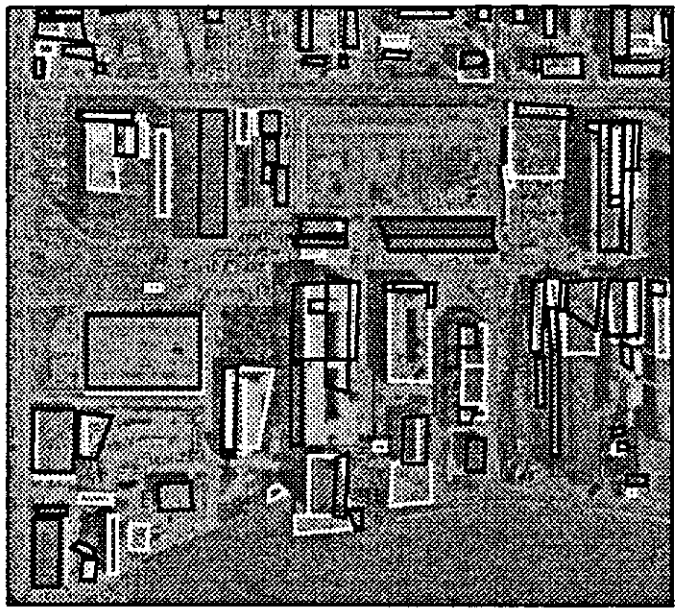

Figure 4-2: Superposition of structures using structure matching registration

Figure 4-3: S1 disparity map using structure matching registration

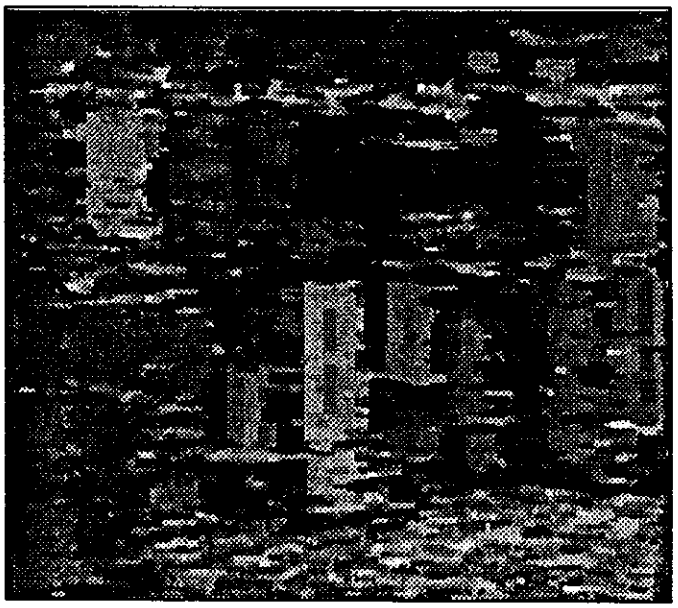

Figure 4-4: S2 disparity map using structure matching registration

5. Conclusions

The importance of scene registration in the automated interpretation of aerial imagery can not be overstated. Scene registration is required for monocular matching, stereo analysis, scene visualization, accurate mensuration, and for many other photo-interpretation tasks. Most work in computational stereo

has ignored the problem of scene registration assuming that the left/right image pairs were already in epipolar geometry. As we have seen, this may limit the utility of many feature-based and some area-based matching techniques, especially in cases where there are significant residual errors in the registration process.

Traditionally we have separated the stereo analysis of digital images in two problems, registration and matching, and have attempted to solve each independently. However, the results of matching, whether structural or using a stereo model, are actually the ultimate form of scene registration since the matching solves the correspondence between different objects in the images. In some sense registration and matching are corresponding processes that are performed at different representational levels. Registration relies on mathematically modeling the image acquisition in the three-dimensional world, while matching defines the relationship between corresponding points in two images. Registration is generally necessary to constrain search during matching, but at least a sparse matching is necessary to perform the registration.

5.1. Future Work

There are several areas for future work focused on improving techniques for scene registration. First, using the image-to-map correspondence, we need to improve the accuracy of the landmarks stored in the CONCEPTMAP database. It may be necessary to include accurately surveyed geodetic control points in addition to those that are acquired from map sheets. However, given the high resolution of the imagery that we are working with, it may actually be quite difficult, using these control points, to improve on the current level of accuracy. More complex known landmarks such as road networks may be utilized to accurately register imagery to maps for automated scene registration [16].

For direct image-to-image correspondence we have seen some limitations in automatic extraction of shadow corner points in complex urban imagery for registration. Additional sources of reliable registration points should be available using monocular extraction of man-made structures such as the road networks. From our previous work in road detection [3] and tracking [19] it seems quite reasonable to use these structures as potential landmarks for scene matching. Furthermore the model of transformation between the two images should be enhanced to get a better registration once we have accurate control points.

Finally, we need to pursue the experiment of iterative refinement of registration via coarse-to-fine matching. The basic idea is to perform a coarse registration using road networks, or known ephemeris data. This coarse registration would be followed by feature matching using structure matching or shadow corners to generate an estimate of the error in the registration. This estimate can be refined using a hierarchical approach using high-level features that are easily matched but have inaccuracies in position, to low level features that are difficult to match but have unambiguous positions in the stereo pair. The process of registration and matching would then iteratively converge to a complete matching of the scene. Whether this approach can achieve high registration accuracy equal to manual correspondence is a topic for further research.

6. Acknowledgements

We thank our colleagues in the Digital Mapping Laboratory for an interesting and congenial working environment. We particularly acknowledge the contribution of Yuan Hsieh and Aviad Zlotnick to the development of several of the examples used in this paper.

7. References

[1] R.D. Arnold.
Local Context in Matching Edges for Stereo vision.
In *Proceedings: DARPA Image Understanding Workshop*, pages 777-791. May, 1978.

[2] Aviad, Z.
Locating Corners in Noisy Curves by Delineating Imperfect Sequences.
Technical Report CMU-CS-88-199, Carnegie-Mellon University, December, 1988.

[3] Aviad, Z. and P. D. Carnine.
Road Finding for Road Network Extraction.
In *Proceedings: Computer Vision and Pattern Recognition*, pages 814-819. Ann Arbor, Michigan, June, 1988.
Extended version available as Tech. Report CMU-CS-88-157.

[4] Aviad, Z., McKeown, D. M., Hsieh, Y.
The Generation of Building Hypotheses From Monocular Views.
Technical Report, Carnegie-Mellon University, January, 1989.
to appear.

[5] S. T. Barnard.
Stochastic stereo matching over scale.
In *Proceedings: DARPA Image Understanding Workshop*, pages 769-778. April, 1988.

[6] S.T. Barnard and M. A. Fischler.
Computational stereo from an IU perspective.
In *Proceedings: DARPA Image Understanding Workshop*, pages 157-167. April, 1981.

[7] H. S. Lim and T. O. Binford.
Structural correspondence in stereo vision.
In *Proceedings: DARPA Image Understanding Workshop*, pages 794-808. April, 1988.

[8] R. A. Brooks, A. M. Flynn, and T. Marill.
Self calibration of motion and stereo vision for mobile robot navigation.
In *Proceedings: DARPA Image Understanding Workshop*, pages 398-410. April, 1988.

[9] L-H Chen, and T. E. Boult.
An integrated approach to stereo matching, surface reconstruction and depth segmentation using consistent smoothness assumptions.
In *Proceedings: DARPA Image Understanding Workshop*, pages 166-176. April, 1988.

[10] Faugeras, O. D., and Toscani, G.
The Calibration Problem for Stereo.
In *Proceedings of Computer Vision and Pattern Recognition*, pages 15-20. June, 1986.

[11] M. J. Hannah.
SRI's Baseline Stereo System.
In *Proceedings: DARPA Image Understanding Workshop*. December, 1985.

[12] B. K.P. Horn.
Relative Orientation.
In *Proceedings: DARPA Image Understanding Workshop*, pages 826-837. April, 1988.

[13] Huertas, A. and Nevatia, R.
Detecting Buildings in Aerial Images.
Computer Vision, Graphics, and Image Processing 41:131-152, April, 1988.

[14] Irvin, R. B., McKeown, D.M.
Methods for Exploiting the Relationship between Buildings and their Shadow in Aerial Imagery.
In *SPIE Proceedings Image Understanding and the Man-Machine Interface II.* January, 1989.
Also available as CMU Computer Science Technical Report CMU-CS-88-200.

[15] B. D. Lucas.
Generalized Image Matching By The Method of Differences.
PhD thesis, Carnegie Mellon University, July, 1984.

[16] Chris McGlone.
Automated image-map registration using active contour models and photogrammetric techniques.
In *SPIE Proceedings on Reconnaissance, Astronomy, Remote Sensing, and Photogrammetry.* January, 1989.

[17] McKeown, D.M.,
Digital Cartography and Photo Interpretation from a Database Viewpoint.
In Gargarin, G. and Golembe, E. (editor), *New Applications of Databases*, pages 19-42. Academic Press, New York, N. Y., 1984.

[18] McKeown, D.M.
The Role of Artificial Intelligence in the Integration of Remotely Sensed Data with Geographic Information Systems.
IEEE Transactions on Geoscience and Remote Sensing GE-25(3):330-348, May, 1987.
Also available as Technical Report CMU-CS-86-174.

[19] McKeown, D.M. and Denlinger, J. L.
Cooperative Methods for Road Tracking in Aerial Imagery.
In *Proceedings IEEE Computer Vision and Pattern Recognition Conference*, pages 662-672. June, 1988.

[20] McKeown, D.M., McVay, C.A., and Lucas, B. D.
Stereo Verification In Aerial Image Analysis.
Optical Engineering 25(3):333-346, March, 1986.
Also available as Technical Report CMU-CS-85-139.

[21] H. P. Moravec.
Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover.
PhD thesis, Stanford University, September, 1980.

[22] N. M. Nasrabadi, Y. Liu, and J-L. Chiang.
Stereo vision correspondence using a multi-channel graph matching technique.
In *IEEE international Conference on Robotics and Automation.* April, 1988.

[23] R. Mohan and R. Nevatia.
Perceptual grouping for the detection and description of structures in aerial images.
In *Proceedings: DARPA Image Understanding Workshop*, pages 512-526. April, 1988.

[25] *The Manual of Photogrammetry*
Fourth Edition edition, American Society of Photogrammetry, Falls Church, VA., 1980.

[26] Price, K. E.
Change detection and Analysis in Multi-spectral Images.
PhD thesis, Carnegie-Mellon University, December, 1976.

[27] Price, K. and Reddy, D. R.
Matching Segments of Images.
IEEE Transactions on Pattern Analysis and Machine Intelligence 1(1):110-116, January, 1979.

[28] K. E. Price.
Relaxation Matching Techniques_A Comparison.
IEEE Transactions PAMI-7(5):617-623, September, 1985.

[29] D. Weinshall.
Qualitative vs. quantitative depth and shape from stereo.
In *Proceedings: DARPA Image Understanding Workshop*, pages 779-785. April, 1988.