

Progress in 3-D Mapping and Localization

Martial Hebert, Matthew Deans, Daniel Huber, Bart Nabbe, Nicolas Vandapel

The Robotics Institute
Carnegie Mellon University
Pittsburgh PA 15213

Abstract

This paper is a summary of results obtained in the past few years in the area of 3-D mapping and robot localization. The emphasis of this work is the reconstruction of three-dimensional representations of the environment from sensor information, assuming inaccurate or absent robot pose information, and assuming general 3-D configurations, e.g., not limited to 2-1/2D elevation maps. The approaches are divided into two broad classes: Surface matching, in which large pieces of 3-D surfaces are matched across observations in order to recover the transformations between observations, and feature matching in which individual features extracted from the input data are matched. Surface matching is most applicable to robots equipped with range sensors such as stereo or ladar, while feature matching is most applicable to video-based systems. We report on experiments and applications in the areas of terrain map building, modeling of interior environments, modeling of individual objects from many views, cooperative stereovision using teams of robots, and simultaneous recovery of structure and motion from bearing-only sensors.

1. Introduction

Constructing accurate 3-D maps of the environment from sensor data is a critical component of mobile robot systems. In addition to providing the necessary information for planning and visualization, 3-D mapping is also a component of localization systems. 3-D mapping is a challenging operation in most realistic applications. In outdoor environments, the volume of data and the large spatial extent of terrain maps makes mapping a computationally challenging operation. At the other extreme, detailed mapping of individual objects may require the registration of a large number of sensor views with a high degree of accuracy. When 3-D information is not directly available from sensors it must be inferred as part of the mapping process.

Broadly speaking the mapping problem can be viewed as a registration problem in which the problem is to find the optimal transformations between multiple sensor views. The registration approaches can be divided into two broad classes: Surface matching in which large pieces of 3-D surfaces are matched across observations in order to recover the transformations between observations, and feature matching in which individual features extracted from the input data are matched. Surface matching is most applicable to robots equipped with range sensors such as stereo or ladar, while feature matching is most applicable to video-based systems.

To illustrate some of the approaches to registration, this paper is a summary of results obtained in the past few years in the area of 3-D mapping and robot localization. The emphasis of this work is the reconstruction of three-dimensional representation of the environment from sensor information assuming inaccurate or absent robot pose information, and assuming general 3-D configurations, e.g., not limited to 2-1/2D elevation maps. We report on experiments and applications in the areas of terrain map building, modeling of interior environments, modeling of individual objects from many views, cooperative stereovision using teams of robots, and simultaneous recovery of structure and motion from bearing-only sensors.

This paper is a survey paper in which we concentrate on illustrative results and summary performance in the various applications. Earlier publications with detailed technical descriptions are referenced as needed.

2. Surface Matching

A first approach to 3-D mapping and localization involves the direct matching of 3-D surfaces. In this approach [13], the input to the mapping algorithm is a set of 3-D surfaces represented by triangulated meshes and, possibly, initial estimates of the relative poses between the surface. Local shape descriptors, or “signatures”, are computed in the neighborhood of points on the surfaces and are compared in order to identified potential correspondences between the surfaces. Once initial correspondences are identified, they are grouped into sets of geometrically consistent correspondences. The final transformations between views are estimated by minimization of the distance between corresponding points.

Although straightforward in principle, this approach raises a number of issues. First of all, the signatures must be carefully designed so that they are invariant by rigid transformations but, at the same time, remain discriminative enough to find correspondences. Also, the signatures must be computable on general surfaces, not limited to smooth, continuous surfaces. Even if the signatures were optimally discriminative, the process of grouping correspondences can become combinatorial if the surfaces exhibit strong symmetries. One simple example is the case of surfaces with large, near-planar sections in which all the points are equivalent insofar as the local shape of their neighborhoods is concerned. Although many techniques exist for registering surfaces when the initial estimate of their relative poses is fairly accurate, it is important to be able to compute correspondences even in the absence of an initial estimate of the transformations. In that case, the search for correspondences must be carefully implemented to avoid unacceptable computation times. Finally, in actual applications, the volume of data to be manipulated may be very large, e.g., in the case of terrain maps. Data structures that enable fast access to 3-D information, nearest neighbors, and local surface evaluation must be employed

The examples presented in the balance of this section illustrate applications of this general approach to problems that range from very large-scale terrain maps, to smaller-scale but higher resolution maps, such as indoor environments, to very high resolution unconstrained modeling of individual objects.

2.1 Terrain Maps

An ideal test for demonstrating the algorithm is the problem of building large, high-resolution three-dimensional representations of unstructured terrain using range sensors that operates over tens to hundreds of meters. Such an application tests the limits of the approach, in particular as it pertains to the computational issues related to data size. In order to test those limits, we have developed algorithms for constructing terrain maps from unregistered 3-D observations [12]. In this application, it is necessary to address the issues of widely varying resolution, absence of reliable features, and very large data sets. Furthermore, the data sets are typically very large, thus demonstrating the use of signature-based surface matching for large-scale problems.

In this application, the input to map registration is a pair of arbitrary polygonal surface meshes, possibly containing holes and disconnected patches, and the output is the 6-DOF rigid-body transform that best aligns the two meshes. In order to fully exercise the surface matching capabilities, we do not assume any prior knowledge of the transformation between individual maps.

In the example experiments reported here, we created meshes from real range data obtained from two sensors, one ground-based and one aerial. The Z+F/Quantapoint (ZF) sensor, a scanning laser range finder mounted on two of CMU's autonomous ground vehicles, produces 360 degree by 30 degree range and reflectance images in a radius of 52 meters. The ZF was configured to generate range and reflectance images of size 6000 x 300, corresponding to angular resolutions of 0.06° and 0.1° in the horizontal and vertical directions, respectively. The initial data is subsampled by a factor of 5 horizontally and 3 vertically and then converted to a mesh.

Once a mesh is formed, it is simplified to a predetermined number of faces using a standard simplification algorithm [15]. A pair of simplified meshes is then presented to the registration engine. Local surface signatures are generated for each point in one mesh and for a fixed percentage of points in the other. Candidate correspondences are found using the correlation-based similarity metric, and the best correspondences are grouped based on geometric consistency. We then determine the rigid-body transform that best aligns the correspondences from each group. Finally, we use high-resolution versions of the surface meshes to refine our transform estimate using a modified version of the ICP algorithm [34][1]. After all pairs of meshes in a sequence have been registered, they are integrated into a single global map using a standard voxel-based method [4].

Figure 1. shows an example of integrated map obtained using the first data from the ZF scanner. The map was created by registering eleven individual maps obtained at 3 to 5 meters interval. The initial data size is 1.8M points for the BF2 data, which is reduced to 15000 for matching. For that data, the variation of resolution from near to far range is greater than 1:10 due to the shallow incidence angle of the laser at that range.

In addition to ground-based data, the technique has been applied to data from the CMU autonomous helicopter. The helicopter is equipped with a single-scanline Riegl laser, which is a time-of-flight range finder with a range of 100m [19]. Combining the ground and aerial data showed that registration is possible between data from different sources with different resolutions and different noise characteristics. Figure 2. shows typical data used for registration from ground and helicopter data.

The experiments show that the surface-based matching techniques perform well with large data sets and that it can tolerate large variations in resolution, two issues that are critical to the practical application of surface matching.

Robustness to variations in resolution is especially critical for application in which high-resolution ground-based data is registered with low-resolution aerial data, for example a standard DTED map. Figure 3. shows an example of registration of a local (15m on the side) map with a global (about 100m) map. The resolutions of the maps differ by a factor of 3. In this example, no constraint was placed on the 6D transformation between the maps.

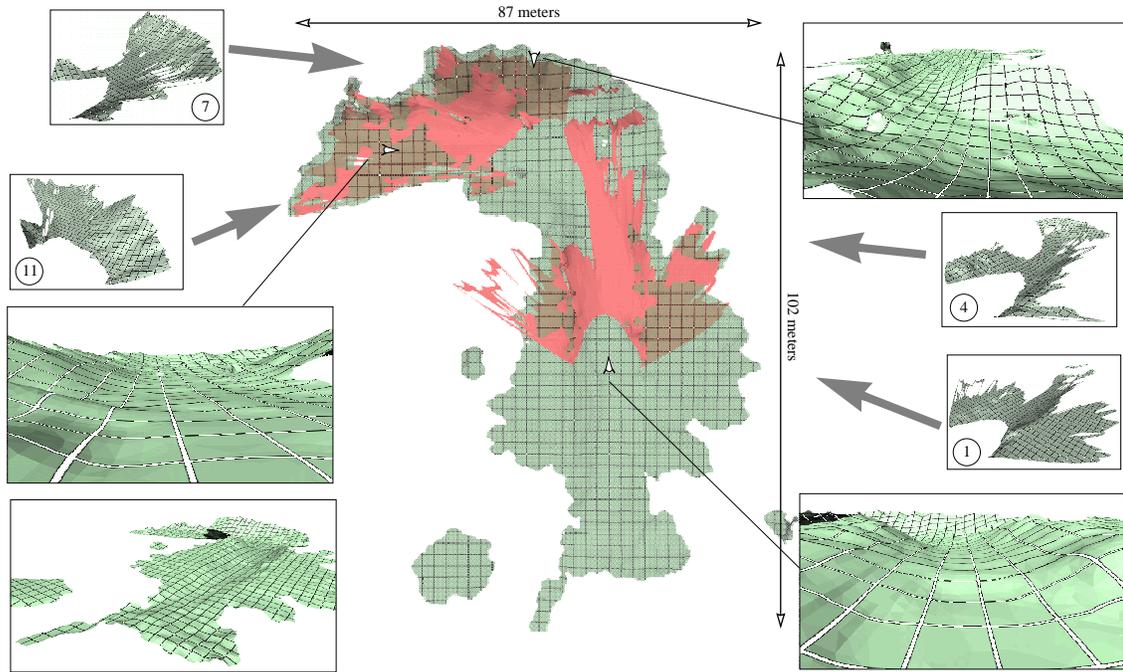


Figure 1. Top view of the integrated terrain map for the eleven data sets in the mesa sequence (center). The numbered insets illustrate individual data sets, and two sets (1 and 11) are overlaid on the top-view. The larger insets show various perspective views of the map, and the white arrows indicate the location and direction of the viewpoint. Grid lines are two-meters apart on the combined map and one meter apart on the individual data sets.



Figure 2. Data from a ground sensor (left) and a helicopter (right) shown as elevation maps. The corresponding area is shown at center.

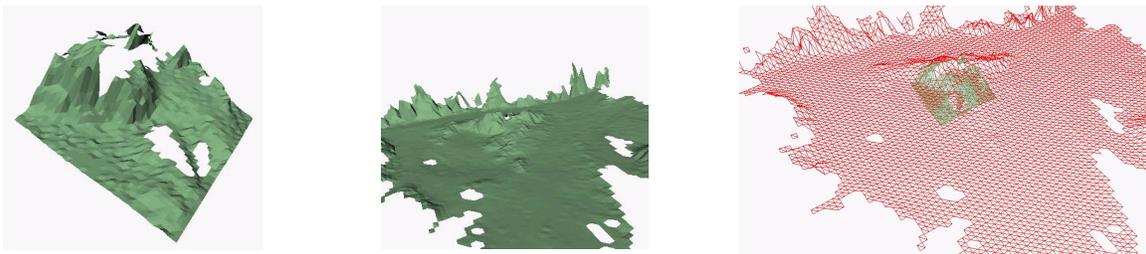


Figure 3. Registration (left) of high-resolution local map (left) with low-resolution global map (right).

2.2 3-D Mapping of Interior Environments

Another example application is the mapping of interior environments. The difference with the previous case of terrain mapping is that the environment is more structured and that the data is typically shorter range but higher resolution.

Given two low-resolution versions of the input meshes, we register them using local signatures as before: First, a

fixed fraction of points is selected at random from both surfaces. Signatures are produced for these points, and all the signatures from one surface are compared to all the signatures from the other using the similarity measure described above. When two signatures are found to have high similarity, the points that produced them are considered to correspond to each other; after all signature comparisons have taken place, we are left with a set of point matches between the two surfaces. Finally, an estimate of the transformation that aligns the two surfaces is computed from the point matches.

For example, this method of surface registration has been successfully applied to build a model of a large warehouse space composed of two adjacent rooms. The building measured roughly 60 meters long by 20 meters wide by 10 meters high, and was filled with an assortment of clutter and debris. We used the ZF laser range finder to collect data at 32 different points in the warehouse; each scan of the sensor covered a 360-degree field of view with 20 degree depression and a maximum range of about 40 meters. Each scan of the range finder returned 1.8 million 3-D points; through naive subsampling this set was reduced to 65000 points. The resulting point cloud was converted into a mesh and resampled to roughly 5000 points for registration. Surface meshes produced from each of the 32 data sets were successfully registered to the meshes of adjacent scans. The complete mesh of the warehouse contained 138000 points at full resolution, while a low-resolution version of the model contained 25000 points.

While it is important that no prior estimate of the transformation between two surfaces is required to successfully register them, a rough estimate of the transformation can be helpful in guiding the search for possible point correspondences and safeguarding against registration mistakes. In our procedure for surface registration, we allow the user to specify information about the ranges of possible translation and rotation between viewpoints to improve registration performance.

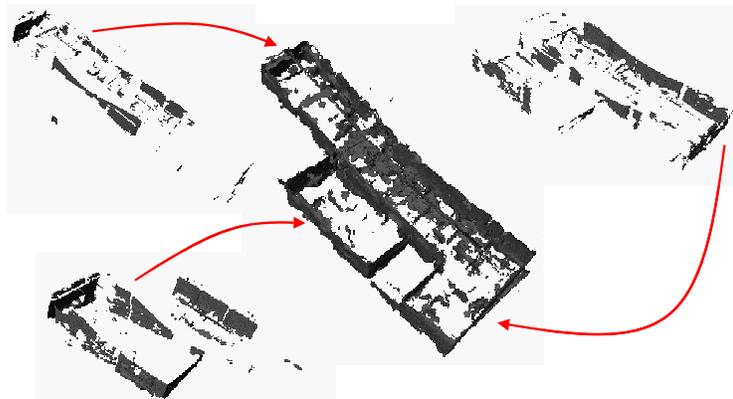


Figure 4. Integrated surface mesh (center) built from 20 individual 3D views.

2.3 View Selection and Global Registration

The last application of 3-D surface matching is object modeling, in which an object is presented in arbitrary poses to a sensor and a consistent model is constructed from the views [10]. We showed that, by using a surface matching approach similar to the one described above, it is possible to fully automate the 3D modeling process without resorting to these restrictive requirements (Figure 5.) The key difference with the previous applications is that we have to deal now with many views, high-resolution, small spatial extent, whereas in the terrain map case, the issues were large spatial extent and large volume of data rather than resolution and number of views.

The registration phase can be considered in terms of three interrelated problems: 1) determining which views contain overlapping scene regions; 2) determining the transform between each pair of overlapping views; and 3) determining the position of all views in a global coordinate system, which is the ultimate goal of the registration phase. Once the overlaps and relative poses are known, the absolute poses can be computed by simultaneously registering all overlapping views (multi-view registration). Multi-view registration can be posed as an optimization problem over the continuous space of absolute pose parameters, which has been solved by others. The objective function seeks to minimize the distance between corresponding points on overlapping surfaces, thereby distributing the errors in the relative pose estimates over the entire model.

For automatic modeling, however, both the amounts of overlap and the relative poses are unknown, which makes the registration problem considerably harder due to the mutual dependency between the overlaps and relative poses. The general approach is to embed the surface matching code into a graph optimization problem: Given a set of views $\{V_i\}$, a graph is created for which the nodes are the views and each arc (V_i, V_j) is weighted by the likelihood that the pair-

wise registration of the two views is correct. The likelihood can be evaluated from the amount of overlap between the views after pairwise registration. Additional criteria used for computing the likelihood includes the visibility consistency which is computed by evaluating the degree of visibility of one surface from the sensor, given the registered position of the other surface.

Starting with a graph G of views, the modeling algorithm attempts to find a covering of the graph by arcs with high registration likelihood. Figure 6 shows an example of a small graph of views. As a minimum, a spanning tree of the graph can be extracted which characterizes the combination of views with the highest likelihood of registration when looking at each pair of views independently. The issue is that views may be consistent locally, but not globally. Therefore, the key to the modeling algorithm is to extract a globally consistent subgraph G' of views, given the pairwise matches. Two approaches can be used for updating the graph. In the incremental update approach, new views are incrementally added to G' until a globally consistent graph is obtained.

The stochastic update approach is similar to simulated annealing in which the total registration likelihood of G' is maximized by adding new views and links to the graph incrementally by using a stochastic process. In both cases, the registration poses are re-estimated whenever the graph is updated by using the global optimization approach described above. Figure 5 shows one result of the automatic modeling applied to the hand-held modeling problem.

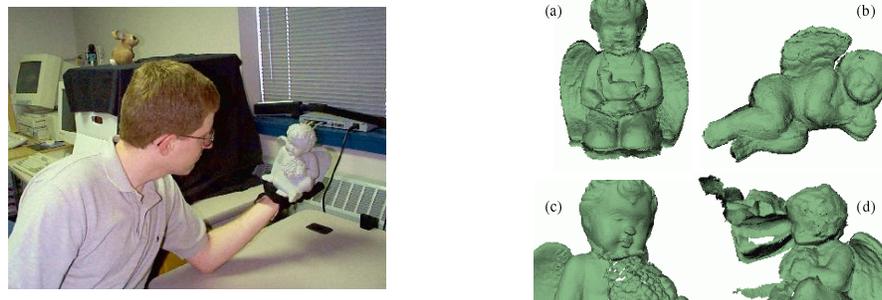


Figure 5. The hand-held modeling application -- a 3D model is automatically constructed from scans of an object (right) by presenting the object to scanner in arbitrary orientations (left).

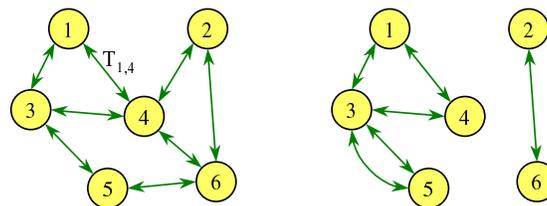


Figure 6. Example model graphs. A complete model (left) and two partial models (right)

3. Feature Matching

Although direct surface matching is appropriate when the data is naturally 3-D, from example when stereo or ladar is used, feature matching is a more natural approach when appearance information can be used. We describe below two possible different aspects of registration and mapping from features. In the first one, 3-D maps of environment from cooperative stereo between multiple robots. The second example involves simultaneous reconstruction of maps and localization using bearing measurements, such as feature tracking in video data.

3.1 Cooperative Stereo

A classical approach to recovering 3D information from cameras is stereovision from an on-board stereo rig. While, in principle, it is possible to recover 3-D data from a single robot with stereo cameras, using cooperating robots has important advantages. One obvious advantage is the savings in hardware and computation time since a single camera is needed on each robot and the full load of the stereo computation does not reside on any given robot. A more important advantage, however, is that stereo reconstruction accurate enough for rendering can be achieved from a long range by taking advantage of the fact that the robots may be separated by a wide baseline. To be useful, the system must generate at least qualitatively correct reconstruction up to 10-20m, which is not possible using the short baseline that is typical on small robots.

Given images of a scene from arbitrary robot positions, the first task is to recover the relative poses of the cameras. These poses must be computed precisely since an error of a few image pixels in mapping points from one image to the next may lead to large errors in the 3-D reconstruction. In particular, the pose estimates from dead reckoning are not precise enough for image matching. The general approach is to extract point features from the images, find correspondences between the two sets of features and recover the relative poses of the cameras from the feature matches. Typical features used for calibration are shown in Figure 4. These features were extracted using an interest operator derived from the Harris operator [43].

After initial feature matches are derived, the geometry of the cameras can be dynamically calibrated by estimating the epipolar geometry from correspondences between features. In the absence of any pose information, it is well-known that standard techniques for epipolar estimation are not robust because of the large percentage of wrong matches [33]. This is particularly a problem in the case of cooperative stereo in which matches are typically very ambiguous. An alternative approach is to first use a standard epipolar estimation technique, and to refine the estimate of camera geometry by using a non-linear, robust estimator [28].

Given the calibration information from the two robots, the 3-D locations of each feature can be recovered by triangulation [27]. For example, Figure 6 (left) shows the depth reconstructed of matched features of Figure 4, as well as the same result (right) in which the 3-D points corresponding to the features are shown in an oblique view. The lines are drawn from the point features to the ground plane.

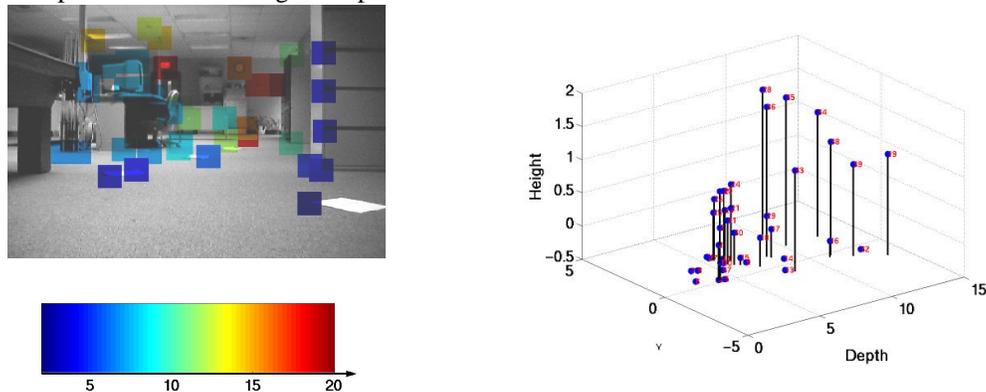


Figure 7. (Left) Sparse depth map computed from the features of Figure 4. The depth is indicated by a color-coded square centered at each feature. The color scale is indicated below the figure. (Right) Oblique view of the reconstructed 3-D points corresponding to features of Figure 4.

Given two robot positions and assuming that the dynamic calibration problem has been solved as described above, the image correspondence problem becomes equivalent to the standard problem of stereo except that the images are taken from different robots instead of from a single camera rig, as is normally the case. While in principle this approach is feasible, there are two key differences with standard stereo from fixed cameras, which are both related to the fact that the relative positions and orientations of the robots are arbitrary. First, as noted, the baseline between the robots may be large. Although necessary for accuracy reasons, this complicates the matching between images because of potentially large distortions between the two images. Second, arbitrary relative positions of the robots may lead to unacceptable image distortion after the rectification necessary to align the scanlines of the images. This second problem is addressed by using a polar rectification method which is guaranteed to rectify images in arbitrary geometry and to guarantee bounds on the amount of warping in the rectification [26][22]. Results are shown in Figure 8. and Figure 9.(a) and (b).

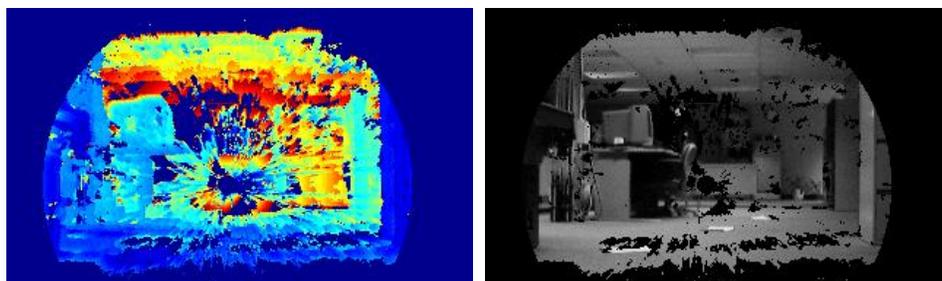


Figure 8. Color-coded depth map after filtering (left) and corresponding mask (right.)

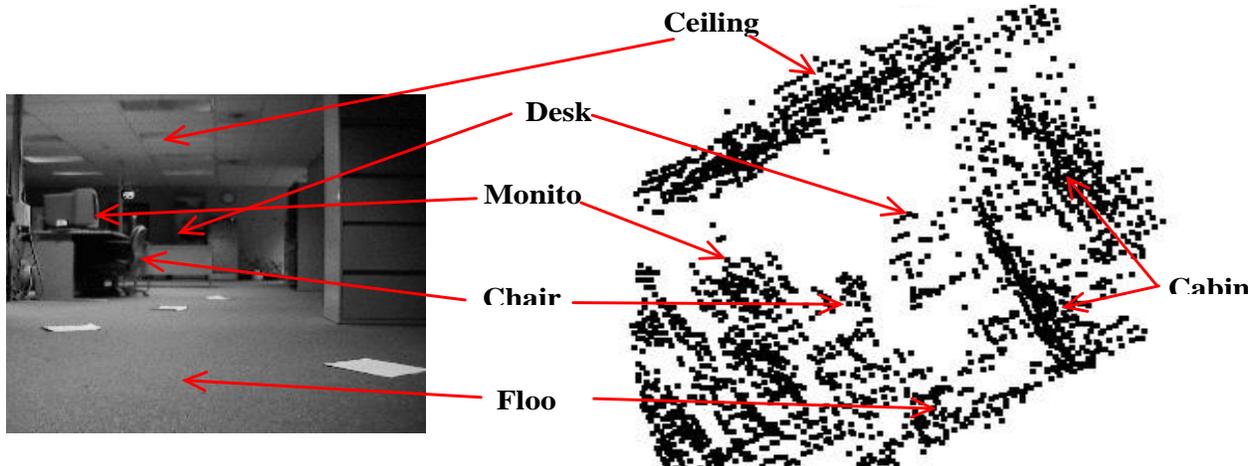


Figure 9. 3-D View of the point cloud constructed by our cooperative stereo mapping.

3.2 Localization from Bearing Measurements

In principle, a more direct way of building maps from features is by accumulating features detected in images and their correspondences across frames as a robot with a single camera moves [5][6]. In this scenario, the robot has no *a priori* knowledge of the environment (no map) and can only observe its egomotion through odometry and the bearings to landmarks in the environment, e.g., with an omnidirectional camera. The robot has no external position or heading reference, and cannot measure the range to landmarks.

This problem is often termed Simultaneous Localization and Mapping (SLAM), or Concurrent Localization and Mapping (CLM). Approaches generally rely on recursive filtering using a Kalman filter or some variant [16]. There have also been attempts to find optimal estimates using batch techniques [18] or to deal with large environments using a hierarchical approach [3][16].

The Computer vision literature contains a large body of work on multi-frame Structure From Motion (SFM). Bundle adjustment is a batch technique that is widely used for SFM under perspective projection [9][32] and is applicable to SLAM. Some attempts have also been made to recursively compute structure and motion using Kalman filters [20][20]. The Kalman filter and SFM views of the SLAM problems have been developed largely independently. In this section, we briefly show that these two classical approaches can be combined into an approach to mapping that retains the accuracy of SFM while taking advantage of the computational efficiency of the Kalman filter formulation.

The SLAM and SFM problems contain features relevant to localization and mapping with a bearing-only sensor. Specifically, the Structure from Motion (SFM) work in computer vision provides insight into what we can expect from bearing-only SLAM. In SFM the 3D structure of an environment is reconstructed using 2D projections (images) from unknown camera locations. The perspective camera projection model is very similar to a bearing-only sensor; depth information is lost in the projection. The only knowledge of the scene from a single image is that scene points lie somewhere along a ray originating at the camera center of projection and passing through the image plane at the point where the feature appears. In bearing-only SLAM, landmarks are projected onto a unit circle around the robot, and a similar constraint applies for the ray intersecting this projection surface.

In most SFM algorithms, little or no *a priori* information about the camera motion is used for reconstruction, camera motion is recovered in the estimation along with scene structure using only the image measurements. Odometry can be seen as providing a prior for the camera motion which is absent in general SFM problems. If we add odometry then we can immediately disambiguate the scale.

There is even a unique solution to the two frame problem if odometry is available since it will provide a direct estimate of the relative pose of the second view. However until the third view is incorporated, bearing measurements cannot begin to correct odometry errors within the estimation algorithm. This can cause initialization problems with Kalman filters which are observed in practice, particularly when odometry is poor.

A final note is that bias in odometry cannot always be corrected in bearings-only SLAM. If the measurements of distance travelled are biased by a scale factor then the estimate of robot motion *and* landmark positions will be scaled by the same factor and no estimator can recover the bias using the bearing measurements. This is an inherent ambiguity. In contrast, SLAM with range-bearing sensors does not suffer from this problem since the range measurements pro-

vide direct information about the scale of the solution.

Finding an estimate for the model parameters $\mathbf{pose} = (\mathbf{pose}_0, \mathbf{pose}_1, \dots, \mathbf{pose}_M)$ and $\mathbf{landmark} = (\mathbf{landmark}_0, \mathbf{landmark}_1, \dots, \mathbf{landmark}_N)$ can be done in many ways. Bundle adjustment, in particular, is a full nonlinear optimization which does not rely on a mean and covariance as a sufficient statistic for previous observations and state estimates. Instead it linearizes the estimation problem at every step using all available observations and the current best estimate. Bundle adjustment may optimize all motion and structure parameters at every step, so the state vector is the entire history of robot pose (trajectory) and the entire map.

We wish to minimize the cost function $E(\mathbf{param}) = (\mathbf{d} - f(\mathbf{param}))^T \mathbf{R}_d^{-1} (\mathbf{d} - f(\mathbf{param})) + (\mathbf{z} - h(\mathbf{param}))^T \mathbf{R}_z^{-1} (\mathbf{z} - h(\mathbf{param}))$ where $\mathbf{param} = (\mathbf{pose}, \mathbf{landmark})$ is the vector of parameters to be estimated, \mathbf{d} and \mathbf{z} are all odometry and all bearing measurements stacked into vectors, and $f()$ and $h()$ are all predicted odometry and bearing measurements stacked into vectors. The first term penalizes robot motion that does not agree well with odometry measurements and the second term penalizes robot motion and landmark map combinations that do not agree well with bearing measurements.

The bundle adjustment iterations involve the inversion of the approximation of the Hessian of E , $\mathbf{J}^T \mathbf{J}$, where \mathbf{J} is the Jacobian of E . The matrix inversions involved in the minimization may be daunting in full-size problems. Fortunately, photogrammetrists have known for decades how to speed up bundle adjustment using the sparse nature of the structure from motion problem. We can take a similar approach here by exploiting the nature of the bearing-only SLAM problem, leading to an efficient solution to the problem.

Figure 10.(a) shows the result obtained by bundle adjustment using data from an RWI ATRV. The robot drove in an approximately elliptical path between six different artificial landmarks. There was a problem with the drive mechanism which caused bias in the odometry measurements. Dead reckoning estimated that the robot drove in a spiral shaped path. Without modeling the bias the bundle adjustment recovers the elliptical path of the robot.

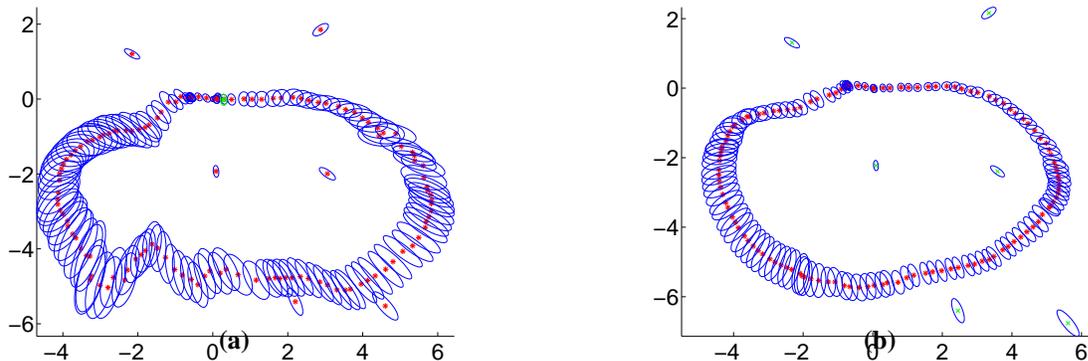


Figure 10. Results on real data (a) Kalman filter (b) Bundle adjustment.

In contrast, a standard Kalman filter approach would diverge in the presence of the odometry bias. The odometry was calibrated by fitting a linear function of the data to the bundle adjustment trajectory estimate. With the calibrated odometry and an initial model from bundle adjustment using the first 20 robot poses, the Kalman filter was able to produce the result in Figure 10.(b). The lower left portion of the trajectory shows some reconstruction behavior which is most likely due to the introduction of the lower left feature. Since it processes one measurement at a time the Kalman filter does not initialize features well with bearing-only measurements.

This example suggested a combined approach to bearing-only localization and mapping that would combine the advantages of classical Kalman filter localization and of map reconstruction through bundle adjustment. Specifically, we developed a recursive formulation so that the computational burden does not grow with the number of sensor scans made, but we also wish to avoid the bias introduced in the linearization phase in the Kalman filter by considering batches of data rather than individual measurements.

One way to accomplish this is to process data in batches, using observations made during a time interval and computing an estimate for the map and for the pose of the robot during the time interval. At the end of the time interval, the state estimate and all of the data from the time interval is collapsed onto a sufficient statistic, a mean and covariance for the map of landmark features and the last robot pose.

The very first batch of data can be processed in exactly the same way as was described for bundle adjustment. At the end of that estimation. As observations are made from new robot positions, the state vector grows to and the information from the bearing and odometry measurements are added into the information matrix. It can be shown that the

update step of the new recursive/batch filter is $O(N^3+kN^2)$ where k is the size of the time window used for the batch update. For small fixed k this is approximately the same computational complexity as the Kalman filter. The algorithm is similar to one reported recently for SFM [21].

Results of applying this filter with a batch size of 10 to the experiment discussed earlier are shown in Figure 11.(a). The uncertainty ellipses on each robot pose are plotted as computed at the end of a batch update, so smoothing does not take place over the whole trajectory as it does with full bundle adjustment. Due to this the groups of 10 poses per batch are apparent from the uncertainty ellipses. Figure 11.(b) shows the results from bundle adjustment and recursive/batch methods plotted against each other for comparison. Results for the example considered are very similar for the two methods.

These examples illustrate the advantages of bundle adjustment over Kalman filtering for the bearing-only SLAM problem. When Kalman filters are used, they need to be initialized with a good estimate of the robot state and landmark locations and bundle adjustment is an efficient way to compute optimal least squares estimates in order to initialize the filter. Bundle adjustment is empirically more robust to bias in the system model, which is to be expected since the linearizations are recomputed each time and are computed near the optimum.

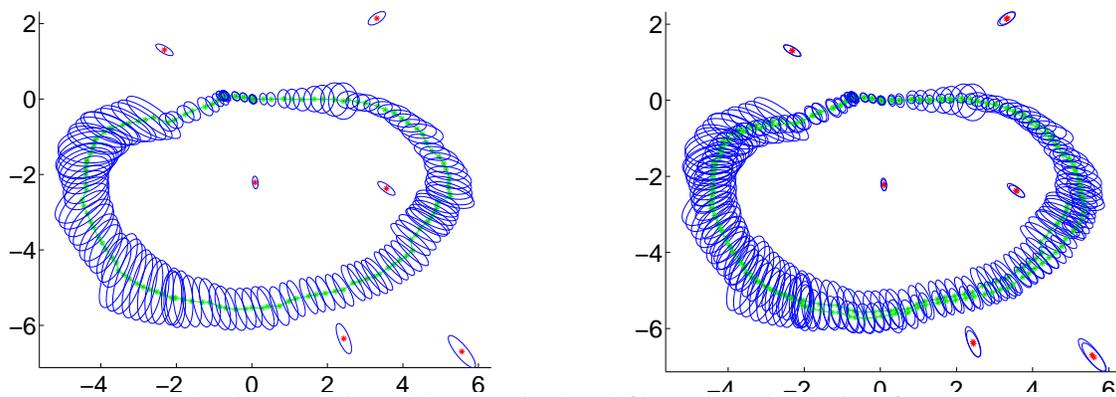


Figure 11. (a) map and trajectory estimated by recursive/batch filter using a batch size of 10. (b) result superimposed with result from bundle adjustment to show agreement.

For computational and memory requirements, bundle adjustment is not practical for large problems. In contrast, the Kalman filter requires $O(N^2)$ memory. The Kalman filter requires $O(N^3)$ computation for each update. Computation required in bundle adjustment is $O(N^3+MN^2)$ which grows linearly in the number of robot positions in the estimate. This is significantly less than a general second order least squares optimization but still not practical for a robot operating for long periods of time. By combining Kalman filtering and bundle adjustment concepts, a new recursive/batch filter was introduced. The filter allows a flexibility in the batch size, trading off computational requirements and performance. If the batch size is set to 1, the extended Kalman filter results, and if the batch size is unlimited, full bundle adjustment results.

4. Conclusion

The examples presented here show that a variety of techniques can be brought to bear on the problem of 3-D registration even in the case of very large data sets (terrain maps), large sets of views (individual objects), or 3-D reconstruction from video (cooperative stereo and bearing-only SLAM.) The underlying approaches can be viewed as building blocks for complete systems that include model recovery as well as localization and state estimation. Much remains to be done before such a system becomes practical. One of the most important tasks in this area is to combine the deterministic registration techniques illustrated above with stochastic localization techniques [31]. This would combine the accuracy of the registration techniques with the flexibility of the localization techniques. For this to occur, however, major advances in the computational efficiency of the registration algorithms need to be achieved. In the case of terrain maps, in particular, the registration algorithms are not yet efficient enough to be performed on-line.

References

- [1] P.Besl and N.McKay. A method of registration of 3-D shapes. IEEE Trans. on Pattern Analysis and Machine Intelligence, 14(2):239--256, Feb. 1992.
- [2] O. Carmichael and M. Hebert. Unconstrained Registration of Large 3D Point Sets for Complex Model Building. 1998 IEEE/RISJ International Conference On Intelligent Robotic Systems, Vancouver, September 1998.
- [3] K.Chong and L.Kleeman. Large scale sonarray mapping using multiple connected local maps. In International Conference on Field and Service Robotics, pages 278--285, 1997.

- [4] B. Curless and M. Levoy. A volumetric method for building complex models from range images. In *Proceedings of SIGGRAPH '96*, pages 303--312, 1996.
- [5] M. Deans and M. Hebert. Invariant filtering for simultaneous localization and mapping. *Proc. Symposium on Simultaneous Localization and Mapping*. IEEE International Conference on Robotics and Automation. May 2000.
- [6] M. Deans and M. Hebert. Experimental Comparison of Techniques for Localization and Mapping using a Bearings Only Sensor. *Proc. of the ISER '00 Seventh International Symposium on Experimental Robotics*. December 2000.
- [7] M. Dissanayake et al. An experimental and theoretical investigation into simultaneous localization and map building (slam). In *Proc. 6th International Symposium on Experimental Robotics*, pages 171--180, 1999.
- [8] S.F. El-Hakim, P. Boulanger, F. Blais, and J.-A. Beraldin. A system for indoor 3-D mapping and virtual environments. In *Proceedings of Videometrics V (SPIE vol. 3174)*, pages 21--35, July 1997.
- [9] R.I. Hartley. Euclidean reconstruction from uncalibrated views. In Zisserman Mundy and Forsyth, editors, *Applications of Invariance in Computer Vision*, pages 237--256. Springer Verlag, 1994.
- [10] D. Huber. Automatic 3D Modeling Using Range Images Obtained from Unknown Viewpoints. *Proc. Intern. Conf. on 3D Imaging*. Quebec City, May 2001.
- [11] Daniel Huber, Owen Carmichael and Martial Hebert. 3-D Map Reconstruction from Range Data. *IEEE Conference on Robotics and Automation*, April, 2000.
- [12] D.F. Huber and M. Hebert. A New Approach to 3-D Terrain Mapping. *Proc. of the 1999 IEEE/RSJ International Conference on Intelligent Robotics and Systems (IROS '99)*, IEEE, October, 1999, pp. 1121-1127.
- [13] A. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(5):433--49, May 1999.
- [14] S. Julier, J. Uhlmann, and H. Durrant-Whyte. A new approach for filtering nonlinear systems. In *Proceedings of the 1995 American Controls Conference*, pages 1628--1632, 1995.
- [15] M. Garland and P. Heckbert. Surface simplification using quadric error metrics. In *Proceedings of SIGGRAPH 97*, 1997.
- [16] J.J. Leonard and H. Feder. Decoupled stochastic mapping. Technical Report 99-1, MIT Marine Robotics Laboratory, Cambridge, MA 02139, USA, 1999.
- [17] J.J. Leonard and H.F. Durrant-Whyte. Simultaneous map building and localization for an autonomous mobile robot. In *IEEE/RSJ International Workshop on Intelligent Robots and Systems IROS '91*, pages 1442--1447, 1991.
- [18] F. Lu and E. Milios. Globally consistent range scan alignment for environment mapping. *Autonomous Robots*, 4(4):333--349, 1997.
- [19] J.R. Miller, O. Amidi, and M. Delouis. Arctic test flights of the CMU autonomous helicopter. In *Proceedings of the Association for Unmanned Vehicle Systems, 26th Annual Symposium*, July 1999.
- [20] P. McLauchlan. Gauge invariance in projective 3d reconstruction. In *Proceedings IEEE Workshop on Multi-View Modeling and Analysis of Visual Scenes (MVIEW'99)*, pages 37--44, 1999.
- [21] P. McLauchlan. A batch/recursive algorithm for 3d scene reconstruction. In *Proc. CVPR 2000*, pages 738--743, 2000.
- [22] M. Pollefeys, R. Koch, L. Van Gool, "A Simple and Efficient Rectification Method for General Motion" *Proc. IEEE International Conference on Computer Vision*, Corfu, 1999.
- [23] K. Pulli. Multiview registration for large data sets. In *Proceedings of the Second International Conference on 3-D Digital Imaging and Modeling (3DIM'99)*, pages 160--8, Oct. 1999.
- [24] P. Robert and D. Minaud. Integration of multiple range maps through consistency processing. In R. Koch and L. van Gool, editors, *3D Structure from Multiple Images of Large-Scale Environments*. European Workshop, SMILE'98, pages 253--65. Springer-Verlag, June 1998.
- [25] R.P.N. Rao. Robust kalman filters for prediction, recognition, and learning. Technical Report 645, Computer Science Department, University of Rochester, 1996.
- [26] S. Roy, J. Meunier, I. Cox, "Cylindrical Rectification to Minimize Epipolar Distortion" *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 1997.
- [27] P. Sturm and R. Hartley, "Triangulation" *Computer Vision and Image Understanding*, 68(2), 1997.
- [28] S.M. Thayer, B. Dias, B. Nabbe, B. Digney, M. Hebert, A. Stentz. Distributed robotic mapping of extreme environments. *Proc. SPIE Mobile Robots XI*. 2000.
- [29] G. Roth. Registering two overlapping range images. In *Proceedings of the Second International Conference on 3-D Digital Imaging and Modeling (3DIM'99)*, pages 191--200, Oct. 1999.
- [30] V. Sequeira, K. Ng, E. Wolfart, J. Goncalves, and D. Hogg. Automated 3D reconstruction of interiors with multiple scan-views. In *Proceedings of Videometrics VI (SPIE vol. 3641)*, pages 106--17, Jan. 1999.
- [31] S. Thrun, D. Fox, W. Burgard, and F. Dellaert. Robust Monte Carlo Localization for Mobile Robots. *Artificial Intelligence Journal*, 2001.
- [32] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle adjustment - a modern synthesis. In *Vision Algorithms: Theory & Practice*. Springer-Verlag, 2000.
- [33] Z. Zhang, R. Deriche, O. Faugeras, Q.T. Luong, "A Robust Technique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry" *Artificial Intelligence Journal*, Vol. 78, 1995.
- [34] Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision*, 13(2):119--152, Oct. 1994.