Registration of Multiple Range Scans as a Location Recognition Problem: Hypothesis Generation, Refinement and Verification

Bradford J. King‡*   Tomasz Malisiewicz*   Charles V. Stewart*   Richard J. Radke†
kingb2@cs.rpi.edu   *malist@cs.rpi.edu*   *stewart@cs.rpi.edu*   *rjradke@ecse.rpi.edu*
* Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY 12180
† Department of ECSE, Rensselaer Polytechnic Institute, Troy, NY 12180
‡ Kitware, Inc., Clifton Park, NY 12065

Abstract

This paper addresses the following version of the multiple range scan registration problem. A scanner with an associated intensity camera is placed at a series of locations throughout a large environment; scans are acquired at each location. The problem is to decide automatically which scans overlap and to estimate the parameters of the transformations aligning these scans. Our technique is based on (1) detecting and matching keypoints — distinctive locations in range and intensity images, (2) generating and refining a transformation estimate from each keypoint match, and (3) deciding if a given refined estimate is correct.

While these steps are familiar, we present novel approaches to each. A new range keypoint technique is presented that uses spin images to describe holes in smooth surfaces. Intensity keypoints are detected using multiscale filters, described using intensity gradient histograms, and backprojected to form 3D keypoints. A hypothesized transformation is generated by matching a single keypoint from one scan to a single keypoint from another, and is refined using a robust form of the ICP algorithm in combination with controlled region growing. Deciding whether a refined transformation is correct is based on three criteria: alignment accuracy, visibility, and a novel randomness measure. Together these three steps produce good results in test scans of the Rensselaer campus.

1. Introduction

This paper addresses a new form of the multiple range image registration problem. The data are range scans taken from a variety of positions throughout a large area — a university campus is our test case. These scans might overlap by a little bit, by a large fraction, or not at all. The scans are not taken incrementally, and there will be at most a few scans of any region. Challenges include handling a variety of structures (such as buildings and landscapes), moving objects, changes in illumination, and significant occlusions. The goal is to determine which scans overlap and to compute the transformations that best align overlapping scans. We approach this goal as a "location recognition problem" because the major issue is a decision about which scans show part of the same environment.

Previous algorithms have approached multiple range image registration by assuming that either coarse initial alignments are available, or that the scans are taken of static scenes and have substantial overlap between them. The algorithms we develop should certainly work when such assumptions are met, but our goal is to place no restrictions on the relative placement of the scans. Similarly, GPS and/or inertial navigation units may be available to provide initial location estimates, but these systems sometimes fail. The techniques we develop here can help detect and correct these failures. Thus, solutions to the location recognition problem will make 3D environment modeling systems more autonomous, more flexible, and more robust.

We present our initial work toward solving this problem. Our overall approach is built on three algorithmic components: hypothesis generation, refinement, and verification. Hypothesis generation is the creation of initial transformation estimates through matching of keypoints — locations of distinctive structure in range scans or in the associated intensity images. We introduce a new range keypoint extraction technique, and we adapt image keypoint extraction techniques from the computer vision literature. Each pair of scans will lead to the creation of many keypoint matches.

A hypothesized 3D rigid transformation between a pair of range scans is generated for each keypoint match individually, even for intensity-image keypoint matches. Creating an estimate from each individual match avoids the need for combinatorial search methods or clustering techniques. This is important because particularly difficult data

sets have relatively few correct matches. Since the initial estimate generated from a keypoint match can only be trusted in a small region of the data surrounding the keypoint in each scan — even if the correspondence is correct — the refinement technique alternates steps of (1) re-estimating the transformation using a robust form of the iterative closest points (ICP) algorithm applied within the region, and (2) expanding the region. This allows inaccurate initial estimates to be locally refined before being applied to entire data sets.

The final component, verification, decides if a given refined transformation is correct and therefore if the location seen in one scan has been recognized in the other. This step is crucial in our approach because we make no attempt to pre-filter matches other than by the match similarity measure. Similar to other techniques [12], our decision criteria include measurements of the alignment accuracy and the fraction of visible surface points matched. Building from [4], we include a novel third measure designed to determine if the alignment is unambiguous. This new measure is based on regenerating keypoints using the estimated transformation. The final decision criterion is based on a combination of these accuracy, coverage and ambiguity measures.

2. Related Work

Several papers have addressed the problem of multiscan registration for building 3D models. In earlier work for modeling single objects, many overlapping scans are acquired and the main problem is high-precision registration starting from reasonable initial estimates (see e.g. [1, 13]). Some recent work has focused on modeling outdoor scenes, either single buildings or larger-scale areas. In [21, 22], registration is initialized by matching 3D line segments between scans, and clustering matched line segments to determine which scans overlap. In [8], data are obtained continually from a scanner mounted on a moving vehicle; airborne imagery is used to guide the alignment of these scans into a complete model. Recent multiscan registration work [12] has used spin images [14] to initialize registration, a graph-based multiscan refinement algorithm, and a verification step based on a combination of accuracy and visibility. While this work has an underlying assumption of static scenes and substantial overlap between scans, several of the techniques developed are applied and extended in our work.

Initialization methods have been the focus of substantial effort in recognition and in registration, using both 2D images and 3D range scans. For 3D scans, early work focused on detecting and matching corners and points of high curvature [5]. More recent work has focused on the use of spin images and related representations as summary descriptions (sometimes called "point descriptors") for matching [7]. Work on intensity images has emphasized both the

detection and description of point locations. Detection algorithms have used the multiscale Laplacian of Gaussian [16], and multiscale corner detection [18], among others. While a host of descriptors has been proposed, affine-invariant versions of Lowe's SIFT descriptor, which is a scale- and contrast-normalized histogram of intensity-gradient, and steerable filters have proven most effective [17]. Finally, researchers have combined both intensity and range data in matching, focusing on filtering matches [19] or using range data to normalize keypoint descriptions [23]. Both have reported preliminary results on alignment of single objects.

Work in *refinement* algorithms for range scan registration has focused on the ICP algorithm, which alternately generates temporary correspondences based on an estimated transformation and refines the transformation estimate based on the correspondences. ICP was proposed almost simultaneously in several papers in the early 1990s (e.g. see [2, 6]), and has been studied extensively since [20].

Automatic *verification* — the problem of determining whether an alignment between a pair of scans or images is correct — has received less attention. Most techniques are simply based on the accuracy of the alignment. More recent work has included tests based on visibility and occlusion [12], with the use of a Bayesian classifier to formulate the decision criteria [11]. In image registration, recent keypoint matching algorithms [4] use a probabilistic argument to determine the minimum number of "good" keypoint matches required for an alignment to be considered correct. We build on all of these techniques in our decision criteria.

3. Sensor and Data Collection

All data for the experiments in this paper were collected around the Rensselaer campus using a Leica HDS 3000 scanner (Figure 1). Bore-sighted RGB images are acquired by the scanner prior to the start of each scan. The images are well-calibrated relative to the scanner position, so that the lines of sight of the range points and image points coincide. Our experiments have verified that errors in the alignment are less than $0.1°$. The data include both buildings and natural scenes. In some cases a building that is clearly visible in one scan is significantly occluded in another.

4. Keypoint Extraction and Matching

Keypoints are distinctive image locations that can be described in an approximately viewpoint independent manner. They have been used for both object recognition and registration. Our work on the location recognition problem uses them as both the basis for hypothesis generation and as part of the hypothesis validation procedure. We propose two keypoint techniques, one based solely on the range data and

Figure 1. Three example range scans taken on the Rennsselaer campus illustrating some of the challenges of the location recognition problem, including occlusions, viewpoint variations, and different types of structures varying from buildings to trees and hillsides (right). The left and center pair has an extremely small overlap region, with the farthest-away building at the top of the center scan being the main building in the left scan.

one based primarily on intensity images. Range keypoints summarize local 3d structure at distinctive locations in the range scan, while intensity keypoints summarize local intensity structure at distinctive locations in the intensity data. These therefore offer independent, almost complementary, approaches to matching and hypothesis generation.

4.1. Range Keypoints and Matching

Our approach to developing range keypoints starts from the idea of a spin image [14], a 2D histogram of range points computed in a cylindrical coordinate system centered at a point of interest. The cylindrical axis may be oriented along an estimated surface normal or along the axis of least second moment. Integrating around the rotation axis condenses the coordinate system from 3D to a 2D histogram (hence the term "spin image"). Spin images are descriptors, not location detectors, and therefore must be combined with a location detection technique. Earlier work used all the data or a uniform sampling of locations [7, 12, 14]. We present an alternative approach designed to handle widely varying viewpoints, low image overlaps, and occlusions.

We detect locations at which to compute keypoints by finding holes in smooth surfaces. The spin image coordinate system is centered on the center of mass of the hole boundary, and the spin image is computed over a region whose size is a small constant multiple of the diameter of the hole (see Figure 2). In this way both the shape of the hole and the geometric structure of the surrounding region are combined in the spin image descriptor. Holes in manmade structures occur primarily at doors, windows, and indentations or outcroppings on the faces of buildings. There are not significant self-occlusions in the regions immediately surrounding the holes as there would be near more substantial depth discontinuities, so the computed spin im-

age should not change substantially with viewpoint (provided the spin-image calculation accommodates sampling differences). One major concern with using holes is that they represent repetitive structures; e.g. different windows on the same building often look the same. We accommodate this in part by expanding the size of the spin image region so that surrounding structure will make the spin image descriptors more distinctive. We also account for this in our matching and decision criteria, described below.

Clearly, the detection of holes requires segmentation, but segmentation of range images is still an unsolved problem [10]. Fortunately, we are more concerned with reliably, repeatably extracting the same region in different scans, rather than computing a segmentation that agrees with human perception. Moreover, our algorithm needs only one successfully matched hole in order to succeed, relaxing constraints on the robustness of the detection algorithm.

The segmentation technique itself extracts nearly planar surface regions using a technique similar to [3]. Robust estimates of local surface normals are first computed, and then seed regions are found where the smallest eigenvalue of the local covariance matrix is sufficiently small. These seeds are grown into surface patches and clustered based on similarity in surface normals and locations. Growth is prioritized so that a surface grows fastest where it is flattest. At the end, all clusters with fewer than 30 data points are eliminated. Sufficiently large regions of "missing pixels" in the planar patches are detected as holes. The missing regions need not be completely enclosed by the patch (e.g. a doorway at the base of a building). Typically 5-15 holes are extracted from range scans of buildings (Figure 2).

Spin images are computed by placing a coordinate origin at the center of the hole and orienting the axis of the spin cylinder along the surface normal of the surrounding surface. The contribution of each point is weighted by the

Figure 2. Automatically detected range keypoints, i.e. holes in planar surfaces where spin images are computed. Black pixels represent points on planar surfaces. In this example all significant holes are found. The circles represent the area over which the spin images are computed.

inverse of an estimate of the local sampling density, similar to [7]. Spin images are matched by computing the cosine of the angle between the vectors formed by each spin image histogram, giving a similarity measure in the interval [0,1].

4.2 Intensity Keypoints and Matching

Following [16], intensity keypoints are found by detecting peaks in the response of the Difference-of-Gaussians operator in the bore-sighted intensity images. Each keypoint has a location, orientation, scale, and descriptor. We use Lowe's SIFT descriptor [16], a 128-component histogram of normalized gradients, which has proven to be effective in experimental evaluations [17].

Equally-spaced samples on the perimeter of the scale-neighborhood around each 2D keypoint are backprojected along its camera's line-of-sight to the range surface and are used for three purposes: (1) If the change in depth between two consecutive samples is greater than some fraction (0.7) of the total variation in depth around the circle, then a depth discontinuity is identified and the keypoint is discarded (Figure 3). (2) The samples' distances from the 3D keypoint center are used to estimate a 3D scale. (3) The samples are used to estimate a local plane. The normal to this plane becomes the z axis of a local 3D coordinate system with its origin at the keypoint center. The x axis is computed by back-projecting the 2D keypoint orientation onto the local plane. A right-handed coordinate system is completed with $y = z \times x$.

Matching these 3D keypoints follows the standard methods in the literature. For each moving scan keypoint $\mathbf{p}_i$ the two keypoints $\mathbf{q}_{i,0}$ and $\mathbf{q}_{i,1}$ with the best matching descrip-
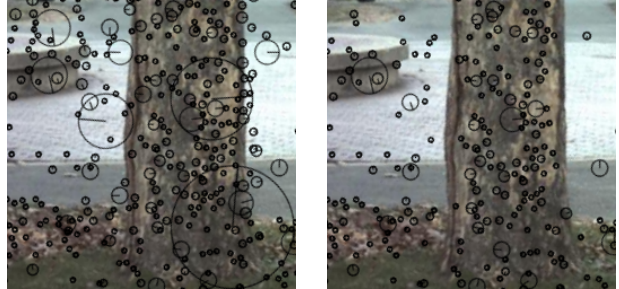

Figure 3. Intensity keypoints before (left) and after (right) range image filtering to eliminate keypoints near depth discontinuities.

tors from the fixed image are found. Let $S(\mathbf{p}, \mathbf{q})$ denote the similarity score between the descriptors associated with $\mathbf{p}$ and $\mathbf{q}$. Then, if

$$\frac{S(\mathbf{p}_i, \mathbf{q}_{i,0})}{S(\mathbf{p}_i, \mathbf{q}_{i,1})} < \theta \qquad (1)$$

for some $\theta < 1$, the match $(\mathbf{p}_i, \mathbf{q}_{i,0})$ is accepted as sufficiently unique. This *ratio test* is applied to the best match for each keypoint $\mathbf{p}_i$, and the surviving matches are sorted by increasing order of the ratio (1).

5. Refinement

At the refinement stage we are given a set of range and intensity keypoint correspondences. Each correspondence matches a 3D point from each of two different scans. The issue is how to turn these correspondences into hypothesized and refined transformation estimates. There are two novelties to our refinement procedure:

- Initialization from a single match
- A region growing variant of ICP

Instead of attempting to combine matches to estimate the transformation, an initial transformation between scans is computed for *each* match and then refined. Other techniques in the literature search for subsets of consistent matches to generate and evaluate the transformation estimate. This involves either a combinatorial search or clustering in parameter space. Both become increasingly problematic as the fraction of correct keypoint matches decreases (e.g. due to occlusions or differences in viewpoint). We don't want to rely on having a sufficient set of correct correspondences, either in percentage or absolute numbers; such a reliance limits the overall robustness of the algorithm. Instead we rely on rapid ICP refinement and a decision procedure that can reliably determine whether or not an alignment between scans is correct. We will therefore be able

to test hypothesized transformations one-by-one and stop when one has been generated that is correct.

Computing the initial rigid transformation from a match between one keypoint from each of two scans is simple if a 3D coordinate system has been established at each keypoint. For an intensity keypoint, this coordinate system is established when the keypoint is backprojected into 3D. For range keypoints, establishing the local coordinate system is straightforward. The local surface normal becomes the z axis. A second axis can either be computed from the moments of the hole boundary points or from the gravity direction obtained from the scanner. Projecting this onto the plane normal to z produces the x axis. The y direction is the cross-product $z \times x$.

A region growing variant of the ICP algorithm addresses a problem caused by using a single correspondence to initialize the transformation estimate. The initial estimate is often only accurate in the region of the two data sets immediately surrounding the correspondences (see Figure 4). This is especially likely for intensity keypoints because they are more local than range keypoints. Applying ICP throughout the data set starting from this initial estimate can lead to mismatches and incorrect convergence, even if the correspondence is correct. Our solution is to refine the estimate in a small region surrounding the corresponding points in each range scan, and then double the radius of the region and repeat the process.

Aside from this region growing, the actual ICP procedure is straightforward. Normal distance constraints are used [6, 20]. Matches are robustly weighted using the Cauchy M-estimator weight function [9], designed for a gradual downweighting of outliers. The robust standard deviation is estimated for the first ICP iteration in each region. While we could easily incorporate intensity measures into the ICP refinement process [15], we currently reserve the use of intensity constraints for the decision-making step. Overall, the algorithm tends to converge quickly.

6. Verification Criteria

The final step, verification, requires decision criteria to determine if a refined transformation is correct. We are interested in an absolute measure of correctness: one that can be applied to a single transformation estimate rather requiring a comparison between different estimates. This allows transformation estimates to be evaluated one-by-one using a greedy approach, but requires a rich measure that clearly separates good transformation estimates from poor ones. We combine three measures:

- Accuracy of the alignment
- Visibility constraints
- A non-randomness score

In the following, we denote I_1 as the moving scan, and I_2 as the fixed scan.

The accuracy measure is a robust mean-square error of the transformation estimate. If $\{\mathbf{x}_i\}$ is the set of points in I_1, $\{\mathbf{y}_i\}$ is the set of closest points in I_2 with associated normals $\{\boldsymbol{\eta}_i\}$, and $\hat{\mathbf{a}}$ is the set of estimated transformation parameters, then the accuracy measure is

$$\sigma^2 = \sum_i w_i[(\mathbf{T}(\mathbf{x}_i; \hat{\mathbf{a}}) - \mathbf{y}_i) \cdot \boldsymbol{\eta}_i]^2 \, / \, \sum_i w_i, \quad (2)$$

for robust weights w_i. This measure comes directly from the ICP process. For a correct transformation estimate, the accuracy must be close to the approximate noise of the sensor. This requirement is not sufficient, however. Alignment of repetitive structures and accidental alignment of subsets of the scene can make incorrect alignments seem accurate, whereas changes in the scene can make correct alignments seem inaccurate.

The visibility measure is designed to address some of these problems. This measure gives the fraction of points that appear to be incorrect when I_1 is mapped onto I_2. If the scene is static, then the mapping of a range point $\mathbf{x}$ from I_1 into the coordinate system of I_2 should not occlude any points from I_2 from the perspective of the scanner — otherwise the location in the scene corresponding to $\mathbf{x}$ would have been imaged in scan I_2. In theory, only scene changes and object motion between the acquisition of the two scans can cause a visibility violation in a correct alignment.

Assume the range scan is represented as a depth image $I_2(u, v)$. Each point $\mathbf{x}_i$ from I_1 is mapped into I_2. If the mapped point is within the field of view of the scanner, let u_i, v_i be its image coordinates in I_2 and z_i be its depth value. If $z_i < I_2(u_i, v_i) - c\sigma$, where $c\sigma$ is a small constant multiplier on the mean-square error, then we say that point i violates the visibility constraint. The visibility error measure from I_1 to I_2 is the fraction of the mapped points within the field of view of I_2 that violate the visibility constraint. A second visibility measure is computed by reversing the roles of I_1 and I_2 in the foregoing. The final visibility measure is the maximum of these measures.

When there are changes in the scene, low overlap, or repetitive structures, even the visibility measure may not be enough to determine if an alignment is correct. A third measure is needed that evaluates the hypothesized transformation to determine if it represents a random alignment or a misalignment of a repetitive structure.

The non-randomness measure places keypoints in a central role. Each keypoint represents a somewhat distinctive structure in the image. The search for keypoint matches is in effect a search over a variety of possible transformations for the keypoint. Hence, if the transformation maps the keypoint onto its best match, then this is local evidence that the transformation is unique and non-random [4]. Trans-
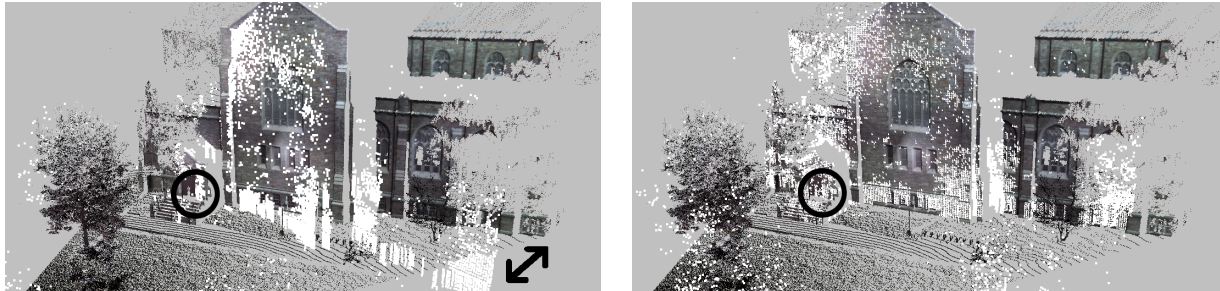
Figure 4. Refining a transformation based on a single keypoint match. The pair is the left two scans from Figure 1. Only 1% of the moving scan and 11% of the fixed scan are in the overlap region. Left: an initial alignment based on a keypoint in the region circled in black. Points in white are from the moving scan. Note the substantial misalignment indicated by the black arrow. Right: the refined transformation estimate. The misalignment is corrected by our refinement procedure producing a final robust error of 4mm.

formations consistent with many keypoint matches are extremely unlikely to be wrong. An important challenge to using this observation, however, is that many original keypoints may not be consistently detected in the two scans, especially when the viewpoints differ substantially. To address this, keypoint matches that are inconsistent with their best initial match are rematched using the estimated transformation. If the new match is better than the initial match for this keypoint, then this is additional evidence that the transformation is non-random and therefore correct.

To be more precise, let $\mathcal{M} = \{(\mathbf{p}_i, \mathbf{q}_i)\}$ be the set of originally matched keypoints. Exactly one of these keypoints was used to initialize the transformation being tested. For each $\mathbf{p}_i$, let $\mathbf{p}'_i = \mathbf{T}(\mathbf{p}_i; \hat{\mathbf{a}})$ be the mapping of $\mathbf{p}_i$ based on the estimated transformation. For intensity keypoints this keypoint mapping includes a projection from 3D to 2D based on camera calibration parameters. We call $(\mathbf{p}_i, \mathbf{p}'_i)$ a "transform match". Correspondences such that $\mathbf{p}'_i$ is not visible in the fixed scan are removed. Let $\mathcal{M}'$ be the resulting reduced set of matches. For each $\mathbf{p}_i$, if $\|\mathbf{p}'_i - \mathbf{q}_i\|$ is less than a small threshold that depends on the scale of the original keypoint, then the original match is considered consistent with the transformation. Otherwise, we compute the similarity score $S(\mathbf{p}_i, \mathbf{p}'_i)$ for the transform match using the scale from $\mathbf{p}_i$ for $\mathbf{p}'_i$. If $S(\mathbf{p}_i, \mathbf{p}'_i)/S(\mathbf{p}_i, \mathbf{q}_i)$ passes the ratio test, then the transform match is considered consistent; it is significantly better than all original matches for $\mathbf{p}_i$. The fraction of matches from $\mathcal{M}'$ that are either originally consistent or transform consistent is the used as the "randomness measure".

We therefore have three decision (verification) measures: robust mean-square error, visibility, and randomness. Empirically, we set thresholds on each and require a transformation to pass all three to be considered verified as correct. We require the alignment error to be within 6 times the scanner error, the visibility violations to be at most 20%, and at least 70% of the initially matched keypoints visible in both scans to be consistent with the final transformation. More sophisticated techniques could (and should) be developed, such as in [11], especially when there could be substantial environmental changes between scans.

7. Experiments and Discussion

We evaluate the three components of our location recognition system — keypoint detection and matching, refinement, and decision making — and validate an overall system that combines them. For the tests we run here, the algorithm attempts to align scans in pairs. (Generalization to more than two scans is conceptually straight-forward by indexing descriptors to simultaneously recover overlapping pairs and the corresponding pairwise transformations, but this is beyond the scope of our current experiments.) For each image, range and intensity keypoints are both extracted. For each pair, matches between keypoints are generated and rank-ordered — by spin-image comparison for range keypoints and by the ratio test for intensity keypoints. Up to 50 range keypoints and 200 intensity keypoint matches are retained. Each match is used to generate a hypothesis, which is then run through the refinement and verification procedures. While we did this here for the purposes of the experimental evaluation, in practice we use a greedy approach, where the matches are tested one-by-one and the procedure ends as soon as a verified transformation is found. One goal of the experiments is to show that this approach is viable.

We have acquired 22 scans from various locations on the Rensselaer campus. Of these, 15 distinct pairs overlap, but 5 of them overlap by an extremely small amount (e.g. see Figure 1). We therefore consider that we have 10 "reasonable pairs" and 5 "challenging pairs". We have manually
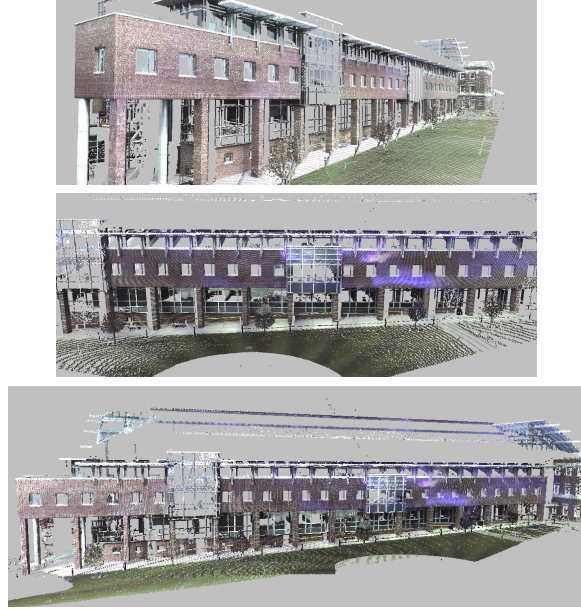
Figure 5. Two scans (top and center) for which range keypoint matching succeeded (bottom), but intensity keypoint matching failed. Range keypoints succeeded despite the highly-repetitive structure of the building since only one correct match was needed.

verified transformations for each of the 15 pairs, and we use these as "ground-truth" for testing. We added another 15 pairs of scans that do not overlap to form an initial test suite of 30 scan pairs.

The overall summary result is that all 10 of the reasonable pairs and 1 of the 5 challenging pairs were correctly registered (Figure 4), and all 15 of the non-overlapping pairs were rejected (none of the transformations survived the verification test). By "correctly registered" we mean that at least one keypoint match was refined to an estimate extremely close to the ground truth. In only 2 cases among all tests did an incorrect match pass the verification test — these were both for the reasonable pairs and both were extremely subtle misalignments. The first involved scans having substantially different viewpoints and scales; the second involved a small vertical translation error. In both cases the error was within the scale of the intensity keypoints, so these matches were counted as correct. Interestingly, for the 4 challenging pairs that were unmatched, hand-selection of a single match followed by refinement and verification produced a correct registration.

These preliminary results indicate the potential of the overall approach. More detailed results on the individual components of the system are summarized as follows:

- Intensity keypoint matching produced at least one verified match for 10 of the 11 aligned pairs, while range keypoint matchings produced at least one verified match for 6 pairs, but all of the ones for which at least one keypoint match existed. Range keypoint matching failed for the other pairs when there were no common building structures or the "holes" were occluded. The pair for which range keypoints succeeded, but intensity keypoints failed is shown in Figure 5. There was a wide baseline between scans and the windows are highly specular, causing the complete failure of intensity keypoints. Range keypoints were able to succeed despite the redundancy of the holes.

- Using the ground-truth, keypoint matches can be labeled as approximately correct. For intensity keypoints among the 10 reasonable pairs, there was a wide variation in the fraction of correct matches, ranging from 29 out of 30 to 1 out of 30. Even worse, for the one extremely challenging pair, there was only one correct match in the top 200 and it was ranked 161. Similar though less extreme results were obtained for range keypoints, with some scan pairs having 40% correct and some less than 5%.

- Among the approximately correct initial matches, 76% were refined to produce transformations that are consistent with the ground-truth (L_2-norm distance within a small multiple of the sensor noise). The region growing aspect of the refinement was generally effective, especially for low overlap (see Figure 4), but sometimes failed when initialized in large planar regions. Although more work is needed to improve refinement, these results show that initializing from a single match is a viable approach: a high fraction of the correct single matches are verified, and when this verification fails the algorithm is likely to pick and verify a different correct match.

- For the randomness component of the verification step, regeneration of keys has a varying impact. In many cases, the improvement was marginal, adding one or two correct matches, whereas in others it was dramatic, e.g. increasing the fraction from 27% to 56%.

- In verification, using the randomness test (with intensity keypoints) together the visibility and the ICP error tests produces 2 false positives (the two discussed above) and 5 false negatives. The same results are obtained with randomness and ICP error alone. ICP error and visibility produces 4 false positives and 2 false negatives. We expect the advantage of using the randomness test to increase when more changes occur in the environment occur between scans.

8. Conclusion

We have described a general version of the problem of registering multiple range scans as a location recognition problem. This approach includes novel techniques for keypoint detection and description, hypothesis generation from single keypoint matches, refinement using a combination of robust ICP and region growing, and a verification procedure that combines measures of accuracy, visibility and randomness. On a preliminary data set taken of the Rensselaer campus, the algorithm has successfully registered all but the lowest overlap scan pairs, and rejected all but the most subtly incorrect transformation estimates.

We conclude by offering three primary observations about how to solve the local recognition (and registration) problem based on our experiments:

- Both range and intensity keypoints are important. Range keypoints based on holes and spin images work extremely well on data sets involving buildings, even for varying magnifications and baselines. Intensity keypoints are effective (a) for scans involving significant occlusions because they are more local and (b) for scans involving natural scenes because they are more generic. Intensity keypoints are less effective for wide baselines and scenes with significant specularities.

- There is no need to cluster matching results. A greedy approach that starts from a single keypoint match and uses a combination of refinement and verification procedures along the lines we've proposed will be effective. When there are many correct matches, the approach will quickly find one and produce a correct transformation estimate. When there are only a few correct matches, clustering is not likely to be effective.

- While significant tuning still must be done on our current refinement and verification procedures, hypothesis generation needs the most future work. In the most difficult examples we've tested, a correct initial correspondence usually exists, but it ranks very low initially. When this keypoint is "pulled out", the refinement and verification procedures usually succeed. We intend to focus on both range keypoints alone and combinations of range and intensity keypoints. Handling low overlaps, occlusions and varying structures will remain our primary focus in this work.

9. Acknowledgements

References

[1] R. Bergevin, M. Soucy, H. Gagnon, and D. Laurendeau. Towards a general multiview registration technique. *IEEE T. Pattern Anal.*, 18(5):540–547, 1996.

[2] P. Besl and N. McKay. A method for registration of 3-d shapes. *IEEE T. Pattern Anal.*, 14(2):239–256, 1992.

[3] P. J. Besl and R. C. Jain. Segmentation through variable-order surface fitting. *IEEE T. Pattern Anal.*, 10:167–192, 1988.

[4] M. Brown and D. Lowe. Recognising panoramas. In *Proc. ICCV*, 2003.

[5] R. J. Campbell and P. J. Flynn. A survey of free-form object representation and recognition techniques. *CVIU*, 81:166–210, Feb. 2001.

[6] Y. Chen and G. Medioni. Object modeling by registration of multiple range images. *IVC*, 10(3):145–155, 1992.

[7] A. Frome, D. Huber, R. Kolurri, T. Buelow, and J. Malik. Recognizing objects in range data using regional point descriptors. In *Proc. Eigth ECCV*, 2004.

[8] C. Früh and A. Zakhor. An automated method for large-scale, ground-based city model acquisition. *IJCV*, 60(1):5–24, Oct. 2004.

[9] P. W. Holland and R. E. Welsch. Robust regression using iteratively reweighted least-squares. *Commun. Statist.-Theor. Meth.*, A6:813–827, 1977.

[10] A. Hoover, G. Jean-Baptiste, X. Y. Jiang, P. J. Flynn, H. Bunke, D. B. Goldgof, K. W. Bowyer, D. W. Eggert, A. W. Fitzgibbon, and R. B. Fisher. An experimental comparison of range image segmentation algorithms. *IEEE T. Pattern Anal.*, 18(7):673–689, July 1996.

[11] D. F. Huber and M. Hebert. 3d modeling using a statistical sensor model and stochastic search. In *Proc. CVPR*, 2003.

[12] D. F. Huber and M. Hebert. Fully automatic registration of multiple 3d data sets. *IVC*, 21:637–650, 2003.

[13] L. Ikemoto, N. Gelfand, and M. Levoy. A hierarchical method for aligning warped meshes. In *Proc. Fourth Int. Conf. on 3DIM*, pages 434–441, 2003.

[14] A. Johnson and M. Hebert. Surface matching for object recognition in complex 3-dimensional scenes. *IVC*, 16(9-10):635–651, July 1998.

[15] A. Johnson and S. Kang. Registration and integration of textured 3-d data. *IVC*, 17(2):135–147, 1999.

[16] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, November 2004.

[17] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. In *Proc. CVPR*, pages II:257–263, 2003.

[18] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *IJCV*, 60(1):63–86, 2004.

[19] G. Roth. Registering two overlapping range images. In *Proc. 2nd Int. Conf. on 3DIM*, pages 191–200, 1999.

[20] S. Rusinkiewicz and M. Levoy. Efficient variants of the ICP algorithm. In *Proc. Third Int. Conf. on 3DIM*, pages 224–231, 2001.

[21] I. Stamos and P. K. Allen. Geometry and texture recovery of scenes of large scale. *CVIU*, 88:94–118, 2002.

[22] I. Stamos and M. Leordeanu. Automated feature-based range registration of urban scenes of large scale. In *Proc. CVPR*, pages II:555–561, 2003.

[23] J. V. Wyngaerd and L. Van Gool. Combining texture and shape for automatic crude patch registration. In *Proc. Fourth Int. Conf. on 3DIM*, pages 179–186, 2003.