Available online at www.sciencedirect.com
ScienceDirect
Computer Vision and Image Understanding 110 (2008) 226–244
Computer Vision and Image Understanding
www.elsevier.com/locate/cviu

Registration of combined range–intensity scans: Initialization through verification

Eric R. Smith *, Bradford J. King, Charles V. Stewart, Richard J. Radke

Rensselaer Polytechnic Institute, Troy, NY 12180, USA

Received 17 April 2007; accepted 15 August 2007
Available online 29 September 2007

Abstract

This paper presents an automatic registration system for aligning combined range–intensity scan pairs. The overall approach is designed to handle several challenges including extensive structural changes, large viewpoint differences, repetitive structure, illumination differences, and flat regions. The technique is split into three stages: initialization, refinement, and verification. During initialization, intensity keypoints are backprojected into the scans and matched to form candidate transformations, each based on a single match. We explore methods of improving this image-based matching using the range data. For refinement, we extend the Dual-Bootstrap ICP algorithm for alignment of range data and introduce novel geometric constraints formed by backprojected image-based edgel features. The verification stage determines if a refined transformation is correct. We treat verification as a classification problem based on accuracy, stability, and a novel boundary alignment measure. Experiments with 14 scan pairs show both the overall effectiveness of the algorithm and the importance of its component techniques.

Keywords: Range registration; Iterative closest point; Keypoint; Decision criteria; Physical changes

1. Introduction

This paper addresses the problem of automatically computing a three-dimensional rigid transformation that aligns two range data sets, a problem that frequently arises in three-dimensional (3D) modeling of large-scale environments and in structural change detection. We present a robust algorithm that can accurately estimate and verify this transformation even in the presence of widely-differing scanning viewpoints (leading to low overlap between scans) and substantial structural changes in the environment between scans. Fig. 1 illustrates two examples of challenging scans successfully registered using our framework. The first example shows the alignment of two scans of a building taken from substantially different viewpoints, with trees occluding much of the common structure. The second example shows two scans of a parking lot and a building taken four hours apart, with different vehicles appearing between the two scans. Later in the paper, we demonstrate the alignment of a scan taken inside a room with one taken through a doorway looking into the room, as well as an alignment of two scans of a building where most of the common structure is repetitive (Fig. 9).

We assume that a range scanner with an associated, calibrated camera acquires the data sets, producing point measurements in 3D and essentially-simultaneous intensity images. Many current range scanners have this data acquisition capability, and it can be added to older scanners. Our registration framework effectively exploits the information available in the range data, the intensity images, and the relationship between the two. It requires neither external measurements such as GPS coordinates nor manual intervention.

Our fully automatic approach to registration involves three distinct stages—initialization, estimation, and verifica-

* Corresponding author.
E-mail addresses: smithe4@cs.rpi.edu (E.R. Smith), kingb2@cs.rpi.edu (B.J. King), stewart@cs.rpi.edu (C.V. Stewart), rjradke@ecse.rpi.edu (R.J. Radke).

1077-3142/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2007.08.004

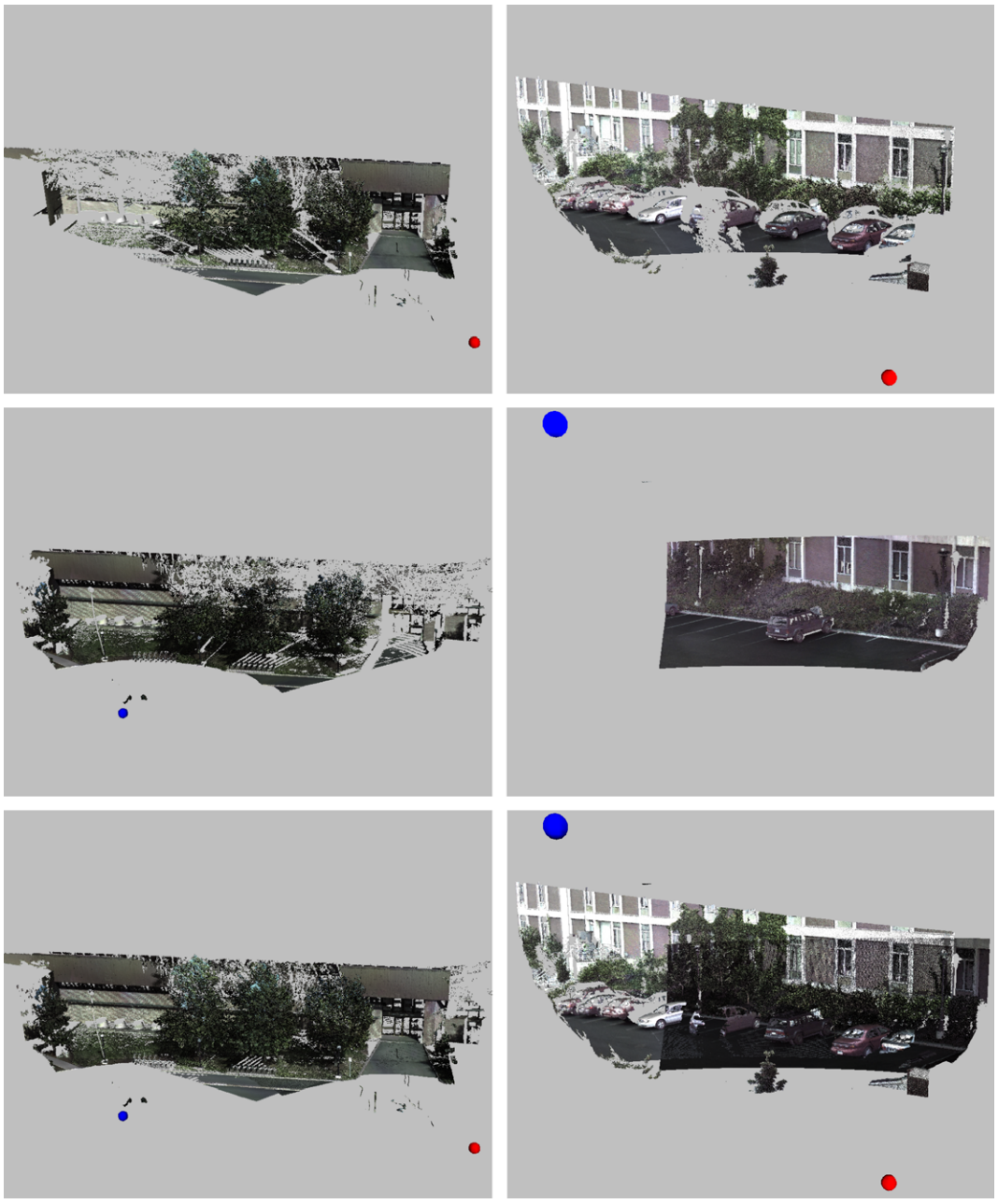

Fig. 1. Two example challenging pairs that our algorithm registers (DCC and MRC Parking lot). The top and center of the left column show a pair of scans taken from vastly different viewpoints, in which trees occlude much of the common structure. The top and center of the right column show two scans of a parking lot and building where all of the vehicles are different between the two scans. In both cases, as shown on the bottom, the algorithm automatically and accurately aligned the scans. The spheres in the figures indicate the scanner's locations.

tion—exploiting the combination of range and intensity information at each stage. During initialization, intensity keypoints and their associated SIFT descriptors [39] are extracted from the images and backprojected onto the range data. A three-dimensional coordinate system is established for each keypoint using its backprojected intensity gradient direction and its locally computed range surface normal. Keypoints are then matched using their SIFT descriptors. We explore several techniques that use the range data to improve the success of matching intensity keypoints.

Since each keypoint has an associated 3D coordinate system, each match provides an initial rigid transformation estimate. This eliminates the need for a RANSAC-style search for minimal subsets of matches in order to generate transformation estimates. Instead, the keypoint matches are rank-ordered based on a distinctiveness measure, and each is then tested using the estimation and refinement procedures. As we will show, this approach is particularly effective for some of the most challenging scan pairs because they have very few correct keypoint matches.

During the estimation stage, initial estimates are refined using a robust form of the Iterative Closest Point (ICP) algorithm [6,14], starting with points taken only from the region surrounding the match in the two scans. The region

is gradually expanded to eventually cover the overlap between the data sets. Region growth is controlled by the uncertainty in the parameter estimate generated by ICP, with more certain estimates leading to faster growth. The basic structure is adapted from the Dual-Bootstrap algorithm previously proposed for 2D image registration [63,69], but incorporates several innovations to robustly solve the 3D rigid registration problem. For example, ICP correspondences are generated between range points and also between backprojected intensity features. The latter produce constraints tangential to the range surfaces, complementing the surface normal distance constraints generated from the range data. As we show experimentally, the growth and refinement procedure nearly always converges to an accurate alignment given a reasonably accurate initial keypoint match. This shows further why a RANSAC-style search is unnecessary.

The final part of the algorithm is a verification test to decide whether a refined transformation estimate is correct. This test uses a simple, yet effective, linear classifier to combine measurements of alignment accuracy for both image and range features, a measure of the stability of the estimate, and a measure of the consistency in the position of range boundaries. These measures, especially the boundary measure, are more sophisticated than have been used in the past [32], but are necessary for effective decisions when the scans involve physical changes, simple geometries, and repetitive structures. Moreover, their effectiveness enables a simple, greedy method of processing the rank-ordered initial keypoint matches. That is, the matches are tested one-by-one; as soon as a candidate estimate passes the verification test, the result is accepted as a "correct" alignment. If all of the top M matches are rejected, the algorithm stops and reports that the scans cannot be aligned.

The overall algorithm is summarized in Fig. 2 and described in detail in the remainder of the paper. Section 2 outlines related research. Section 3 discusses the range/image data and the preprocessing of this data. Section 4 presents the initialization technique based on detection

(1) Preprocess the range scans to filter outliers, compute surface normals, and estimate scales (Sec. 3.1).

(2) Preprocess the intensity images to detect intensity keypoints and edgel features (Sec. 3.2), and then backproject both onto the range surfaces (Sec. 3.3).

(3) For each scan and its associated images,
   (a) Filter out keypoints that overlap depth discontinuities (Sec. 4.1).
   (b) Compute SIFT descriptors for each keypoint using the surface orientation of the backprojected keypoint to reorient the image neighborhood over which the descriptor is computed, making it viewpoint invariant (Sec. 4.1).
   (c) Filter keypoints in the moving scan to ensure an even spatial distribution of keypoints, while favoring higher strength keypoints.

(4) Match the keypoint descriptors between scans, compute a distinctiveness measure for each match, and order the resulting matches by distinctiveness (Sec. 4.2)

(5) For $m = 1$ to M do
   (a) Use keypoint match m to compute the parameters, $\hat{\mathbf{\Theta}}_0$, of an initial 3D rigid transformation and establish a small 3D region, Ω_0, centered on the matched keypoint location from the moving scan. (Sec. 4.2)
   (b) $t = 0$
   (c) Repeat
      (i) Apply robust ICP using a sampling of both range points and backprojected intensity features in Ω_t in the moving scan (Sec. 5.1). ICP produces a new vector of estimated transformation parameters, $\hat{\mathbf{\Theta}}_{t+1}$, and an associated covariance matrix $\mathbf{\Sigma}_{\hat{\mathbf{\Theta}}_{t+1}}$.
      (ii) Compute an expanded bootstrap region Ω_{t+1} (Sec 5.4).
      (iii) $t = t + 1$
   (d) Until Ω_t covers the apparent overlap volume between the fixed and moving scans.
   (e) Run ICP to convergence, producing an estimate $\hat{\mathbf{\Theta}}_m$.
   (f) Compute each of the measures in the decision criteria, and combine using a linear discriminant. If $\hat{\mathbf{\Theta}}_m$ passes, terminate with a successful alignment (Sec. 6).

(6) If all M initializations are rejected by the decision criteria, indicate that the scans cannot be aligned and terminate.

Fig. 2. Outline of the complete registration algorithm.

and matching of augmented keypoints. Section 5 describes the region-growing refinement algorithm based on both range correspondences and backprojected intensity correspondences. Section 6 formulates the verification criterion. The overall effectiveness of the algorithm, together with a detailed analysis of the individual components, is presented in Section 7. Section 8 discusses the results, summarizes our contributions, and concludes the paper.

2. Related work

The problem we address is a variation on the range scan registration problem that has been studied extensively over the past several decades [12,55,56], with solutions used in contexts as diverse as object modeling [4,7], the study of architecture [1], digitization of cultural artifacts [5,38], and industrial inspection [48]. Work on range image registration can be roughly divided into techniques for (1) initial registration or coarse alignment of a pair of scans, (2) fine registration of scan pairs, and (3) simultaneous registration of multiple scans given the results of pairwise scan registration. While the latter is an important problem [4,7,32,51,59], we focus here on the problems of initialization and refinement for a scan pair, while adding the substantial-but-not-well-studied issue of determining when the scans are well-aligned or can even be aligned at all.

A wide variety of techniques have been proposed for initial registration of range scans. All are based on extracting and matching summary descriptions of the scans. Some methods use extraction and matching of distinctive locations, such as points of high curvature [15] or other extremal locations [66]. Other feature-based methods have been developed for specific contexts, for example exploiting the presence of windows and building corners in the registration of architectural scans [13,61]. A second set of approaches, mostly used for aligning scans of single, smooth objects, includes methods that extract and match summary descriptions over small regions of the data. These include point signatures [16], "splashes" [62], and the more recent integral point descriptors [23]. A third set is comprised of methods that summarize larger regions or even the entire scan as the basis for registration and 3D object recognition. This includes work on extended Gaussian images (EGIs) [31] and spherical attribute images [29], where shapes are described using a mapping of a scan onto a sphere based on estimated surface normals. This work has mostly been superseded by spin images [33,35], extensions of the shape-context work [2] to 3D [21], and tensor-based methods [43]. Recently, Makadia et al. [40] resurrected the use of EGIs by reducing the EGI to a "constellation" of peak locations on the Gaussian sphere and matching these constellations between scans. This matching leads to the generation of several relative pose hypotheses between scans, each of which is then tested. This theme of hypothesis generation and testing is echoed in our approach.

When calibrated intensity images are available in addition to range scans, intensity features may be combined with range information to develop different initialization methods. One approach is to augment large region descriptors such as spin images with intensity data, producing a hybrid descriptor [11]. A more common approach is to build on the rapid progress that has been made recently in the matching of intensity keypoints between images taken over large changes in viewpoint [46,45]. Proposed keypoint detection techniques have been based on the Laplacian-of-Gaussian operator [39], information theory [36], Harris corners [44], and intensity region stability measures [41]. Following detection, summary descriptions of keypoints are constructed as the basis for matching. While a number of methods have been proposed [2,20,25,39], the SIFT descriptor [39], a spatial histogram of intensity gradients, has proved the most effective [45]. Several authors have proposed combinations of keypoint descriptor methods with range data for the initialization problem. Roth [53] used a combinatorial search to match keypoints that have been backprojected into 3D, but did not exploit the intensity information in the matching process. Wyngaerd and Van Gool [67] extracted and matched combined intensity and range feature vectors, computed at the intersection of a sphere with the range surface. Bendels et al. [3] matched 2D SIFT features, backprojected them onto the range data, and then used RANSAC on these points in 3D to identify an initial rigid transformation and filter incorrect matches. This does not exploit 3D information during the computation of the SIFT descriptor, as has been shown in [57] as well as our own earlier work [37]. In particular, Seo et al. [57] used 3D geometry to eliminate keypoints in non-planar regions and to rectify SIFT regions prior to computing the descriptor. We explore similar techniques here and also analyze in detail the effectiveness of this method.

We now turn to the refinement stage of registering a pair of range scans. In the literature, refinement usually depends on application of the well-known iterative closest point (ICP) algorithm [6,14,42]. ICP starts from an initial transformation estimate, and uses this to map points from one scan (the "moving" scan) onto the second scan (the "fixed" scan) in order to establish temporary correspondences between points in the two scans. These correspondences are then used to refine the estimate. The process iterates until a convergence criterion is met. The simplest and most common basis for establishing correspondence is the original idea of minimizing geometric distance between points, although a variety of shape properties has been exploited as well [58]. Distance metrics used in the estimation step of ICP have focused on the Euclidean distance and point-to-plane distance ("normal distance") measures, but more recently, quadratic approximations [47] have been proposed. Efficient variations of ICP have been studied extensively [55]. Earlier work [17,70] studied robustness to missing components of scans due to differences in viewpoints. The matching and refinement loops of ICP were cast into a Expectation Maximization (EM) framework in [26].

Genetic algorithms have been applied to improve the domain of convergence of ICP [10]. Convergence properties of ICP were studied recently by Pottmann et al. [50]. This work clearly shows the advantage of using normal distances instead of Euclidean distances, and suggests even faster convergence with higher-order distance approximations. A version of normal-distance ICP that is robust to mismatches and missing structures forms the core of our refinement algorithm.

Several researchers have combined intensity information with ICP. In one approach, color is treated as an additional geometric attribute to be combined with position in the determination of correspondences in ICP matching [19,34]. An alternative approach uses intensity and other geometric attributes to filter closest point matches [24,49]. Projection-based methods have been proposed as well, using image-based optical flow for matching [68] and, more recently, using color and depth differences in 2D coordinate systems to drive 3D registration [52]. Overall, two difficulties arise when combining range and intensity information to drive ICP that are especially significant in our particular problem domain: (1) intensity/color information and range data have different metric properties, and (2) intensities can change substantially between scans. We address both of these problems below through a novel technique that uses geometric constraints derived from matching backprojected image features.

The final concern is the problem of determining if two scans are actually well-aligned. This issue has received much less attention than initialization and refinement, in large part because of the limited domain in which fully automatic registration has been applied. Silva, Bellon, and Boyer proposed an ''interpenetration measure'' that checks the fine alignment of two surfaces [60]. The most extensive work is that of Huber and Hebert [32], whose algorithm tests both alignment accuracy and the fraction of mismatches that do not violate visibility constraints. In our context, visibility constraints must be carefully defined to accommodate for the possibility of changes between scans. Our approach does not exploit visibility directly, but instead looks at measures of accuracy, stability, and the alignment of visible depth discontinuities. Our decision measures substantially extend our earlier work both for range image registration [37] and intensity image registration [69]. Finally, in registration and mosaic construction based on keypoint matching, combinatorial analysis of keypoint match count provides a probabilistic measure of correctness [8]. Such an approach does not work here due to the low number of correct keypoint matches in difficult scan pairs.

3. Preprocessing and feature extraction

This section discusses Steps 1 and 2 of our algorithm (Fig. 2), the preprocessing of the range data and the intensity images, including backprojection of intensity features onto surfaces approximated from the range data.

We denote the set of points associated with a scan as $\mathcal{P}$, and the individual point locations as $\mathbf{P}_i \in \mathcal{P}$. As mentioned above, we assume that a range scanner with an associated, calibrated camera acquires the data sets. Since all points in a scan are acquired from a single viewpoint, the points can be easily mapped into a spherical coordinate system and binned in two dimensions using the spherical angles. One use of this binning structure is to find the nearest scan point $\mathbf{P}_i$ to any backprojected image point, which is straightforward because the image and range scanner viewpoints nearly coincide. A second use is to accelerate matching through computation of approximate closest points as in [55].

3.1. Range scan preprocessing

The scan points in 3D are preprocessed to eliminate outliers, to estimate a surface normal, $\boldsymbol{\eta}_i$, associated with each point, and to estimate the uncertainty in feature position along this normal in real units, represented as a scale (standard deviation) σ_i. The normal estimation algorithm uses a robust least-squares approach applied over a small, adaptive neighborhood that avoids rounding of sharp edges. The values of the scale estimates depend on both the noise in the scanner (2–4 mm for our scanner) and the degree to which the surface varies from planarity. The latter is particularly important because the intersample spacing can be relatively high (e.g., 2–5 cm) and surfaces in outdoor scenes can be quite rough. Thus, the estimated scale values tend to vary across the data set and can be as high as 1–2 cm. This consideration is important because even for perfectly-aligned scans, ''correct'' correspondences may be separated by up to half the sample spacing on the surface; therefore, interpreting any measure of distance between corresponding points must consider this locally-estimated scale.

The final step in preprocessing the range scans is to locate non-trivial depth discontinuities. Adjacent discontinuity locations in the scan grid are linked to form chains; the very short chains and smaller discontinuities are discarded. In outdoor scans, these discarded points usually correspond to boundaries in trees or grass, which are unlikely to be stable in subsequent scans. The remaining boundaries are used in the verification tests, as discussed in Section 6.

3.2. Keypoint and intensity feature detection

Next, keypoints are located in the intensity images. These will be used as the basis for initialization of the transformation estimate. We also locate intensity features, which will be used in ICP refinement of the transformation estimate.

We define *keypoints* as scale-space peak locations in the magnitude of the Laplacian-of-Gaussian operator, as in the well-known SIFT algorithm [39]. We use techniques similar to those in [9] to ensure widespread feature distributions. In particular, the minimum spacing between keypoints is set

to be linearly proportional to the product of the field of view angles of the scanner. This ensures a sufficient number of keypoints for small fields of view without generating an overly large number of keypoints for larger fields of view.

At the end of the keypoint computation, each keypoint has an associated location $\mathbf{u}_j$, gradient direction $\hat{\mathbf{g}}_j$, and detection scale s_j, all measured in image coordinates.

The *intensity features* are sampled edge elements (edgels)—image locations having a significant peak in the gradient magnitude along the gradient direction. The detection technique is similar to the edge detection method proposed in [27]. Adaptive thresholding is used to ensure a widespread distribution of features. Just as with keypoints, each feature has an associated normal direction $\hat{\mathbf{n}}_j$, location $\mathbf{u}_j$, measured to subpixel accuracy along $\hat{\mathbf{n}}_j$, and detection scale s_j.

3.3. Backprojection

Both keypoints and features are backprojected into 3D as illustrated in Fig. 3. The line of sight associated with image point $\mathbf{u}_j$ is found and the closest scan point $\mathbf{C}_j$ is found to this line of sight. A plane is then formed in the scene by a point, $\mathbf{P}_j$, that is the mean of the range samples inside the volume bounded by the backprojection of a circle of radius $2s_j$ about the keypoint center on the image plane, and the robustly-estimated normal, $\boldsymbol{\eta}_j$, as calculated in the preprocessing step at $\mathbf{C}_j$. The intersection of the line of sight with the plane determines the backprojected location, $\mathbf{U}_j$. The backprojection of $\hat{\mathbf{g}}_j$ is the unique direction vector, $\boldsymbol{\gamma}_j$, that is normal to $\boldsymbol{\eta}_j$ and projects into the image at $\mathbf{u}_j$ parallel to $\hat{\mathbf{g}}_j$. A 3D coordinate system is established at $\mathbf{U}_j$ by making $\boldsymbol{\gamma}_j$ the x axis, $\boldsymbol{\eta}_j$ the z axis and $\boldsymbol{\eta}_j \times \boldsymbol{\gamma}_j$ the y axis. Finally, a scale S_j is computed from s_j by backprojecting the point $\mathbf{u}_j + s_j\hat{\mathbf{g}}_j$ onto the plane and computing the distance of the resulting point to $\mathbf{U}_j$. This is the physical analog to the detection scale in the image.

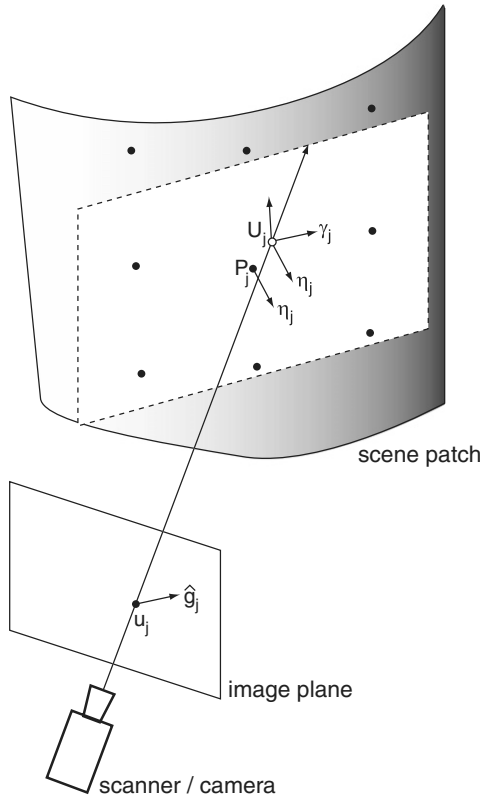

Fig. 3. Keypoints and associated image gradients are backprojected to form a 3D coordinate frame around each backprojected point.

4. Initialization based on keypoint matching

This section describes the procedure for generating a sequence of initial transformation estimates—Steps 3, 4, and 5a of the algorithm. In doing so, we investigate ways to augment keypoint matching with 3D information, building on work in [37,57].

4.1. Keypoint filtering and descriptor computation

Step 3a is a filtering of keypoints that are near depth discontinuities. The image regions surrounding these keypoints incorporate information from surfaces at different depths and therefore change discontinuously with changes in viewpoint. The filtering process is a simple planarity check [57], implemented by comparing the variance of the points used to calculate $\mathbf{P}_j$ with respect to the plane formed in the scene. If the region is rough, highly curved, or incorporates multiple surfaces, the variance of the plane from the local region will be much higher than the variance of the range data, σ_i^2, and the keypoint will be eliminated. Fig. 4 illustrates an example result.

In Step 3b our goal is to create a descriptor that is invariant to changes in viewpoint. Our strategy is to com-

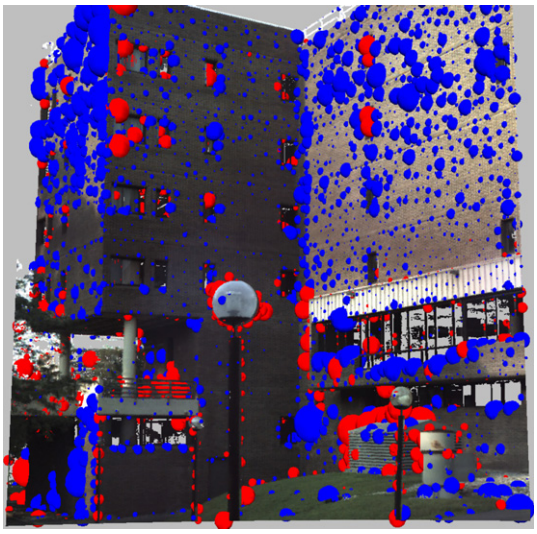

Fig. 4. Illustrating the effect of discontinuity filtering on keypoints. This figure shows the range data from the point of view of the scanner, where the red spheres are the filtered keypoints and the blue spheres are the remaining keypoints. The sizes of the spheres are proportional to the scales of the backprojected keypoints.

pute a SIFT descriptor [39] using the planar surface centered at the backprojected keypoint (see Section 3.3) to create an affine re-mapping of the image region. Algorithms that work with intensity images alone use the image data itself to infer such an affine mapping [46].

Recall that the SIFT descriptor is formed as a 4×4 spatial grid of orientation histograms, with 8 orientation bins per histogram. Concatenating the histograms and normalizing gives the 128-component descriptor vector. In our technique, the SIFT 4×4 square grid is aligned with the x and y axes of the keypoint's backprojected coordinate system, and the size of the grid is determined by the backprojected detection scale. Using an affine approximation to the mapping between this tangent plane and the image plane, the boundaries of this grid are mapped into the image plane, and the image locations and gradients falling within these boundaries are mapped back onto the surface. The backprojected locations are used to assign the points to the 4×4 SIFT grids and the affine-mapped gradient vectors are entered into the 8-bin orientation histogram associated with each grid block it maps to. After all gradients have been entered into the histograms, each histogram entry is divided by the amount of interpolated weight that its grid block received to normalize for sampling differences across the spatial bins. The 128-component descriptor is then extracted, and it is normalized in the same way as in [39]. Modulo sampling and other imaging artifacts, the resulting descriptor should be invariant to image scaling and to rotation and translation in 3D. This is a similar approach to Seo et al. [57]; however, we use only affine approximation and do not resample the actual intensities. The process is illustrated in Fig. 5.

4.2. Matching and initialization

The keypoints extracted from two scans are matched based on the SIFT descriptors of the keypoints as in [39]. The backprojected scales of the keypoints are approximate physical scales and therefore, in theory, should be invariant to the position of the scanner. A keypoint match is only allowed if the ratio of the larger to the smaller scale is less than a constant. (This type of comparison is possible because the constraints are based on back-projected and therefore physical scale measurements.) We have experimentally determined an effective value of this constant to be 2.0, reflecting the uncertainty in the estimates of scale due to sampling and the lack of affine correction of the scale estimate.

For each keypoint from one scan, among the matches allowed by the scale ratio test, the two keypoints whose descriptor vectors are closest in Euclidean norm are located. We define the distinctiveness measure of the best match as the ratio of the best to the second best descriptor distance [39].

The keypoint matches are ordered by increasing value of the distinctiveness measure and tested in this order in the next step. For each match tested, simply computing the transformation between the backprojected coordinate systems of the two keypoints generates an initial 3D rigid transformation. The initial region, Ω_0, starts as a cube aligned with the axes of the moving scanner's coordinate system, with side length 8 times the maximum of the point spacing and the backprojected pixel spacing on the surface. This initial region is expanded if necessary to contain at least 100 range points.

5. Refinement

The goal of refinement is to substantially improve the initial keypoint-based alignment between scans. This is done iteratively by alternating steps of (a) re-estimating the transformation using only matches for points in the bootstrap region Ω_t and (b) expanding the bootstrap region. Several iterations of refinement are illustrated in Fig. 6. Aside from this region growth procedure, the most important novel contribution in this section is the

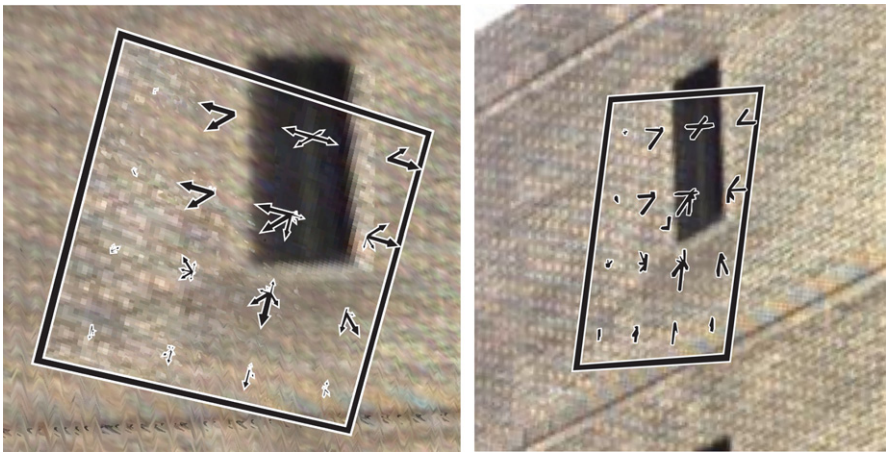

Fig. 5. Illustrating the computation of the augmented keypoint descriptor. The affine mapping from the surface tangent plane to the image plane determines a mapping between a square grid on the surface and an affine grid in the image. On the left, the augmented keypoint descriptor is shown on the range plane it is computed. The affine approximation is visible here as the descriptor region is not a perfect square. The right image shows the descriptor region projected into the image plane.

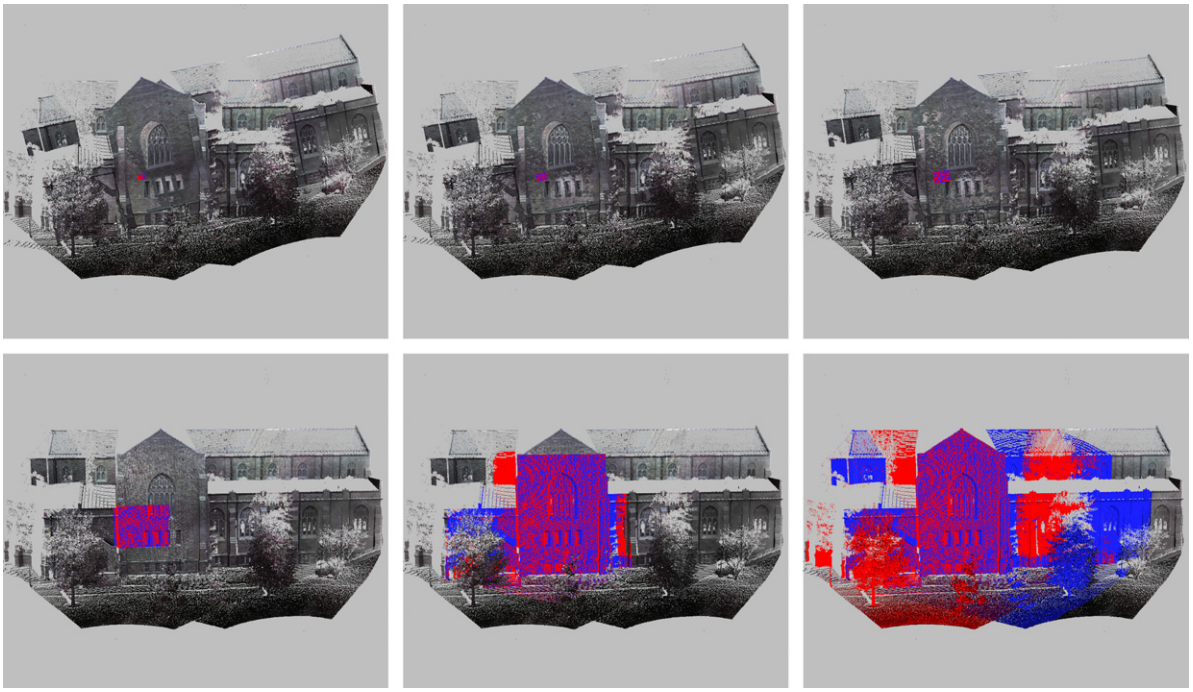

Fig. 6. Several iterations of ICP estimation and region growth of a relatively easy pair called "VCC North". The colored regions represent the area inside the current bootstrap region, Ω_t. The first frame shows the initial alignment of the scans and the initial region. Subsequent frames show the alignment and the region following iterations 1, 2, 5, 8, and 14, respectively. The correction of the strong initial misalignment can be clearly seen. Regions that are solidly one color or the other were only seen from the viewpoint of one scan.

combination of range scan correspondences and back-projected intensity feature correspondences in estimating the rigid transformation parameters.

5.1. Robust ICP

The core idea of ICP is well known. An initial transformation estimate is used to map points from the moving scan to the fixed scan. For each mapped point, the closest point in the fixed scan is located and these two points are used to form a temporary correspondence. The set of all such correspondences is used to re-estimate the parameters of the transformation. The whole process then iterates until some convergence criterion is met. In our refinement step, we do not wait for ICP to converge for each bootstrap region Ω_t, but instead stop and expand the region after two ICP iterations. After region growth has converged ICP continues to run until convergence.

Let $\widehat{\mathbf{\Theta}}$ be the current parameter estimate and, for moving point $\mathbf{P}_i$, let

$$\mathbf{P}'_i = \mathbf{T}(\mathbf{P}_i; \widehat{\mathbf{\Theta}})$$

be the mapped point. Next, let $\mathbf{Q}_j$ be the closest point, in Euclidean distance, to $\mathbf{P}'_i$. This forms the correspondence $(\mathbf{P}_i, \mathbf{Q}_j)$. Let $\boldsymbol{\eta}_j$ be the precomputed (unit) surface normal at $\mathbf{Q}_j$. The objective function for estimating the parameters from this set of correspondences is

$$F(\mathbf{\Theta}) = \sum_{(\mathbf{P}_i,\mathbf{Q}_j)} \rho([\mathbf{T}(\mathbf{P}_i;\mathbf{\Theta}) - \mathbf{Q}_j]^\top \boldsymbol{\eta}_j / \sigma_{i,j}), \tag{1}$$

where

$$[\mathbf{T}(\mathbf{P}_i;\mathbf{\Theta}) - \mathbf{Q}_j]^\top \boldsymbol{\eta}_j \tag{2}$$

is the "normal distance"—the distance from $\mathbf{T}(\mathbf{P}_i;\mathbf{\Theta})$ to the plane through $\mathbf{Q}_j$ with normal $\boldsymbol{\eta}_j$—and ρ is a robust loss function. We use the Cauchy ρ function,

$$\rho(u) = \frac{C^2}{2} \log\left(1 + \frac{u^2}{C^2}\right), \tag{3}$$

with $C = 2.5$ (consistent with typical values from the statistics literature [30]). Finally, $\sigma^2_{i,j}$ is the variance in the normal distance alignment error of the match. Estimation of $\sigma^2_{i,j}$ is discussed in detail below.

Minimization of $F(\mathbf{\Theta})$ to produce the next estimate $\widehat{\mathbf{\Theta}}$ is accomplished using iteratively reweighted least-squares (IRLS) [64], with weight function

$$w_{i,j} = \frac{1}{\sigma^2_{i,j}} \frac{\rho'(u_{i,j})}{u_{i,j}} = \frac{1}{\sigma^2_{i,j}(1 + u^2/C^2)} \tag{4}$$

where $u_{i,j} = [\mathbf{T}(\mathbf{P}_i;\mathbf{\Theta}) - \mathbf{Q}_j]^\top \boldsymbol{\eta}_j / \sigma_{i,j}$, and weighted least-squares formulation

$$F_w(\mathbf{\Theta}) = \sum_{(\mathbf{P}_i,\mathbf{Q}_j)} w_{i,j} [(\mathbf{T}(\mathbf{P}_i;\mathbf{\Theta}) - \mathbf{Q}_j)^\top \boldsymbol{\eta}_j]^2 \tag{5}$$

In each iteration of IRLS, the update to the parameter estimate is achieved using the small-angle approximation to

the rotation matrix. Since the incremental changes to the parameters tend to be small, this approximation works well. At the end of these two iterations the covariance matrix, Σ_{Θ}, of the transformation parameters is estimated from the inverse Hessian of (5). See [22,65] for more details.

5.2. Backprojected intensity constraints

The algorithm description thus far has not included constraints from matching backprojected intensity features. A novel step in our work is to use these as geometric constraints in the same way as correspondences from range point matching. Our goal is to create constraints on the 3D rigid transformation that complement the constraints from range matches. Since the latter are based on distances normal to surfaces, we design the intensity constraints to operate tangentially to surfaces (Fig. 7). The importance of this is most easily seen in the extreme case of aligning scans of a single, patterned, planar surface: range correspondences would only determine three degrees of freedom (DoF) of the transformation, leaving two translational and one rotational DoF in the plane. The intensity feature correspondences are designed to determine these three DoF.

As mentioned earlier, within the 2D image, each intensity feature has a location and a normal. These are backprojected onto the estimated surface in 3D, giving a location $\mathbf{U}_j$, and direction $\boldsymbol{\gamma}_j$. The latter is in the tangent plane of the surface. When this backprojected feature is correctly matched to a backprojected feature $\mathbf{V}_i$ from the moving scan and the transformation correctly aligns the points, then (ignoring noise) we should have

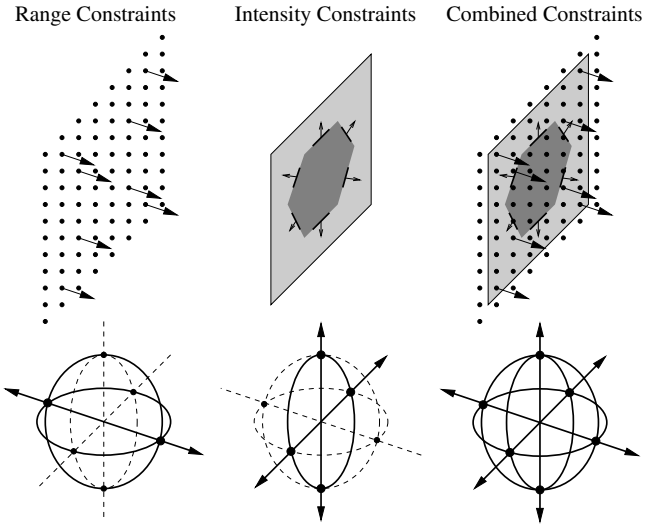

Fig. 7. Illustrating the complementary role of constraints from matching range points and from matching backprojected intensity features. Range matches provide constraints normal to the surface, while backprojected intensity feature matches provide constraints in the tangent plane of the surface. The top row shows the features providing constraints. The bottom row shows constraints placed on a rigid transform by each group of features. Circles represent rotational degrees of freedom and lines represent translational degrees of freedom. Solid circles and lines indicate those that are constrained.

$$[\mathbf{T}(\mathbf{V}_i;\boldsymbol{\Theta}) - \mathbf{U}_j]^{\top}\boldsymbol{\gamma}_j = 0. \tag{6}$$

Since this has the same form as the normal distance constraint (2), we can match backprojected image features and use them to constrain the estimate of transformation parameters in the same manner as range correspondences and their constraints. This makes implementation straightforward.

There are several important details to making this approach succeed. The first and the most important issue is estimating the variance $\sigma_{i,j}^2$ of the alignment error for both types of correspondences. This is used both to separate inliers from outliers and to assign relative strength to the two types of constraints—larger variances on one type of constraint lead to lower overall weight (4), even for inliers, and therefore less influence on the transformation parameter estimate. Estimation of variance is discussed in detail in the next section. Second, when sampling points in the current moving scan region Ω_t to form matches and thereby constraints on the transformation, the algorithm attempts to choose the same number of each type of feature. We rely on the computation of the variances and the resulting robust weights (4) to balance the relative influence of the two types of constraints. Finally, for intensity feature matches, we use the Beaton–Tukey biweight function

$$\rho(u) = \begin{cases} (1 - u^2/B^2)^3 & |u| < B \\ 0 & |u| \geqslant B \end{cases}, \tag{7}$$

as in [69], in place of the Cauchy loss function (3), with $B = 4.5$ (consistent with parameters developed in the statistics literature [30]). The Beaton–Tukey function more aggressively downweights potential outliers, which is more appropriate for image-based constraints where there tend to be more outliers—mostly due to viewpoint and illumination changes—and the alignment errors of these outliers tends to be only marginally higher than that of the inliers.

5.3. Correspondence variances

When estimating the rigid transformation parameters for a fixed set of matches, the variance in the alignment error, through its effect on the weight function (4), is used to eliminate the influence of outlier matches and to balance the influence of constraints from range point correspondences and from backprojected image feature correspondences. This prominent role of the variance marks a different approach to robustness than prior work on range registration [17,70], but it is consistent with traditional robust estimation methods [64] as well as our own prior work on registration [69]. The particular challenge faced here is that the variance must be treated as both unknown and varying from point to point. The latter follows from two observations: (1) as discussed in Section 3.1, due to intersample spacing, surface roughness and surface curvature, the variance in the locally-estimated planar surface approximation tends to change substantially from range

point to range point, and (2) the variance in the positions of backprojected intensity features tends to increase with distance from the scanner.

In order to make the problem of estimating the alignment error variances tractable, we make two simplifying assumptions. First, the alignment error variance is assumed to be proportional to the sum of the variances of the point positions along their normal direction. Thus, if σ_i^2 and σ_j^2 are the variances of the moving and fixed points, $\mathbf{P}_i$ and $\mathbf{Q}_j$, respectively, then the variance in the normal distance alignment error, $[\mathbf{T}(\mathbf{P}_i;\boldsymbol{\Theta})-\mathbf{Q}_j]^\top\boldsymbol{\eta}_j$, is modeled as

$$\sigma_{i,j}^2 = k_c^2(\sigma_i^2+\sigma_j^2), \tag{8}$$

with k_c as yet unknown. This assumption (which also depends on the point position errors in the two scans being independent of each other and primarily along the surface normal directions) is reasonably accurate at convergence, when the surfaces' normals are aligned and the primary factor in the remaining error is due to point position errors, rather than uncertainty in the parameter estimate. Moreover, since the parameter estimate tends to be quite accurate within the bootstrap region throughout the computation, the approximation holds during all iterations. Indeed, we have found experimentally that the estimated value of k_c stays quite close to 1.0 for range feature correspondences.

The second assumption is that the error in the true position of the image features in the image coordinate systems is proportional to the image smoothing scale at which they are detected. This ignores a number of properties of the imaging process and the imaged surfaces, focusing only on the well-known effect that smoothing has on edge-position uncertainty. We have found this assumption to be a reasonable first approximation both in past work [69] and in the algorithm described here. Given this assumption, the variance of a backprojected image feature at location $\mathbf{U}_i$ is $k_e^2S_i^2$ along the backprojected direction $\boldsymbol{\gamma}_i$ (which is in the surface tangent plane), where S_i is the backprojection of the smoothing scale for image feature i. Combining this with (8) yields the variance for correspondences between backprojected image features as

$$\sigma_{i,j}^2 = k_c^2k_e^2(S_i^2+S_j^2) = k_c'^2(S_i^2+S_j^2). \tag{9}$$

In the latter, we have combined the two unknowns into a single value k_c'. Thus, we have the same form for the variance in the normal-distance alignment error for both range point correspondences and backprojected intensity feature correspondences.

The Appendix describes how the scaling factor k_c is estimated from a given set of correspondences. Estimation of k_c' is identical, and we never explicitly estimate k_e.

5.4. Region growth

The region growth technique is a simple extension of the region growth technique for two-dimensional images used in the Dual-Bootstrap algorithm [63,69]. As mentioned earlier, the region is an axis-aligned rectangular solid in 3D, and all computation is done in a coordinate system centered on this solid. Growth occurs by considering the location $\mathbf{Y}$ in the center of each face of the solid, and computing the transfer error covariance [28, Chapter 4] of the mapping of this point. If the Jacobian of the mapping function is $\mathbf{J}=\partial\mathbf{T}/\partial\mathbf{Y}$, the transfer error covariance matrix of the mapped point $\mathbf{Y}'=\mathbf{T}(\mathbf{Y};\widehat{\boldsymbol{\Theta}})$ is

$$\boldsymbol{\Sigma}_{\mathbf{Y}'} = \mathbf{J}\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\Theta}}}\mathbf{J}^\top. \tag{10}$$

Here, the rotation component of the transformation is parameterized using the small angle approximation, making $\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\Theta}}}$ a 6×6 matrix, the Jacobian 3×6, and the resulting transfer error covariance, $\boldsymbol{\Sigma}_{\mathbf{Y}'}$, 3×3. The latter covariance is projected onto the outward face normal of the rotated Ω_t at $\mathbf{Y}'$. Growth is inversely proportional to the resulting scalar variance value. Each side is allowed to expand by up to 50% of its distance from the center, indicating that the solid can at most double in size in each dimension. This only occurs when there is a great deal of confidence in the estimate, and growth is typically slower earlier in the computation.

5.5. Iterations and convergence

Since normal-distance ICP does not provably converge, neither does our overall refinement algorithm. However, we have not encountered a problem that does not converge in practice. Our criterion for region growth convergence is that Ω_t has expanded sufficiently to cover the region of overlap (as computed from the estimated transformation) between the two scans. Our criterion for ICP convergence is that the mapping of points on the region boundary does not change substantially between ICP iterations. Region growth typically converges in about 10 iterations, and ICP converges almost immediately afterward. Since many of the earlier iterations occur with significantly fewer correspondences, overall convergence is quite fast (see Section 7.4).

6. Decision criteria

After the refinement procedure converges, the algorithm checks the result to determine if the alignment is sufficiently accurate and reliable to be considered correct. The challenge in doing so is handling potential structural/illumination changes as well as low scan overlap. Thus, measures such as Huber's visibility test [32], which assumes no changes between the scans, cannot be used. Our test incorporates seven measures: accurate alignment of both image feature and range correspondences; accurate alignment of normal directions for both image feature and range correspondences; stability of the transformation estimate; and two novel boundary alignment measures. These are all combined using a linear classifier to determine whether the registration is correct or incorrect.

6.1. Positional accuracy

Accuracy is the most natural measure. We use the normal-distance alignment error of the correspondences from the last iteration of refinement after region growth has converged. For range scan correspondences this is

$$\mathrm{PA}_R = \left(\sum_{(\mathbf{P}_i,\mathbf{Q}_j)} w_{i,j} [(\mathbf{T}(\mathbf{P}_i;\boldsymbol{\Theta}) - \mathbf{Q}_j)^\top \boldsymbol{\eta}_j]^2 k_c^2 (\sigma_i^2 + \sigma_j^2) \Big/ \sum_{(\mathbf{P}_i,\mathbf{Q}_j)} w_{i,j} \right)^{1/2}. \tag{11}$$

An identical measure, denoted by PA_I, is used for image feature correspondences.

6.2. Angular accuracy

The angular accuracy measure ensures that the feature directions—range surface normals and backprojected feature normals—are well-aligned, complementing the positional accuracy measure. This is most effective for testing image feature correspondences, since their normal directions vary much more quickly than the normals of smooth surfaces. For any correspondence $(\mathbf{P}_i, \mathbf{Q}_j)$ from the final ICP iteration, let $w_{i,j}$ be the final weight, let $\boldsymbol{\eta}_i'$ be the mapping of $\mathbf{P}_i$'s normal direction into the fixed image, and let $\alpha_{i,j}$ be the angle between $\boldsymbol{\eta}_i'$ and $\boldsymbol{\eta}_j$. Then the angular accuracy measure is

$$\mathrm{AA}_R = \sum_{(\mathbf{P}_i,\mathbf{Q}_j)} w_{i,j} |\alpha_{i,j}| \Big/ \sum_{(\mathbf{P}_i,\mathbf{Q}_j)} w_{i,j}. \tag{12}$$

The analogous measure AA_I is computed for image feature correspondences, except that the angles are mapped into $[0 \ldots \pi/2)$ prior to computing AA_I to account for the possibility of contrast reversals [69].

6.3. Stability

Incorrect transformations tend to be less stable than correct ones, especially when low inter-scan overlaps are allowed. Stability is measured using the transfer error covariance computation, which was already used for region growth in a previous stage (Section 5.4). A bounding box is placed in the moving scan around the points that have correspondences with non-zero weight. The trace of the transfer error covariance $\boldsymbol{\Sigma}_{\mathbf{Y}_i'}$ (Eq. (10)) is then maximized over this box, with the result being the stability measure, ST. The trace is used because it captures the uncertainty in any direction.

6.4. Boundary alignment check

While the four accuracy measures and the stability measure are strong indicators of correctness, they still produce false positives, most notably in scenes involving repetitive structure. A common example is aligning two scans of the face of a building that has regularly-spaced windows. An estimated alignment that produces a shift in the scans by the inter-window spacing looks very similar to a correct alignment. If we could assume that there were no changes in the scene between the two scans, then this misalignment could be caught by Huber's visibility test [32]. Since we are considering the potential for structural scene changes (along with small overlap and substantial viewpoint changes), a more sophisticated test is needed.

The test we propose is based on surface boundaries, detected as part of the preprocessing step (3.1). We do not take the natural step of establishing correspondences between boundary points in the two scans, in large part because viewpoint differences may cause a boundary (a depth discontinuity) in one scan to be missed in the other. Instead, each sampled boundary point, j, from the moving scan is tested as follows (see Fig. 8). Two boxes are formed: B_i on the surface interior, where samples occur, and B_e on the surface exterior, where no points are measured. If the two scans are properly aligned, then when B_i is mapped into the fixed scan, it should contain fixed-scan points, whereas when B_e is mapped into the fixed scan, it should be empty. Since there is one box per boundary point and the spacing between boundary points is at the sampling resolution of the scanner, upwards of several thousand boundary boxes are formed per scan, allowing for an aggregate computation of our boundary measures and tolerance to changes and transient objects that may affect the measure in a small number of boxes.

Here are details on the formation of surface interior box B_i and exterior box B_e for one boundary point. The axes of B_i are the estimated surface normal at the boundary, the boundary chain direction (lies on the estimated surface), and a third axis perpendicular to the other two. This con-

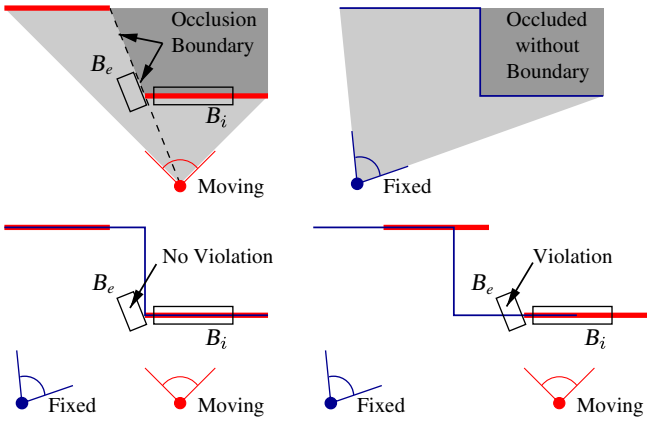

Fig. 8. Construction and use of the boundary test boxes. In the moving scan (top left), B_i, the interior box, is aligned with and centered on the surface near to but not overlapping the boundary, and B_e, the exterior box, is aligned with the line-of-sight direction and centered at the depth of the boundary point. When these boxes are mapped into the fixed scan (top right), B_i should contain surface points and B_e should be empty, even when the location corresponding to the boundary point is along an edge and the surfaces on both sides of the edge are visible. The bottom left and right images illustrate cases with no violation and violation of the boundary constraints, respectively.

struction ensures the inclusion of sample points. For B_e, one axis is parallel to the edge tangent, one is *parallel to the scanner's line of sight* at the boundary, and the third is perpendicular to the other two. By orienting B_e in this way, it should be empty in the fixed scan regardless of the type of boundary—a true surface discontinuity, a crease edge, or a location where the surface is tangent to the scanner's line of sight. The boxes are both wide enough (in the tangential directions) to accommodate sample spacing in the *fixed scan* and tall enough to account for noise (8).

Three different boundary point counts are formed: N is the total number of boundary points tested; N_i is the number of boundary points, j, such that j has at least a minimum number of fixed points in its interior box B_i after mapping into the fixed scan; N_{ie} is the number of boundary points, j, such that (a) j is counted in N_i and (b) j's exterior box B_e is also (nearly) empty in the fixed scan.[1] In other words, boundary points counted in N_{ie} have occupied interior boxes and empty exterior boxes in the fixed scan. The three counts are used to form two boundary measures for a given alignment: $BT_1 = N_{ie}/N_i$ and $BT_2 = N_{ie}/N$. Ideally, both scores will be near 1, but when there are substantial structural changes between scans, BT_2 may be low. On the other hand, BT_1 cannot be used alone because severe misalignments can cause it to be high and BT_2 to be very low. We let the linear classifier determine the trade-off between these measures (and the other decision measures) automatically.

6.5. Classification

For each rigid transformation parameter estimate produced by the refinement procedure, seven measures are computed from the parameters, the parameter covariance matrix, the matches and the boundary points—the two positional accuracy measures, PA_R and PA_I, the two orientation accuracy measures, AA_R and AA_I, the stability measure, ST, and the two boundary measures BT_1 and BT_2. These are each computed with neither re-matching nor re-estimation. The values are input into a linear classifier, which outputs a binary accept/ reject decision. This classifier is trained using the Ho-Kashyap algorithm [18] on the data described below.

7. Experimental evaluation

Our experiments evaluate both the overall effectiveness of the algorithm and the importance of the individual algorithm components. We collected a set of 14 test scan pairs exhibiting many challenging aspects of range scan registration, including three with substantial illumination differences, four with large changes in viewpoint, four with changes in structure, two having low overlap, one taken of an entirely planar scene, and three dominated by repetitive structures. Only three of the scan pairs, all taken outdoors, would be considered "easy". Two "hard" pairs and their resulting alignments are shown in Fig. 1, and four other "hard" pairs are shown in Fig. 9. The remaining eight pairs are shown in Fig. 10. The abbreviations indicated in the figures are used in the result tables below. Six of the 14 pairs were used for tuning the initialization and refinement parameters. Six separate pairs are used for the training involved with the decision criteria (Section 7.3). There were no substantial differences in the results for the training and test sets, so the results are lumped together for the sake of brevity.

The range scans were collected using a Leica HDS 3000 scanner. The scans in our experiments were typically acquired at 2–5 cm linear sample spacing on the surfaces. The RMS error of the measured points is 2–4 mm, independent of depth, on matte planar surfaces. The scanner acquires intensity images using a fixed, bore-sighted camera and a rotating set of mirrors to effect different camera orientations. The resulting images overlap to cover the field of view of the scanner. Our algorithms automatically partition the images to avoid redundancy in the intensity keypoints and features.

For each scan pair we are able to manually verify alignments. Any transformation estimated by our algorithm may then be compared to this manually-verified alignment by computing the RMS mapping error between the two transformations taken over the interscan region of overlap. When the RMS error is within a few standard deviations of the sensor noise, the estimated transformation is considered "correct".

As described in Fig. 2, the algorithm runs until either (a) one estimate is refined and verified by the decision criteria, or (b) M initial estimates have been tested (for our experiments $M = 50$). In the latter case, the scan pair is rejected. For the purposes of evaluating the performance of various stages of the algorithm, all of the top M initial estimates were refined and run through the verification procedure, as described below.

Overall, the algorithm aligned all 14 scan pairs correctly, and verified 13 of the 14 registrations as correct. The decision criteria were unable to verify the correct alignment of the BioTech scans. In the fixed scan the front face of the building is almost parallel to the line of sight, resulting in a sample spacing of nearly a half meter on the far part of the face. This led to overly high values in the accuracy and stability measurements.

7.1. Initialization and keypoint matching

The first round of experiments focuses on the effectiveness of keypoint matching, including three enhancements: (a) discontinuity filtering, (b) computing each keypoint descriptor on the backprojected surface plane rather than

[1] B_e will not always be completely empty, even for correct boundaries, because of boundary curvature and because scanners tend to produce points that interpolate between foreground and background surfaces at discontinuities.

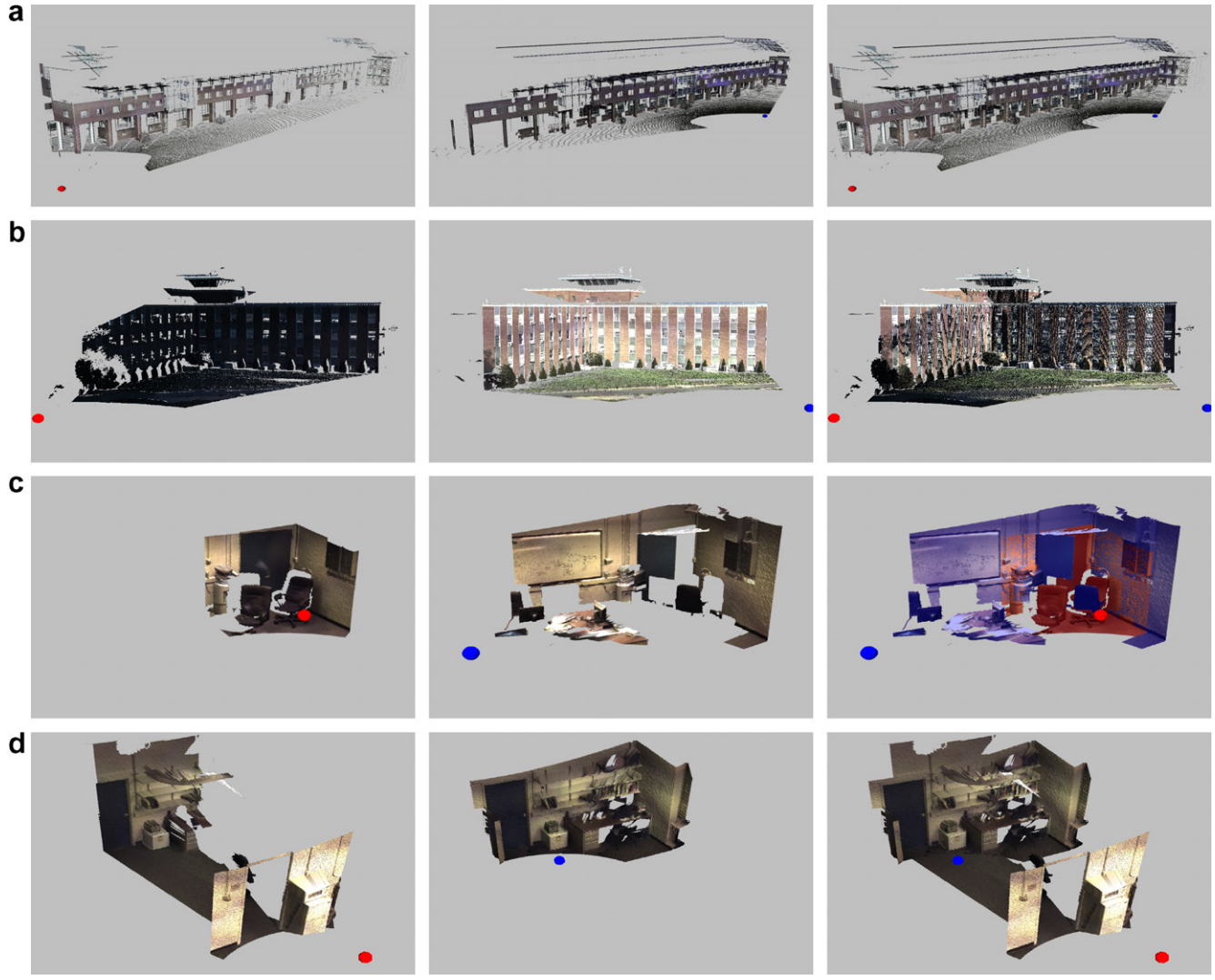

Fig. 9. These are the (a) BioTech, (b) Sci Center, (c) Lab Changes, and (d) M301 into M336 scan pairs. Scan pairs (a) and (b) have highly repetitive structure; the pair (b) is further complicated by illumination differences. Scan pair (c) demonstrates the algorithm's ability to register scans with many changes; the right-hand picture is tinted to show the changes between the scans. The only overlapping parts between the two original scans are the walls and the water cooler. Scan pair (d) demonstrates the registration of the inside of a room and the same room seen through a doorway.

in the original image, and (c) using physical scale in keypoint matching and filtering. Each of these techniques may be added or removed from the algorithm, so we can explore the effectiveness of each separately. In our evaluation, the top 50 matches are examined for each test pair and each is marked "correct" or "incorrect" based on applying the manually-verified transformation to the back-projected moving image keypoint and comparing to its matched, backprojected keypoint in the fixed scan. If the positions are within $4 \times \mathbf{S}_j$, and both the gradient directions and surface normals are within 15 degrees, the match is considered correct.

The results for all scan pairs are shown in Table 1. The scan pairs are ordered top-to-bottom, by increasing difficulty, with the last row providing the totals across all scan pairs. The columns of the table show all combinations of activating the enhancements, with the left showing all enhancements on and the right showing all enhancements off. The abbreviations used in the table are (a) "disc flt" for discontinuity filters, (b) "aff-dscr" for computing the affine-mapped descriptor on the planar surface of the scan, and (c) "phys scl" for using the physical scale in matching.

Several inferences may be drawn from the table:

- The overall percentage of correct matches on our data set increases by 33% using all of the enhancements, compared to using no enhancements. This is less than we expected overall.
- The number of correct matches varies dramatically between scans, reflecting our intuition that some pairs are quite easy, while others are very difficult.
- As can be seen by comparing the left four columns with the right four columns, discontinuity filtering has very little overall effect on the effectiveness of keypoint

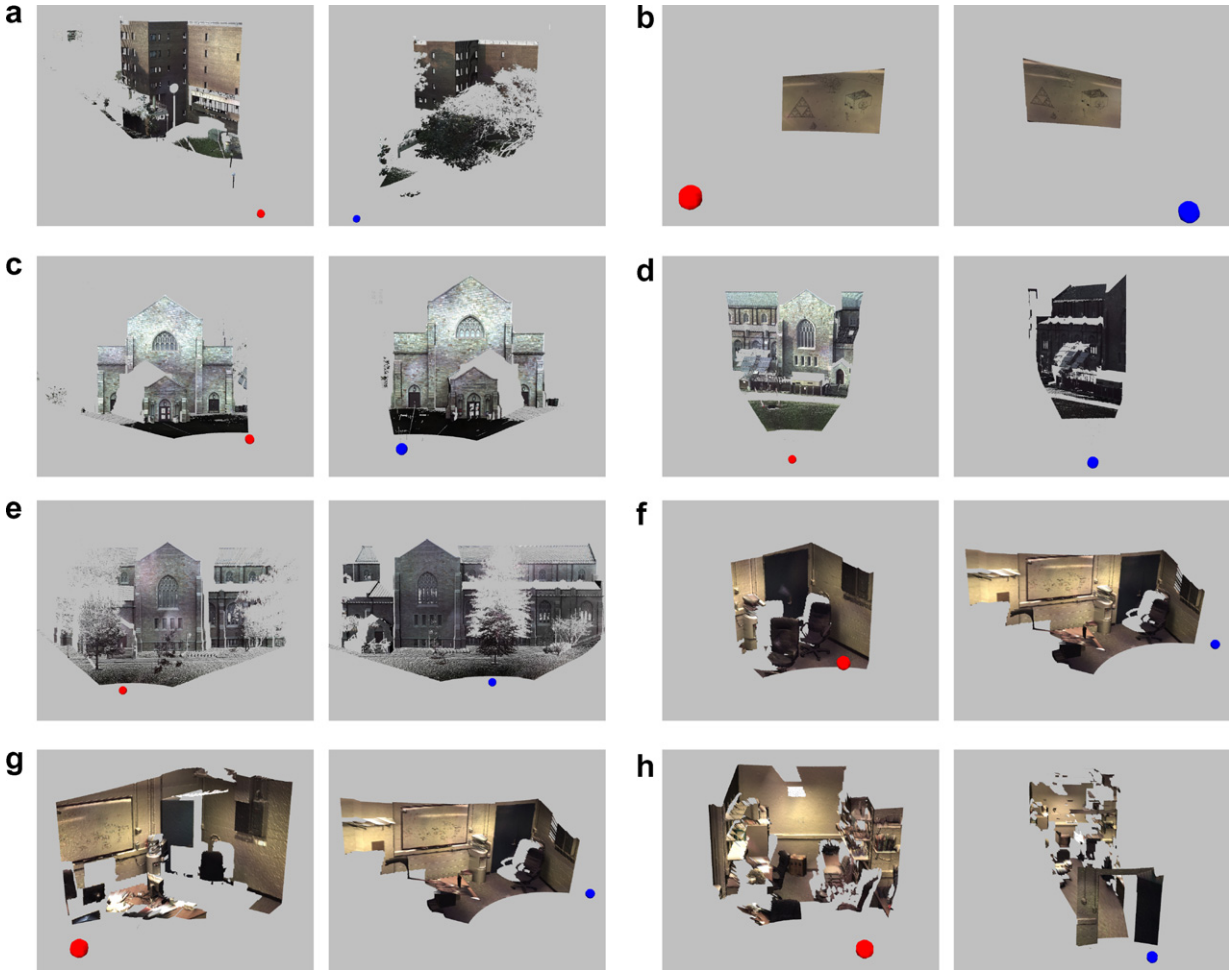

Fig. 10. (a) JEC—highly repetitive structure and large viewpoint difference. (b) Whiteboard—a flat scan that demonstrates the power of image-feature matching. (c) VCC West—easy pair, but has noisier images due to rain while scanning. (d) VCC South—illumination differences and smaller overlap. (e) VCC North—easy pair, has one main overlapping face that provides many good initializations. (f) Lab Regular—indoor scan with varying viewpoints and smaller overlap. (g) Lab Large—indoor scan with varying viewpoints. (h) M336 into M301—a difficult scan pair with a large sample spacing difference between scans and a doorway that blocks part of one scan's view.

matching. We attribute this to the simple observation that the descriptors for keypoints found along depth discontinuities change so much between views that they are unlikely to be matched anyway.

- While the general trend is an improvement in the number of matches when the other two enhancements are added, this is not monotonic for all scan pairs (e.g., "Lab Normal").
- Finally and most importantly, the enhancements do make an impact on the most difficult scan pairs shown at the bottom of the table. The enhancements produce enough keypoint matches to allow our algorithm to align these scans accurately, even when a random-sampling search [8,57] would fail due to an insufficient number and density of correct matches.

We have three possible explanations for why the improvements are not more substantial. First, when computing scale we do not adjust for affine distortions in the original image computation, as in [46]. We have avoided this expensive computation in favor of simpler enhancements. Second, since the ranking of the matches is based on *relative* descriptor distance, improvements in descriptor computation and associated reductions in descriptor distances do not necessarily lead to better rankings. Third, the quality of images taken by the scanner is not as high as ordinary digital images and varies substantially with illumination (often with substantial glare), suggesting that there is a limited range of potential improvements. Despite these concerns, the overall trend, especially for the difficult scan pairs where the enhancements matter most, is towards improved performance. As a result, we use all enhancements in both our remaining tests and in our complete algorithm.

7.2. Refinement

In evaluating the refinement algorithms, we focus on the effectiveness of (a) starting registration from a single keypoint match, (b) region growth during estimation and

Table 1
Number of correct keypoint matches among the first 50

	All enhancements	disc fltr, aff-dscr	disc fltr, phys scl	disc fltr	aff-dscr, phys scl	aff-dscr	phys scl	No enhancements
VCC West	46	40	42	35	46	40	44	36
Whiteboard	46	46	45	45	46	46	45	45
VCC North	42	34	38	27	47	34	37	27
VCC South	41	33	28	27	40	32	28	27
Lab Large	32	24	28	20	33	24	28	20
MRC Pk lot	28	28	22	22	25	24	15	17
Lab Normal	11	12	17	18	11	12	17	17
Lab Changes	13	13	11	14	15	13	11	14
M301 into M336	13	10	10	10	13	10	10	10
M336 into M301	5	3	7	3	5	3	7	3
BioTech	5	4	1	1	2	3	1	1
Sci Center	3	3	1	1	3	2	1	0
JEC	2	2	2	0	1	2	0	0
DCC	2	1	1	0	2	1	0	0
Totals	289	253	253	223	289	246	244	217

refinement, and (c) using image features in the registration process. In testing these, we ran the refinement process on all 50 initial keypoint matches.

The results are summarized in Table 2. The first column of results is the number of "correct" keypoint matches in the top 50 and is included as a basis for comparison. The second column, "no region grow" indicates that there is no region growth and no image feature correspondences are used. In other words, robust ICP is run scan-wide starting from the initial estimate generated from the single keypoint match. The third column indicates that image feature correspondences were added, but region growth was still not used. In the fourth and fifth columns, region growth was added.

As before, we make several important observations about these results:

- Clearly, the addition of both image features and region growth substantially improves the number of successful alignments, with region growth playing the more substantial role. More importantly, adding region growth allows the alignment of three of the most challenging pairs. However, we note that robust ICP initialized from a single match can accurately align the scans in a surprising number of cases.
- Using image feature correspondences is most helpful on scans involving substantial planar regions ("Whiteboard" and "Lab Changes"). In most other scans, the effect of image features is negligible. This is partially due to the fact that variations in surface orientation generate constraints in similar directions to the image-feature constraints, making the image features less necessary. Another reason is that the low image quality renders the image feature constraints less reliable.
- The total number of successful final alignments is *larger than the number of correct keypoint matches*. This means for many of the scan pairs nearly all initializations from these matches are refined to correct final estimates. Also, that some matches that are too far off to be considered good are actually refined to a correct final estimate. This demonstrates the power of using image features and region growth in the refinement stage.

Table 2
Number of correct alignments produced by the refinement stage when applied to the top 50 keypoint matches

	# kpt matches	no region grow	no region grow + img ft	region grow	region grow + img ft
VCC West	46	23	20	47	47
Whiteboard	46	18	36	12	38
VCC North	42	27	19	43	43
VCC South	41	11	21	42	42
Lab Large	32	29	16	32	31
MRC Pk lot	28	9	19	29	30
Lab Normal	11	11	15	19	17
Lab Changes	13	2	14	6	14
M301 into M336	13	16	15	19	18
M336 into M301	5	4	7	6	7
BioTech	5	0	0	1	2
Sci Center	3	0	0	3	3
JEC	2	2	2	2	2
DCC	2	0	0	2	3
Totals	289	152	184	263	297

An additional experiment was performed to determine how well the initialization and refinement stages perform as the amount of noise in the scans increases. We chose five scan pairs representing a range of difficulties, and added independent, normally-distributed noise to each measurement along its line of sight. We then ran the algorithm on the resulting scans. Fig. 11 plots the number of initial estimates in the top 50 that the algorithm refines to a correct transformation as the standard deviation of the added noise increases from 0 to 5 cm (recall that the noise in the scanner is 2 mm). The results clearly demonstrate that adding noise to the scans does not substantially affect the number of successful registrations. There is a noticeable downward trend for several of the scan pairs, mostly on the easier pairs. However, since the algorithm only needs a single success, the decrease in the easier scans is insignificant. Furthermore, the more difficult scans were not greatly affected by the noise, because their difficulty lies in the image based initialization.

7.3. Decision

A linear discriminant combining the seven decision measures was trained using six of the scan pairs and the top 50 keypoint match refinements, each compared to the manually-verified transformation. The results were then tested on all 14 pairs, again using the top 50 matches and refinements for all pairs as the basis for testing. Out of the resulting 700 alignments, there were 3 false negatives, no false positives, and 294 true positives. Unfortunately, as discussed above, two of the false negatives are on the BioTech scan and these are the only two alignments. Therefore, we cannot claim that this scan was aligned.

When one or more of the measures is removed and the linear discriminant is retrained, performance degrades. Results for these experiments are summarized in Table 3. The first four rows of results measure the performance of the basic building blocks of the decision criteria, including from top to bottom (1) using position accuracy, PA_R only, (2) using both range criteria, PA_R and AA_R, and the stability measure, ST, (3) using the image-feature-based criteria, PA_I and AA_I, and the stability measure, ST, and (4) just using the boundary measures, BT_1 and BT_2. Row (1) shows that using the alignment accuracy measure alone produces a significant number of false positives, while row (4) shows that the boundary measures alone have nearly equivalent performance to alignment accuracy. Both are insufficient overall. The remaining rows show, in order, (5) leaving out only the boundary measures, (6) leaving out only the image-feature measures, (7) leaving out only the range measures, and (8) using all measures. It is clear from these results that all criteria are needed, including both the image-feature measures and boundary measures. In addition, it is interesting to note that the image features play a significant overall role, as can be seen by comparing rows (6) and (8) of the results.

Table 3
Verification results of total number of false positives, false negatives, true positives, and true negatives over all scans in top 50

Measures used	False pos.	False neg.	True pos.	True neg.
PA_R	19	34	263	384
PA_R, AA_R, ST	2	106	191	401
PA_I, AA_I, ST	68	44	253	335
BT_1, BT_2	22	43	254	381
All but BT_1 and BT_2	4	8	289	399
All but PA_I, AA_I	8	7	290	395
All but PA_R and AA_R	28	7	290	375
All	0	3	294	403

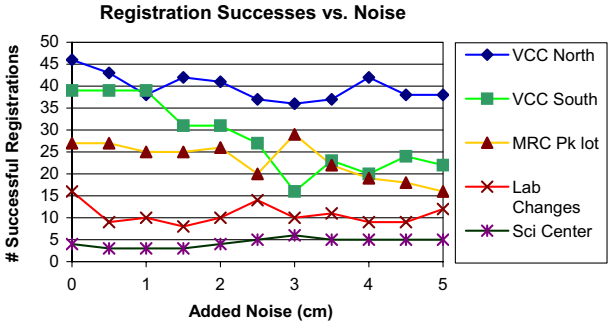

Fig. 11. Investigating the effect of additional noise on the performance of the algorithm. The number of registration successes in the top 50 initializations is plotted as a function of the standard deviation of the added noise variance for five representative scan pairs. The additional noise has little effect on the overall power of the algorithm because only a single success is needed per scan pair.

7.4. Performance

The time to register two preprocessed scan pairs is mainly dependent on the rank of the first initialization that is refined to a verified alignment. For our experiments, 9 of 14 scan pairs are correctly aligned and verified on the first ranked initialization. The other four first successes are on the 2nd, 6th, 8th, and 37th ranked initializations. It should be noted that many of the failure refinements before the first success are quickly terminated because of extremely poor initialization.

We evaluated the performance of the algorithm on our full data set using a PC with a 3.2 GHz Pentium 4 processor and 2 gigabytes of RAM. The sizes of the scans range from 13K to 1.2M points and 2K to 73K intensity features. The preprocessing step of the algorithm is currently the slowest, taking from 2.5 min up to 10.7 min per scan, largely because of large number of overlapping, redundant, and sometimes useless images produced by the scanner—we are currently working on dramatic speed improvements here. The keypoint matching and initialization calculation takes on average 26.4 s. The average running time of a single registration and verification over all scans is 11.5 s with an 8.7 s standard deviation. Finally, the median time from initialization until the verification detects a successful alignment is 8.7 s. Of the four scans whose first success is not on

the first keypoint match the total registration times are 1.9 s, 2.7 min, 1.5 min, and 10.2 min. Many steps of the algorithm can be trivially parallelized, which would decrease registration time on multiprocessor machines, especially when the first successful alignment is poorly ranked.

8. Conclusion

We have presented a three-stage algorithm for fully automated 3D rigid registration of combined range/intensity scans that is extremely effective in the presence of substantial viewpoint differences and structural changes between scans. The algorithm succeeds through the combination of (a) the matching of intensity keypoints that have been backprojected onto range scans, (b) the initialization of registration estimates from single keypoint matches, (c) an estimation and refinement procedure that combines robust ICP, region growth and a novel combination of intensity feature and range point correspondences, and (d) a decision criteria that includes measures of positional and orientation alignment accuracy, stability, and a novel boundary-based measure, all combined using a linear discriminant trained from a subset of the test data.

While the primary contributions of our work are the overall algorithm and its demonstrated effectiveness, we also have several more specific contributions, including (1) an experimental evaluation of the effectiveness of keypoint matching based on backprojected surface information, (2) adapting the Dual-Bootstrap approach—starting from preliminary alignments between small regions of the data—for range scan registration, (3) a novel method for combining constraints based on intensity edges and on range points during estimation, and (4) novel, sophisticated decision criteria for automatically determining when two scans are well-aligned despite structural changes.

In terms of the individual stages of our algorithm, we can conclude from our experiments that:

- Matching of backprojected keypoints which have been affine-corrected based on the 3D planar surface is an improvement over matching the keypoints based on image information alone. However, the improvements are not as dramatic as might be expected. The most important effect is in the most difficult scans where only a small number of correct matches are obtained. Further improvements may be possible using a better estimation of viewpoint-invariant scale.
- Region growth and the use of image-feature correspondences both play an important role in registration, especially on scans of near-planar surfaces and the scan pairs involving the most difficult viewpoint differences and structural changes.
- The combined decision criteria produced no false positives on our data set and only three false negatives out of 700 tests. While the accuracy measures are important, our novel boundary measures play a crucial role, especially on repetitive structures and scans involving structural changes.

As an additional note, we have shown the important role that image features play at all stages of the alignment process, including initialization, refinement and verification. In each stage, backprojected features contribute to the algorithm's success despite substantial viewpoint and/or illumination differences between scans.

Finally, as compared to other approaches, our experiments, including the addition or removal of specific features of our algorithm, have shown that (1) on particularly challenging scan pairs, using matching of backprojected keypoints alone [8,57], perhaps followed by robust ICP, is not enough to register the scans effectively, (2) just running a robust version of ICP [6,14,17,70], initialized using a transformation obtained from a keypoint match, does not converge to the correct solution reliably, and (3) more sophisticated decision criteria [32] than just using alignment error are indeed necessary.

The weakest point of our algorithm is initialization, since we have shown that once an initial estimate is obtained, region growth and re-estimation nearly always converge to a correct alignment. Therefore, one area of future investigation is new methods for initialization, most likely using range data in addition to image data [33,40]. Preliminary results along these lines were presented in [37]. Aside from initialization, the next stage of our research is extending the algorithm to multiscan registration and automatic detection of structural changes. Here the decision criteria will be even more important, as shown in [32] for static scenes, perhaps necessitating more sophisticated measures and non-linear classifiers.

Appendix A. Modeling variances in correspondences

Recall that the variance in a scan point correspondence is modeled as $\sigma_{i,j}^2 = k_c^2(\sigma_i^2 + \sigma_j^2)$, with σ_i^2 and σ_j^2 being the previously-estimated feature position variances and k_c being an unknown constant multiplier (Eq. (8)). We use the fact that k_c is the same for all correspondences to estimate its value for each set of correspondences and transformation parameters. For a correspondence between $\mathbf{P}_i$ and $\mathbf{Q}_j$, let $d_{i,j} = (\mathbf{T}(\mathbf{P}_i; \hat{\boldsymbol{\Theta}}) - \mathbf{Q}_j)^\top \boldsymbol{\eta}_j$ be the signed normal distance measured in the fixed scan's coordinate system. Suppose the values of $d_{i,j}$ are each normally-distributed, with variance $\sigma_{i,j}^2$. Although $\sigma_{i,j}^2$ varies across the matches, the values $d_{i,j}/\sqrt{\sigma_i^2 + \sigma_j^2}$ are i.i.d. with variance k_c^2. Since k_c is the only unknown, we can form the set of values $d_{i,j}/\sqrt{\sigma_i^2 + \sigma_j^2}$ and robustly compute the variance to estimate k_c. In the very first iteration of matching in the initial bootstrap region, the variance of $d_{i,j}/\sqrt{\sigma_i^2 + \sigma_j^2}$ is estimated

using the median absolute deviation scale estimator [54,64]. Subsequently, for the first match set for each region, Ω_t, after weights have been calculated, k_c is re-estimated as

$$k_c^2 = \frac{\sum_{(\mathbf{P}_i,\mathbf{Q}_j)} w_{i,j} d_{i,j}^2 / (\sigma_i^2 + \sigma_j^2)}{\sum_{(\mathbf{P}_i,\mathbf{Q}_i)} w_{i,j}}.$$

References

[1] P.K. Allen, A. Troccoli, B. Smith, S. Murray, I. Stamos, M. Leordeanu, New methods for digital modeling of historic sites, IEEE Comp. Graph. Appl. 23 (6) (2003) 32–41.

[2] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, IEEE Trans. Pattern Anal. Mach. Intell. 24 (4) (2002) 509–522.

[3] G.H. Bendels, P. Degener, R. Wahl, M. Körtgen, R. Klein, Image-based registration of 3d-range data using feature surface elements, in: Proceedings of the 5th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST (2004), 2004.

[4] R. Bergevin, M. Soucy, H. Gagnon, D. Laurendeau, Towards a general multiview registration technique, IEEE Trans. Pattern Anal. Mach. Intell. 18 (5) (1996) 540–547.

[5] F. Bernardini, I.M. Martin, H. Rushmeier, High-quality texture reconstruction from multiple scans, IEEE Trans. Visual. Comput. Graphics 7 (4) (2001) 318–322.

[6] P. Besl, N. McKay, A method for registration of 3-d shapes, IEEE Trans. Pattern Anal. Mach. Intell. 14 (2) (1992) 239–256.

[7] G. Blais, M. Levine, Registering multiview range data to create 3d computer objects, IEEE Trans. Pattern Anal. Mach. Intell. 17 (8) (1995) 820–824.

[8] M. Brown, D. Lowe, Recognising panoramas, in: Proceedings of ICCV, 2003.

[9] M. Brown, R. Szeliski, S. Winder, Multi-image matching using multi-scale oriented patches, in: Proceedings of CVPR, vol. 1, 2005, pp. 510–517.

[10] K. Brunnstrom, A. Stoddart, Genetic algorithms for free-form surface matching, in: ICPR96, vol. 4, 1996, pp. 689–693.

[11] N. Brusco, M. Andreetto, A. Giorgi, G.M. Cortelazzo, 3D registration by textured spin-images, in: Proceedings of the Fifth International Conference on 3DIM, 2005.

[12] R.J. Campbell, P.J. Flynn, A survey of free-form object representation and recognition techniques, Comput. Vis. Image Understand. 81 (2001) 166–210.

[13] C. Chao, I. Stamos, Semi-automatic range to range registration: a feature-based method, in: Proceedings of the Fifth International Conference on 3DIM, 2005.

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

[15] C. Chua, R. Jarvis, 3D free-form surface registration and object recognition, Int. J. Comput. Vis. 17 (1) (1996) 77–99.

[16] C. Chua, R. Jarvis, Point signatures: a new representation for 3d object recognition, Int. J. Comput. Vis. 25 (1997) 63–85.

[17] G. Dalley, P. Flynn, Pair-wise range image registration: a study in outlier classification, Comput. Vis. Image Understand. 87 (2002) 104–115.

[18] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, John Wiley and Sons, 2001.

[19] J. Feldmar, J. Declerck, G. Malandain, N. Ayache, Extension of the ICP algorithm to nonrigid intensity-based registration of 3d volumes, Comput. Vis. Image Understand. 66 (2) (1997) 193–206.

[20] W.T. Freeman, E.H. Adelson, The design and use of steerable filters, IEEE Trans. Pattern Anal. Mach. Intell. 13 (9) (1991) 891–906.

[21] A. Frome, D. Huber, R. Kolurri, T. Buelow, J. Malik, Recognizing objects in range data using regional point descriptors, in: Proceedings of the Eighth ECCV, 2004.

[22] N. Gelfand, L. Ikemoto, S. Rusinkiewicz, M. Levoy, Geometrically stable sampling for the ICP algorithm, in: Proceedings of the Fourth International Conference on 3DIM, 2003, pp. 260–267.

[23] N. Gelfand, N.J. Mitra, L.J. Guibas, H. Pottmann, Robust global registration, in: Eurographics Symposium on Geometry Processing, 2005.

[24] G. Godin, D. Laurendeau, R. Bergevin, A method for the registration of attributed range images, in: Proceedings of the Third International Conference on 3DIM, 2001, pp. 179–186.

[25] L.V. Gool, T. Moons, D. Ungureanu, Affine/photometric invariants for planar intensity patterns, in: Proceedings of the Fourth ECCV, 1996.

[26] S. Granger, X. Pennec, Multi-scale EM-ICP: a fast and robust approach for surface registration, in: Proceedings of the Seventh ECCV, 2002, pp. 418–432.

[27] C. Harris, M. Stephens, A combined corner and edge detector, in: Proceedings of The Fourth Alvey Vision Conference, Manchester, UK, 1988, pp. 147–151.

[28] R. Hartley, A. Zisserman, Multiple View Geometry, Cambridge University Press, 2000.

[29] K. Higuchi, M. Hebert, K. Ikeuchi, Building 3-d models from unregistered range images, GMIP 57 (4) (1995) 315–333.

[30] P.W. Holland, R.E. Welsch, Robust regression using iteratively reweighted least-squares, Commun. Statist. Theor. Meth. A6 (1977) 813–827.

[31] E. Horn, N. Kiryati, Toward optimal structured light patterns, IVC 17 (2) (1999) 87–97.

[32] D.F. Huber, M. Hebert, Fully automatic registration of multiple 3d data sets, IVC 21 (2003) 637–650.

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

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

[35] A.E. Johnson, M. Hebert, Using spin images for efficient object recognition in cluttered 3d scenes, IEEE Trans. Pattern Anal. Mach. Intell. 21 (5) (1999) 433–449.

[36] T. Kadir, A. Zisserman, M. Brady, An affine invariant salient region detector, in: Proceedings of the Eighth ECCV, 2004.

[37] B.J. King, T. Malisiewicz, C.V. Stewart, R.J. Radke, Registration of multiple range scans as a location recognition problem: hypothesis generation, refinement and verification, in: Proceedings of the Fifth International Conference on 3DIM, 2005, pp. 180–187.

[38] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, S. Anderson, J. Davis, J. Ginsberg, J. Shade, D. Fulk, The Digital Michelango Project: 3d scanning of large statues, in: SIGGRAPH, New Orleans, July 2000, pp. 131–144.

[39] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.

[40] A. Makadia, A. Patterson IV, K. Daniilidis, Fully automatic registration of 3d point clouds, in: Proceedings of CVPR, New York, NY, USA, 2006.

[41] J. Matas, O. Chum, M. Urban, T. Pajdla, Robust wide-baseline stereo from maximally stable extremal regions, IVC 22 (10) (2004) 761–767.

[42] C.-H. Menq, H.-T. Yau, G.-Y. Lai, Automated precision measurement of surface profile in CAD-directed inspection, IEEE Trans. Robotics Automation 8 (2) (1992) 268–278.

[43] A. Mian, M. Bennamoun, R. Owens, A novel representation and feature matching algorithm for automatic pairwise registration of range images, Int. J. Comput. Vis. 66 (1) (2006) 19–40.

[44] K. Mikolajczyk, C. Schmid, Scale and affine invariant interest point detectors, Int. J. Comput. Vis. 60 (1) (2004) 63–86.

[45] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, IEEE Trans. Pattern Anal. Mach. Intell. 27 (10) (2005) 1615–1630.

[46] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, L.V. Gool, A comparison of affine region detectors, Int. J. Comput. Vis. 65 (1–2) (2005) 43–72.

[47] N.J. Mitra, N. Gelfand, H. Pottmann, L. Guibas, Registration of point cloud data from a geometric optimization perspective, in: Eurographics Symposium on Geometry Processing, 2004.
[48] V.-D. Nguyen, V. Nzomigni, C. Stewart, Fast and robust registration of 3d surfaces using low curvature patches, in: Proceedings of the Second International Conference on 3DIM, 1999, pp. 201–208.
[49] I.S. Okatani, A. Sugimoto, Registration of range images that preserves local surface structures and color, in: Proceedings 3DPVT, 2004.
[50] H. Pottmann, Q.-X. Huang, Y.-L. Yang, S.-M. Hu, Geometry and convergence analysis of algorithms for registration of 3d shapes, Int. J. Comput. Vis. 67 (3) (2006) 277–296.
[51] K. Pulli, Multiview registration for large data sets, in: Proceedings of the Second International Conference on 3DIM, 1999, pp. 160–168.
[52] K. Pulli, S. Piiroinen, T. Duchamp, W. Stuetzle. Projective surface matching of colored 3d scans, in: Proceedings of the Fifth International Conference on 3DIM, 2005.
[53] G. Roth, Registering two overlapping range images, in: Proceedings of the Second International Conference on 3DIM, 1999, pp. 191–200.
[54] P.J. Rousseeuw, Least median of squares regression, J. Amer. Stat. Assoc. 79 (1984) 871–880.
[55] S. Rusinkiewicz, M. Levoy, Efficient variants of the ICP algorithm, in: Proceedings of the Third International Conference on 3DIM, 2001, pp. 224–231.
[56] J. Salvi, C. Matabosch, D. Fofi, J. Forest, A review of recent range image registration methods with accuracy evaluation, IVC 26 (2006).
[57] J.K. Seo, G.C. Sharp, S.W. Lee, Range data registration using photometric features, in: Proceedings of CVPR, 2005.
[58] G. Sharp, S. Lee, D. Wehe, ICP registration using invariant features, IEEE Trans. Pattern Anal. Mach. Intell. 24 (1) (2002) 90–102.
[59] G. Sharp, S. Lee, D. Wehe, Multiview registration of 3d scenes by minimizing error between coordinate frames, IEEE Trans. Pattern Anal. Mach. Intell. 26 (8) (2004) 1037–1050.
[60] L. Silva, O.R.P. Bellon, K.L. Boyer, Enhanced, robust genetic algorithms for multiview range image registration, in: 3DIM 2003, pp. 268–275,.
[61] I. Stamos, M. Leordeanu, Automated feature-based range registration of urban scenes of large scale, in: Proceedings of CVPR, vol. II, 2003, pp. 555–561.
[62] F. Stein, G. Medioni, Structural indexing: efficient 3-d object recognition. IEEE Trans. Pattern Anal. Mach. Intell. 14 (12) (1992) 1198–1204.
[63] C. Stewart, C.-L. Tsai, B. Roysam, The Dual-Bootstrap iterative closest point algorithm with application to retinal image registration, IEEE Trans. Med. Imag. 22 (11) (2003) 1379–1394.
[64] C.V. Stewart, Robust parameter estimation in computer vision, SIAM Rev. 41 (3) (1999) 513–537.
[65] C.V. Stewart, Uncertainty-driven, point-based image registration, in: Y.C. Nikos Paragios, O. Faugeras (Eds.), Handbook of Mathematical Models in Computer Vision, Springer, 2006, pp. 221–236.
[66] J. Vanden Wyngaerd, L. Van Gool, Automatic crude patch registration: toward automatic 3d model building, Comput. Vis. Image Understand. 87 (2002) 8–26.
[67] J. Vanden Wyngaerd, L. Van Gool, Combining texture and shape for automatic crude patch registration, in: Proceedings of the Fourth International Conference on 3DIM, 2003, pp. 179–186.
[68] S. Weik, Registration of 3-d partial surface models using luminance and depth information, in: Proceedings of International Conference on 3DIM, 1997, pp. 93–100.
[69] G. Yang, C.V. Stewart, M. Sofka, C.-L. Tsai, Registration of challenging image pairs: initialization, estimation, and decision, IEEE Trans. Pattern Anal. Mach. Intell. 29 (11) (2007) 1973–1989.
[70] Z. Zhang, Iterative point matching for registration of free-form curves and surfaces, Int. J. Comput. Vis. 13 (2) (1994) 119–152.