Physical Scale Intensity-Based Range Keypoints

Eric R. Smith, Richard J. Radke, and Charles V. Stewart*
Rensselaer Polytechnic Institute
110 Eighth Street, Troy, NY USA 12180
smithe4@cs.rpi.edu, rjradke@ecse.rpi.edu, stewart@cs.rpi.edu

Abstract

This paper presents a method for detecting, describing, and matching keypoints in combined range-intensity data. We extend a 2D image-based detection and description framework to 3D using an image back-projected onto a range scan. A key feature of the framework is a physical scale space for detecting keypoints, which eliminates errors in scale during both detection and matching. We develop smoothing, differentiation, and description techniques that are focused on making the keypoint invariant to viewpoint, sampling, and intensity changes. We demonstrate the power of our algorithm with comparisons to a back-projected SIFT algorithm, showing that it is able to find and match keypoints in a variety of challenging scan pairs.

1. Introduction

Identifying and matching 3D keypoints are integral steps in many algorithms that address range registration and 3D object detection. Regardless of its application, the quality measure of a keypoint algorithm is its repeatability. This means that a keypoint's location and description must be as invariant as possible to viewpoint, sampling, and intensity differences. In this paper, we combine information from calibrated cameras with range data to generate and match intensity keypoints in three dimensions, increasing keypoint repeatability and matchability. The top row of Figure 1 illustrates an example of a keypoint match pair generated by our approach. Even though the keypoints straddle a depth discontinuity, they are correctly matched by our algorithm. On the other hand, back-projection of a SIFT keypoint [10, 14] detected in 2D can easily generate incorrect matches due to scale mismatch and improper handling of depth discontinuities, as illustrated in the lower row of Figure 1.

There are currently two general classes of keypoint algorithms that work with 3D data; some are strictly point/geometry based [4, 7, 19] while others form descriptors based on functions defined on the geometry (e.g., geometric invariants or intensities from associated images) [1, 6, 14, 15, 16, 18, 20]. Our paper fits into the second class of such algorithms, by using images from a calibrated camera associated with the range scanner. Much of the work in the second class of algorithms is focused on datasets that are full-3D models (e.g., models constructed by combining data from multiple viewpoints [18, 16]). Our focus is on single viewpoint range data, typically acquired from large-scale outdoor scenes, which differs from typical full-3D models in two ways. First, the problem of depth discontinuities is much more apparent in single viewpoint data. In outdoor scenes both large-scale (e.g., ends of buildings) and small-scale (e.g., doors and windows) discontinuities are preva-

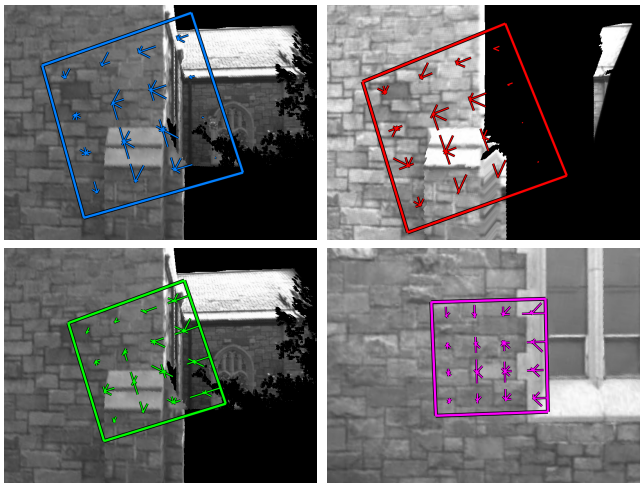

Figure 1. The top row shows a correct match between two automatically-detected physical scale keypoints from two different scans computed near a depth discontinuity. The bottom row shows a back-projected SIFT keypoint (green) computed at the same location, and its (incorrect) match in purple. The physical scale keypoints are exactly the same size; however, the purple keypoint is 70% smaller than its matched keypoint (green).

*This work was supported in part by the DARPA Computer Science Study Group under the award HR0011-07-1-0016.

lent. Second, a range scanner is able to produce a physical distance to each point in the scan, while models reconstructed from multiview stereo, for example, lack a defined scale. This means that the distance between structures in one scan can be reproduced in another scan that sees the same structures. These two aspects of single-scan data are important in the design of our algorithm.

The algorithm presented in this paper extends SIFT [10] by detecting keypoints from back-projected intensities on 3D range surfaces. We compute SIFT-like descriptors for the keypoints using scales and neighborhoods defined in the range scanner's physical 3D space. Our algorithm has three main advantages:

- Keypoints are described in a viewpoint invariant manner. The descriptors are invariant to projective foreshortening and favor contributions from neighboring values (typically from the same surface) that are more likely to be seen when the viewpoint changes. Our algorithm uses a novel application of a bilateral filter to accomplish this.
- Keypoints whose regions of support overlap depth discontinuities are robustly detected, described, and matched.
- Keypoints only match other keypoints at the same physical scale. This is in contrast to typical 2D keypoints that achieve scale invariance by allowing matching at different pixel scales.

To understand the significance of the third advantage, consider a simpler approach that uses the standard SIFT algorithm and then back-projects the keypoints and their descriptors onto the range data. After the back-projection, a pair of keypoints that should match will typically differ in the area they cover on the physical surface due to their detection using pixel scales. Because of this size difference, gradients will be entered at different relative locations on each descriptor; furthermore, some values may only appear in one descriptor. While these problems are somewhat mitigated in the design of the original SIFT algorithm through the use of a fine sampling in scale, and through partial volume interpolation and Gaussian weighting in the formation of the descriptor, the use of physical scales when range data are available offers a much stronger solution, allowing use of many fewer scales and matching only at the same scale (Section 3). We show that the physical scale space increases the repeatability of our detection and description techniques, yielding keypoints that are invariant to scale and have additional robustness to discontinuities.

This paper is organized as follows. Section 2 discusses related work. Section 3 describes the detection process, and Section 4 discusses the descriptor computation and matching. Section 5 presents experimental results on several large-scale datasets, and Section 6 offers concluding remarks.

2. Related Work

Shape descriptors that collect points into a histogram have been used extensively, most notably Johnson and Hebert's spin images [7]. Spin images collect points into a 2D histogram that is spun around a center point's normal. Mian et al. extended spin images to a tensor representation vector and developed a method for matching them [12]. Frome et al. [4] extended 2D shape contexts to 3D, using an oriented basis point to construct a 3D histogram. 3D shape context descriptors have a degree of freedom in rotation about their normal, which is addressed by creating additional descriptors at sampled rotations. Recently, Zhong [19] developed Intrinsic Shape Signatures, which histograms points in a spherical coordinate system around a basis point. The coordinate frame is based on the eigenvectors of the scatter matrix of the neighborhood. While geometry-only keypoints have been successful, the difficulty in establishing a highly repeatable coordinate frame leads these techniques to require more matches than combined 3D-intensity keypoints to estimate surface matches.

In 2D, scale space was studied thoroughly by Lindeberg [9] and used notably by Lowe [10] to detect scale invariant keypoints. 2D SIFT keypoints detected in image space and back-projected into 3D data have proven successful as well. Of the SIFT-like 3D keypoint algorithms, our previous work in range registration [14] used geometry to provide affine invariance by mapping gradients from the image onto a plane in the range data. We also used the 3D information to form filtering heuristics for keypoints and matches. Wu et al. [16] extended SIFT using a framework called VIP (viewpoint-invariant patches) by mapping the intensities onto a plane fit onto a surface in 3D and detecting and describing the keypoint there.

Assuming full, uniformly sampled meshes, Zaharescu et al. [18] extended Wu's viewpoint invariant patches to use full 3D gradients and descriptors. They built a scale space of intensities on a mesh using repeated convolutions of a fixed-width 3D Gaussian. Extrema are detected at one-ring neighborhoods of the difference-of-Gaussian meshes. Gradients are computed using weighted directional derivatives around a vertex, which are binned on the three axes of the keypoint to form a descriptor. While our approach shares similar concepts with this work, our focus is on single view range data.

A strictly geometric scale space was developed by Novatnack and Nishino [13], in which a mesh is embedded into 2D and smoothed in this space using geodesic Gaussians. Interest points are detected based on a normal map. Akagunduz and Ulusoy used a scale space of mean and Gaussian curvatures to detect interest points on parameter-

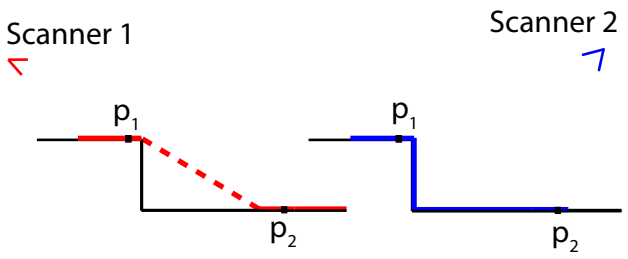

Figure 2. The red lines indicate the mesh formed by scanner 1, and the blue by scanner 2. Since scanner 1 did not see the corner, it believes the shortest path between p_1 and p_2 is over the dotted line, which disagrees with scanner 2's view. Because of this phenomenon we do not use geodesics as a distance measure.

ized 3D surfaces [1]. Other geodesic based scale spaces have been proposed. For example, Hua et al. [6] proposed a conformal mapping of a mesh, detected interest points on it using geodesic diffusion, and used 2D SIFT descriptors to characterize them. Zou et al. [20] developed an intrinsic geometric scale space using Ricci flow on 3D surfaces. Starck and Hilton proposed a geodesic intensity histogram for manifold matching [15]. In this paper, we do not use geodesics because they are not possible to properly define when discontinuities are present, as illustrated in Figure 2.

Our decision to use geometry-augmented photometric features instead of strictly geometric or strictly photometric features is largely due to the nature of the data. For many datasets that lack distinguishable local geometry, point-based keypoints such as spin images [7] often have to be quite large to be successful. This can be problematic in datasets with low surface overlap or many discontinuities. In these cases, intensity-based keypoints are more successful for local surface matching [8].

3. Keypoint Detection

We assume that a range scanner with an associated, calibrated camera acquires the datasets, producing point measurements in 3D and essentially-simultaneous intensity images. A unit normal vector for each range data point is computed using a prioritized iteratively-reweighted least-squares (IRLS) plane fitting algorithm. The first step in keypoint detection is to bring the intensities from the image's pixel space to the range scanner's physically measured space. Since the image sampling is generally denser than the range sampling, most image pixels do not have automatic range correspondences. Using the calibration of the camera, we map pixel locations from the image into the range scanner's 3D coordinate system, as illustrated in Figure 3. Consider an image pixel $\mathbf{x}_{im}$. We find the closest (within a threshold) range point to the 3D ray r formed by $\mathbf{x}_{im}$ and the camera origin; let this point be $\mathbf{x}_{rg}$. We compute the 3D location $\mathbf{p}$ corresponding to $\mathbf{x}_{im}$ as the intersection of the ray r with the plane described by $\mathbf{x}_{rg}$ and its normal. The location $\mathbf{p}$ inherits its normal from $\mathbf{x}_{rg}$. Connectivity between the back-projected pixels in 3D can be simply derived from pixel neighbor relationships in the image. However, since our range data often contains holes from missing returns, in practice we use a 2D Delaunay triangulation so we can form connections across small holes. We refer to the connected set of back-projected pixels as the *image mesh*. This image mesh can be thought of as a piecewise 2D manifold embedded in 3D. Our detection algorithm extends 2D SIFT operators to work in this space. The first problem we address is how to develop a physical scale space.

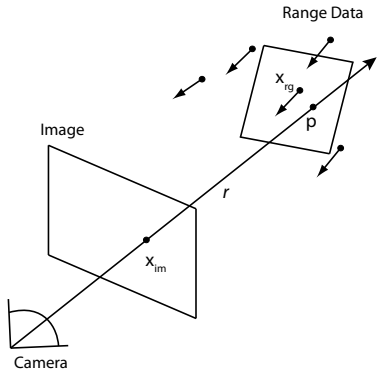

Figure 3. This figure demonstrates the back-projection of image pixels into the range space. The points $\mathbf{p}$ form the image mesh.

3.1. Physical Scale Space

Our image mesh allows us to compute a repeatable physical distance between image pixels. This physical space gives our algorithm its intrinsic invariance to scale. Therefore, we can a priori choose a set of physical scales, $S = \{s_1, \ldots, s_k\}$, at which to detect keypoints for any image mesh.

In order to build a physical scale space, we must define (1) a smoothing kernel to be applied to the image mesh, (2) a discrete convolution method for applying this kernel to the mesh, and (3) a downsampling method to create the image meshes corresponding to the different physical scales. The smoothing kernel that we use is similar to a mesh bilateral filter [3]. We chose this filter because it reduces contributions from nearby surfaces that are less likely to be seen as the viewpoint varies. Specifically, the kernel we use is

$$B(\mathbf{x}, \mathbf{n}; \mathbf{x}_0, \boldsymbol{\eta}, \sigma, \nu) = \exp(-\|\mathbf{x} - \mathbf{x}_0\|^2 / 2\sigma^2) \cdot \exp(-(1 - \mathbf{n} \cdot \boldsymbol{\eta})^2 / 2\nu^2)$$

where $\mathbf{x}$ is the 3D location at which B is being evaluated and $\mathbf{n}$ is the normal for $\mathbf{x}$. $\mathbf{x}_0$ is the kernel's center point and $\boldsymbol{\eta}$ is its normal. The first factor of the kernel is the standard Gaussian kernel using 3D Euclidean distances and σ^2 as its variance. The second factor lowers the contribution

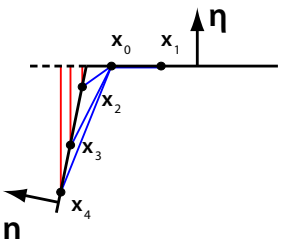

Figure 4. The blue lines represent the distances used for the first (Gaussian) term in the bilateral filter. Fleishman et al. [3] used the red distances for their second term. Using these distances, $\mathbf{x}_2$, which may not exist in a different scan will make a sizable contribution to $\mathbf{x}_0$. Our technique mitigates this problem using the dot products of the normals $\mathbf{n}$ and $\boldsymbol{\eta}$.

of points whose normal $\mathbf{n}$ significantly differs from $\boldsymbol{\eta}$. We used a value of 0.4 for ν throughout the paper, which gives near-zero weight to points whose normals differ from $\boldsymbol{\eta}$ by 90° or more. This term improves the repeatability of our scale space, especially near structural corners, as the presence of a point on the adjoining surface in the image mesh will depend on the viewpoint at which the surface was seen. It also increases robustness to changes in the light source's position, as it will mitigate contributions from a surface in shade to an adjacent surface in direct light and vice versa. This surface normal comparison approach differs from the analogous "intensity" difference term from [3], which used a point to plane distance. Figure 4 illustrates how the proposed surface normal term does a better job of reducing contributions between nearly-perpendicular surfaces.

Convolution is performed on the intensities and the normals of the points in the image mesh. Bilateral filter weights are computed for points within 2σ of the target point. The resulting intensities are normalized by the total weight and the surface normals are normalized to unit vectors. The scale s_i for each layer of physical scale space is realized by smoothing scale s_{i-1} where $\sigma = \sqrt{s_i^2 - s_{i-1}^2}$.

In contrast to SIFT, our physical scale space does not use octaves. We can eliminate octave computation because we do not need to locate extrema with respect to the scale dimension. In order to compute subsequent layers of scale space we downsample the points used for smoothing. We do this by labeling a subset of the points (starting with all) of the image mesh as control points, and using only control points as support for convolution. After each layer has been smoothed, we remove points from the control point subset such that no two points within the subset are closer than $s_i/2$ from each other. Figure 5 demonstrates four layers of the physical scale space.

3.2. Extrema Detection

Since we treat each scale independently, we compute the Laplacian of the image intensities on the mesh and then find maxima and minima. We compute the Laplacian using the

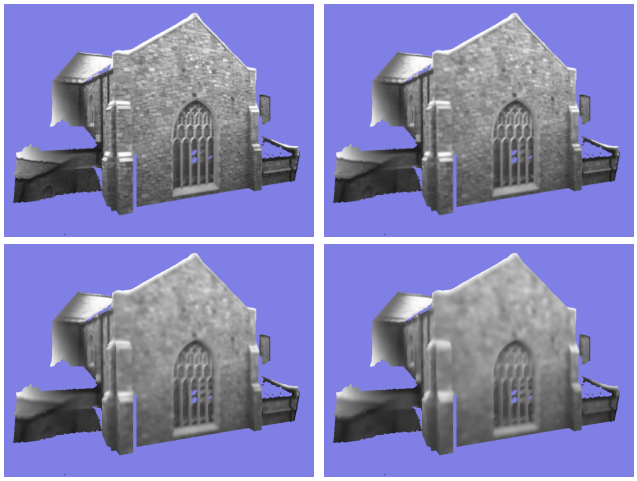

Figure 5. This figure shows the effects of the bilateral filter at different scales, the scales used were 3cm, 6cm, 12cm, and 17cm.

Laplace-Beltrami Operator (LBO), a generalization of the Laplacian to functions on manifolds. We use the indirect discretization by Xu [17], which first requires a gradient to be computed for each vertex. We use Xu's discrete linear approximation to the gradient. This technique works by computing the intensity gradient on each triangle adjacent to vertex $\mathbf{p}_i$, and then using an area-weighted average of the adjacent triangle gradients to approximate the gradient at $\mathbf{p}_i$. Figure 6 illustrates the process. Adopting the notation from [17], the gradient on a triangle is

$$\nabla_{T_j} I = \frac{1}{2A_j}[I(\mathbf{p}_i)\mathbf{v}_i + I(\mathbf{p}_j)\mathbf{v}_j + I(\mathbf{p}_{j+})\mathbf{v}_{j+}], \quad (1)$$

where A_j is the area of triangle $[\mathbf{p}_i, \mathbf{p}_j, \mathbf{p}_{j+}]$. $I(\mathbf{p_i})$ is the intensity at point $\mathbf{p}_i$, and $\mathbf{v}_i$ is the inward normal of the edge opposite $\mathbf{p}_i$ (i.e. edge $[\mathbf{p}_j, \mathbf{p}_{j+}]$) scaled by the length of that edge. It is computed by

$$\begin{aligned}\mathbf{v}_i \quad = \quad & [((\mathbf{p}_i - \mathbf{p}_j)\cdot(\mathbf{p}_j - \mathbf{p}_{j+}))(\mathbf{p}_{j+} - \mathbf{p}_i) + \\ & ((\mathbf{p}_i - \mathbf{p}_{j+})\cdot(\mathbf{p}_{j+} - \mathbf{p}_j))(\mathbf{p}_j - \mathbf{p}_i)]/2A_j,\end{aligned}$$

and similarly for the other edge normals. Thus the intensity gradient at a vertex with respect to the mesh is simply approximated by

$$\nabla_M I(x_i) = \frac{1}{A(\mathbf{p}_i)} \sum_{j \in N_1(i)} A_j \nabla_{T_j} I, \quad (2)$$

where $A(\mathbf{p}_i)$ is the sum of areas of the 1 ring of triangles connected to $\mathbf{p}_i$ ($N_1(i)$, see Figure 6). The LBO can now be approximated as

$$\Delta_M I(x_i) = \frac{1}{2A(\mathbf{p}_i)} \sum_{j \in N_1(i)} -\mathbf{v}_i^T [\nabla_M I(\mathbf{p}_j) + \nabla_M I(\mathbf{p}_{j+})]. \quad (3)$$

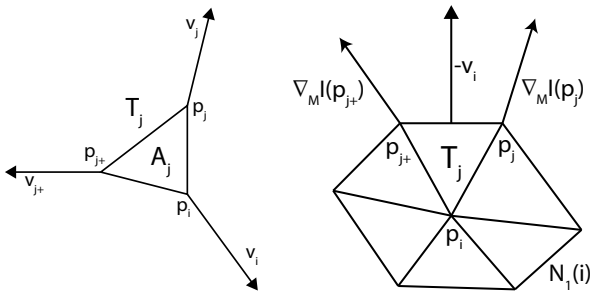

Figure 6. Left: the terms in the gradient's calculation. Right: a 1-ring neighborhood and the vectors used to compute the contribution of triangle T_j to the Laplacian.

Extrema are detected as the maxima and minima of the scale-normalized LBO by comparing a vertex with its 1-ring neighbors. Extrema are filtered by thresholding the strength of the Laplacian response to be within the top 10% of all responses. Finally, extrema are further filtered by a non-maximal suppression [2] with a suppression radius of $3s_i$. This spatial filter is applied to each scale independently, thus keypoints of differing scales cannot suppress each other.

As mentioned earlier, range data often has holes due to surfaces that scatter the return away from the scanner; we only compute values for the LBO when a full ring of vertices surrounds a point. We require at least 5 neighboring vertices with Laplacians computed within distance s_i to detect extrema. Since we do not require a point to be an extremum across scales, the number of neighbors that a vertex is compared against is significantly fewer than regular SIFT. This results in more candidate keypoint locations than regular SIFT before spatial filtering.

After the extrema have been detected, the gradients computed above are projected into the tangent space of their point for later use in computing the keypoint's orientation and descriptor. These gradients are also made more repeatable by computation in the physical scale space.

3.3. Keypoint Coordinate Frame

A full, repeatable 3D coordinate frame can now be computed for each keypoint [14, 18, 16]. One axis is taken to be the normal of the extremal vertex. The second axis is defined to be the dominant gradient direction of the intensities at the keypoint, computed as follows. We project the gradients of nearby vertices (e.g., within $3s_i$) into the tangent plane of the keypoint, weight them using the same bilateral filter as before with $\sigma = s_i$, and enter them into a histogram. The maximum of this histogram is found and a parabolic fit to the values around this maximum determines the dominant gradient direction [10]. The third axis of the keypoint is merely the cross product of the first two.

This coordinate frame is an improvement over previous work [14] because the normals in the image mesh are smoothed as the scale increases, giving more support to the normals for larger scale keypoints. The computation of the dominant gradient direction is improved by the physical scale space's invariance to strong gradients in the image caused by structural discontinuities. It is also improved by down-weighting the contribution of gradients from locations that are more sensitive to viewpoint variation. Finally, the size of the contribution region is fixed by the physical scale.

4. Keypoint Description

The SIFT-like descriptor that we compute lies in the tangent space of the keypoint, with its x-axis aligned with the dominant gradient direction. The support for the descriptor consists of the image mesh points that lie within a sphere around the keypoint with radius $8\sqrt{2}s_i$ (i.e. half the length of the descriptor's diagonal). The descriptor is a 4×4 spatial grid with 8 orientation bins per spatial bin, resulting in a standard 128-dimensional descriptor. Gradients of vertices from the support region are projected onto the descriptor's plane, then weighted with the bilateral filter (with σ as the descriptor's radius) and placed in their appropriate bins using partial volume interpolation. The descriptor is normalized and thresholded as in the original SIFT [10]. Example descriptors can be seen in the top row of Figure 1.

We choose to only histogram gradients on a single plane to be robust to changes in viewpoint. Consider a keypoint near a structural corner; if the keypoint were to wrap around the corner, then the keypoint could only be matched when both planes are seen. By only using one plane, any view that sees that plane will create a similar descriptor at that location. An alternate descriptor built using all three axes of the keypoint would also suffer from this problem; it would also have many of its bins empty for the common planar keypoint.

Our descriptor has advantages over previous methods (e.g., [14, 16]) in terms of structural discontinuity invariance and exact scale matching. Since the support of a descriptor is much larger than the scale at which the keypoint is detected, descriptor bins frequently extend across depth discontinuities into the scan's free space. In our algorithm, distant pixels from such structural discontinuities are either not in the support region or are downweighted due to the bilateral filter. Such bins will have nearly empty orientation histograms, which is a repeatable property for any view of the keypoint. The other improvement is that the physical sizes of two descriptors will exactly agree for a correct match, removing effects of differences in scale dimension during matching.

4.1. Matching

Keypoints are only matched against keypoints of the same scale. A separate k-d tree is built on the descriptors for each scale used by the image mesh being matched. As in SIFT, we find the nearest 2 keypoints in the k-d tree to each query keypoint at the scale being matched. Matches are ranked by the ratio of the distance between the query keypoint and the first and second closest keypoints. Lowe's 0.8 threshold on this ratio is used for accepting matches.

Matching keypoints only against other keypoints of the same scale significantly culls the search space for a match. The reduced search space improves the power of the ratio metric, by ensuring that the second best keypoint is of the same scale and is not an implausible match. Also, it allows a similarly-sized set of physical scale keypoints to be matched faster than multiscale keypoints. Finally, since keypoints at one scale can be computed and matched independently of keypoints at another scale, increasing the number of computed scales cannot decrease the total number of correct matches.

Candidate 3D transformations from one scan to another can be computed easily using a single keypoint match, since each match implies a full 3D coordinate system. Furthermore, because of the improvements in assigning the coordinate system to a keypoint, these transformations are generally very accurate. As we demonstrated in [14], high-quality 3D similarity transformations can be efficiently estimated using a single good 3D keypoint match as a starting point using the dual-bootstrap ICP method. Alternate RANSAC methods based on the proposed keypoints would require a relatively small number of trials to compute an accurate transformation. The RANSAC algorithm presented by Wu [16], ignoring the scale, would work here as well.

5. Results

We demonstrate the quality of our keypoint algorithm using pairwise range scan matching on real-world outdoor scans. The datasets were collected using a Leica HDS-3000 scanner. For each scan, we built the image mesh from the image that views the greatest portion of the range data. These images were acquired by the scanner at roughly the same time as the scan and are 1024×1024 in resolution.

We selected a set of scales fixed for all experiments beforehand as $S = \{3\text{cm} \times (2^{1/2})^i; i = 0, \ldots, 5\}$. For each scan, we then determined which should be the base (smallest) scale, computed as the scale closest to the median edge length in the image mesh.

Our dataset, illustrated in Figure 7, consists of eight outdoor range scan pairs, which include large intensity differences, low overlap, repeated structure, large viewpoint differences, and numerous discontinuities. Two of the scan pairs, VCC North and Parkinglot, are considered easy, since the difference in viewpoint between the scans is low for both pairs. The VCC South scan pairs are more difficult, in that the VCC South1 pair demonstrates a significant intensity difference while VCC South2 has a small overlap. The remaining four datasets are all very difficult for intensity-based keypoints; we include them to demonstrate improved matchability and to show the limits of our algorithm. The DCC scan pair has numerous occluding trees, and is made more difficult by a large viewpoint difference. The JEC, JROWL, and Biotech pairs all have significant repeated structure. The JROWL pair has one scan taken during the day and the other at dusk, and is an example of a pair that would be more suited for a global matching algorithm such as [11]. The Biotech pair is the most difficult pair, not only due to its repeated structure but also due to a narrow, distant view from one scan as well as glare in the images.

We verified keypoint matches using a ground truth registration between the scan pairs. We define the fixed keypoints to be from the scan for which the k-d tree is built and matched, and the moving keypoints to be from the scan that is querying it for matches. Matches are considered to be correct if the ground truth transformed moving keypoint's orientation vector is within 15° of the fixed keypoint's and if its location is within $4s_i$ of its matched keypoint.

The algorithm is compared to a back-projected SIFT algorithm. In back-projected SIFT (bpSIFT), keypoints are detected in the images, and back-projected into a plane fit on the range data. Then the 2D gradients from the images are mapped onto the plane in the range data to give the descriptor affine invariance [14]. This algorithm is also similar to VIP patches [16]. Correct matches for bpSIFT are measured in the same way, except that the scale used for the location threshold is the larger of the two keypoints' back-projected scales.

Since, as mentioned above, a single high-quality keypoint can be used to seed the registration of the entire scan pair, we consider the number of correct matches in the list of top 50 matches sorted by the ratio criterion to be a key measure of the quality of the matching algorithms.

Results are presented in Table 1. On the easy scan pairs, both algorithms perform well. The VCC South1 pair presents some difficulty for bpSIFT. Physical scale keypoints (PSK), however, are able to find almost three times as many correct matches ranked among the top 50 than bpSIFT. The VCC South2 pair is matched well in both algorithms. However, there are more matches for PSK at smaller scales in the small overlap region. The improved robustness to discontinuities helps improve results for the DCC pair. Improvement on the Biotech scan pair is minor, since the back wall is practically the only distinctive location in the images.

The number of keypoints ranked in the top 50 by PSK improved over bpSIFT for all non-trivial scan pairs, and

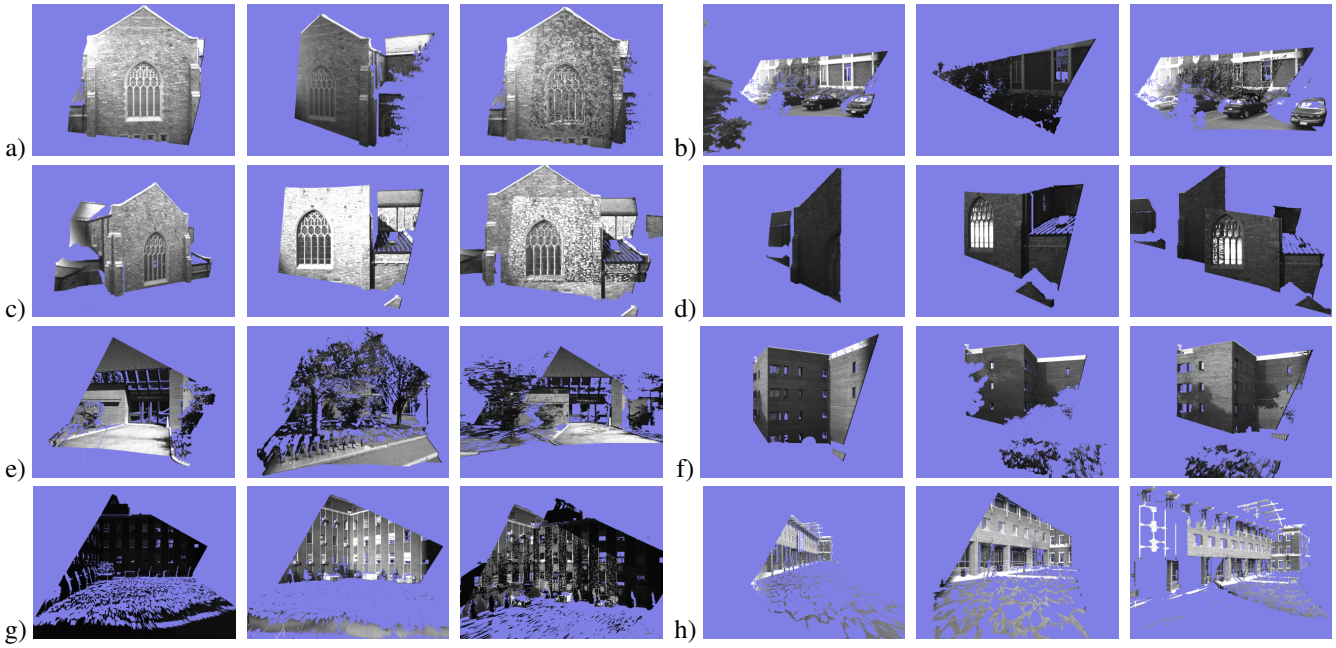

Figure 7. The experimental dataset. Columns 1, 2, 4, and 5 show the individual image meshes, and columns 3 and 6 show them aligned. The scans are: a) VCC North, b) Parkinglot, c) VCC South1, d)VCC South2, e) DCC, f) JEC, g) JROWL, and h) Biotech.

sometimes by a large margin (e.g., VCC Souths and JEC). This measure demonstrates that PSK generates more distinctive correct matches than bpSIFT. This suggests that for finding a rough transformation between the scans (e.g., to initialize ICP), an algorithm that considers the matches' ranks would quickly find the correct transformation. Even though the number of false positives increased for each pair with PSK (due to the increased number of keypoints computed, see Table 2), PSK's ratio of correct matches to false positives increased for every pair except for the Biotech pair.

Referring back to Figure 1, we see how bpSIFT fails to robustly compute a descriptor over a discontinuity, as it merely maps gradients from the image onto the plane at that keypoint, disregarding the fact that the right side of the descriptor lies over a discontinuity. Figure 8 presents some additional keypoint matches that further demonstrate PSK's robustness to discontinuities.

	PSK #correct	PSK #false pos	PSK #top 50	bpS #correct	bpS #false pos	bpS #top 50
VCC North	2983	352	50	1090	170	50
Parkinglot	262	153	50	40	30	32
VCC South1	275	174	46	21	54	17
VCC South2	92	154	42	12	36	13
DCC	28	304	9	3	50	3
JEC	177	490	30	3	107	1
JROWL	43	729	9	5	174	1
Biotech	9	429	2	5	157	1

Table 1. Results across 8 image pairs. PSK columns correspond to the proposed physical scale keypoints, and bpS to back-projected SIFT. # correct indicates the number of keypoint matches that passed the ratio test and were verified by a ground truth as correct. # false positives indicate mismatched keypoints that pass the ratio test. # top 50 are the number of correct matches ranked in the top 50 when sorted by ratio.

6. Conclusions

We proposed a physical scale based intensity keypoint algorithm that is able to increase matchability for range scans of outdoor scenes. Through fixing the scales of keypoint matches, we improve the repeatability of both the detector and the descriptor, as well as the distinctiveness of the matches. We further reduced the effect of the viewpoint on the detector and descriptor by using a bilateral filter and physical scale. We used difficult experimental datasets to demonstrate how far we could push intensity-based keypoints.

While we improved the keypoint's robustness to discontinuities, more work is required to obtain true discontinuity invariance. Our descriptor is only invariant to discontinuities in which the entire missing area lies in the free space of the scan. Future work may include a different matching scheme for descriptors that partially lie in the hidden space of the scan, as well as a progressive meshes [5] type algorithm to reduce the computation time at larger scales.

We note that the quality of the initial normal computation affects the proposed algorithm in both detection and description. In the future we plan to test our algorithm on sparser datasets where normal estimation can be difficult. We speculate that small keypoints may fail due to poor normals, but larger keypoints should still be usable due to the normal smoothing.

	PSK #fixed	PSK #moving	bpS #fixed	bpS #moving
VCC North	16	20	13	9
Parkinglot	21	8	6	0.6
VCC South1	18	13	7	9
VCC South2	8	9	0.9	4
DCC	14	18	6	10
JEC	28	27	5	7
JROWL	38	26	2	7
Biotech	15	26	3	7

Table 2. Numbers of keypoints computed (in thousands) in the fixed and moving scans under both algorithms. PSK computes many more keypoints as the filtering is mainly spatial and is not across scales, this can be seen especially in scans whose images project onto a much larger area.

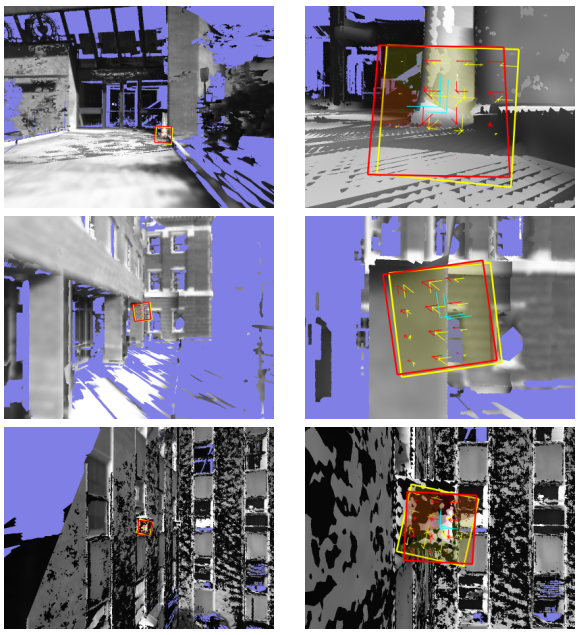

Figure 8. Interesting keypoint matches. The left column is a wider view of the area being matched; the keypoint is shown close up in the right column. These examples are taken from the DCC, Biotech, and JROWL respectively. The moving keypoint is shown in yellow with its matched fixed keypoint in red.

References

[1] E. Akagndz and I. Ulusoy. Scale and orientation invariant 3d interest point extraction using hk curvatures. In *3dRR, ICCV Workshops*, 2009.

[2] M. Brown, R. Szeliski, and S. Winder. Multi-image matching using multi-scale oriented patches. In *CVPR*, 2005.

[3] S. Fleishman, I. Drori, and D. Cohen-Or. Bilateral mesh denoising. *ACM Trans. Graph.*, 22(3):950–953, 2003.

[4] A. Frome, D. Huber, R. Kolluri, T. Blow, and J. Malik. Recognizing objects in range data using regional point descriptors. In *ECCV*, 2004.

[5] H. Hoppe. Progressive meshes. In *SIGGRAPH*, 1996.

[6] J. Hua, Z. Lai, M. Dong, X. Gu, and H. Qin. Geodesic distance-weighted shape vector image diffusion. *IEEE Trans. Vis. Comput. Graph.*, 14(6):1643–1650, 2008.

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

[8] B. J. King, T. Malisiewicz, C. V. Stewart, and R. J. Radke. Registration of multiple range scans as a location recognition problem: Hypothesis generation, refinement and verification. In *3DIM*. IEEE Computer Society, 2005.

[9] T. Lindeberg. Scale-space theory: A basic tool for analysing structures at different scales. *Journal of Applied Statistics*, 21:224–270, 1994.

[10] D. Lowe. Distinctive image features from scale-invariant key-points. *IJCV*, 60:91–110, 2004.

[11] A. Makadia, A. Patterson, and K. Daniilidis. Fully automatic registration of 3d point clouds. In *CVPR*. IEEE Computer Society, 2006.

[12] A. S. Mian, M. Bennamoun, and R. A. Owens. Matching tensors for automatic correspondence and registration. In *ECCV*, 2004.

[13] J. Novatnack and K. Nishino. Scale-dependent 3d geometric features. In *ICCV*. IEEE, 2007.

[14] E. Smith, B. King, C. Stewart, and R. Radke. Registration of combined range-intensity scans: Initialization through verification. *Computer Vision and Image Understanding*, 110(2):226–244, 2007.

[15] J. Starck and A. Hilton. Correspondence labelling for wide-timeframe free-form surface matching. In *ICCV*, 2007.

[16] C. Wu, B. Clipp, X. Li, J.-M. Frahm, and M. Pollefeys. 3d model matching with viewpoint-invariant patches (vip). In *CVPR*, 2008.

[17] G. Xu. Convergent discrete laplace-beltrami operators over triangular surfaces. In *GMP*, pages 195–204. IEEE Computer Society, 2004.

[18] A. Zaharescu, E. Boyer, K. Varanasi, and R. Horaud. Surface feature detection and description with applications to mesh matching. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[19] Y. Zhong. Intrinsic shape signatures: A shape descriptor for 3d object recognition. In *ICCV Workshops*. IEEE, 2009.

[20] G. Zou, J. Hua, Z. Lai, X. Gu, and M. Dong. Intrinsic geometric scale space by shape diffusion. *IEEE Trans. Vis. Comput. Graph.*, 15(6):1193–1200, 2009.