In Proceedings of IEEE International Conference on Robotics and Automation (ICRA), Shanghai, China, May 2011.

Interest Point Detection in Depth Images through Scale-Space Surface Analysis

Jörg Stückler and Sven Behnke

***Abstract*— Many perception problems in robotics such as object recognition, scene understanding, and mapping are tackled using scale-invariant interest points extracted from intensity images. Since interest points describe only local portions of objects and scenes, they offer robustness to clutter, occlusions, and intra-class variation.**

In this paper, we present an efficient approximate algorithm to extract surface normal interest points (SNIPs) in corners and blob-like surface regions from depth images. The interest points are detected on characteristic scales that indicate their spatial extent. Our method is able to cope with irregularly sampled, noisy measurements which are typical to depth imaging devices. It also offers a trade-off between computational speed and accuracy which allows our approach to be applicable in a wide range of problem sets. We evaluate our approach on depth images of basic geometric shapes, more complex objects, and indoor scenes.

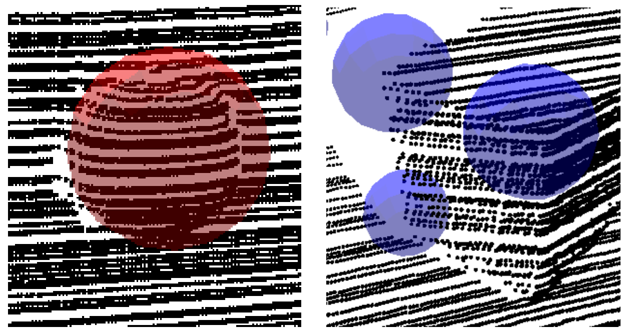

Fig. 1. Characteristic examples for detected blobs (left) and corners (right). The interest points are detected on characteristic scales which reflect their spatial extent. For blobs the spatial extent corresponds to the curvature radius of the surface. Corners are extracted on their maximum spatial extent.

I. Introduction

Interest points provide a compact representation of image content. They describe only local parts of the observed scene and, hence, offer robustness to clutter, occlusions, and intra-class variation. For such properties, interest points are a favorable choice to solve perception problems in robotics such as object recognition, scene understanding, and mapping.

The computer vision community has developed efficient means for the extraction of scale-invariant interest points from intensity images. These approaches are often based on multi-scale pyramid representations of the image, in which the image is successively subsampled with increasing scale. While methods to detect interest points on multiple scales in point clouds and depth images have been developed, they lack the computational efficiency of their intensity image counterparts. In intensity images, the interest operators are applied to the intensity values which are sampled on a 2D lattice, whereas in depth images, surface geometry is the quantity of interest. It is measured in the implicit form of the point cloud distribution such that the direct transfer of concepts from the image processing domain to depth images is not possible.

In this paper, we present an efficient approximate algorithm to extract surface normal interest points (SNIPs) in corners and blob-like regions from depth images. We detect these interest points on characteristic scales that indicate their spatial extent. Our method extends the approach of Unnikrishnan et al. [1]. They derive an interest operator that is based on convolving surfaces with Gaussian kernels on multiple scales. It is directly applied to the point cloud data without relying on a mesh parametrization. We apply their approach to noisy depth images from a single view and present approximations to achieve a trade-off between computational efficiency and accuracy.

In order to gain computational efficiency, our approach allows to specify an image neighborhood size for the computations involved. Similar to intensity image algorithms, we build a pyramid representation of the depth image. Range-sensing devices sample the observed surface irregularly with varying sampling density throughout the depth image. Sampling density depends on sensor characteristics like angular resolution as well as on distance to and impact angle on the reflecting surface. To cope with the varying sampling density, we propose to estimate at each image location and scale an optimal lowest resolution for computation within the specified image neighborhood.

Occlusions constitute a further difficulty for interest point detection in depth images from single views. Approaches to interest point detection in point clouds often assume that the object is densely sampled in a complete view without holes. We propose means to take special care of occlusion effects.

We evaluate our approach on depth images of basic geometric shapes, more complex objects, and indoor scenes. We demonstrate the behavior of our interest point detector qualitatively, and measure its repeatability.

The remainder of this paper is organized as follows. After a short review of related work in Sec. II, we will introduce the concepts of multi-scale interest point detection in unorganized point clouds in Sec. III. We detail our approach in Sec. IV and assess its quality in experimental evaluation in Sec. V.

All authors are with the Autonomous Intelligent Systems Group, University of Bonn, Germany. Email: stueckler@ais.uni-bonn.de

II. Related Work

Various methods have been recently developed to extract interest points from dense, complete-view point clouds on a characteristic scale. The scale reflects the spatial extent of the underlying geometric structure.

Pauly et al. [2], for example, measure surface variation at a point by considering the eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$ of the local sample covariance in the n-point neighborhood of the examined point. The scale is chosen to be the neighborhood n for which the variation measure attains a local extremum. As the variation measure is susceptible to noise, heuristics have to be applied to detect extrema with satisfying robustness.

Novatnack et al. [3] extract multi-scale geometric interest points from dense point clouds with an associated triangular connectivity mesh. In this approach geodesic distance between points is computed from shortest paths through the mesh. They build a scale-space of surface normals given by the mesh and derive edge and corner detection methods with automatic scale selection. Our approach does not require a mesh for connectivity information and surface normals.

Unnikrishnan et al. [1] derive an interest operator and a scale selection scheme for unorganized point clouds. They do not rely on connectivity information given by a mesh. In contrast to our approach, they extract geodesic distances between points using disjoint minimum spanning trees in a pre-processing stage. They present experimental results on complete views of objects without holes.

Some approaches detect multi-scale interest points in depth images, since range-sensing devices obtain depth images from a single view and such approaches do not require registration of multiple views. Stable interest points can even be used to select points for sparse feature-based point cloud registration.

Lo et al. [4] directly apply the interest point operator of SIFT to the normalized depth image. While the intensity of a point is not affected by projection into the image, measured depth naturally depends on the view point. In our approach, the appearance of surfaces only changes implicitly in the sampling densities for which we account.

Novatnack et al. [5] transfer the approach in [3] to depth images. They approximate geodesic distances by computing shortest distances between points through the image lattice. Surface normals are computed by triangulating the range image. They evaluate their approach for the feature-based registration of depth images with high sampling density and low noise. In our approach, we detect multi-scale corner- and blob-like interest points on surfaces. We explicitly take the sampling density at each depth image location into account.

Flint et al. [6] generate a 3D voxel image which represents the sampling density of the depth image. By building a scale-space representation of the 3D image analogously to 2D images, they find scale-space maxima in sampling density. They use the Determinant of Hessian as interest operator. In their approach, Flint et al. assume that changes of sampling density are only caused by changes of surface geometry. However, for range-sensing devices, the sampling density also depends on the impact angle on the surface.

Steder et al. [7] extract interest points from depth images without scale selection, based on a measure of impact angle variation. Similar to our approach, they also reject unstable interest points in virtual geometric structure at occlusions. Recently, they proposed the NARF feature descriptor [8] to describe depth image interest points.

III. Multi-Scale Interest Points from Irregularly Sampled Point Clouds

Unnikrishnan et al. [1] derive a multi-scale interest operator for irregularly sampled point clouds. First, they define an integral operator which yields local mean curvature estimates of the underlying surface. They normalize this operator for sampling irregularity and develop a scale selection mechanism to detect interest points on characteristic scales.

A. Multi-Scale Interest Operator for Smooth Surfaces

The approach is based on an integral operator $A : \mathbb{R}^d \times \mathbb{R}^+ \rightarrow \mathbb{R}^d$ that acts on d-dimensional surfaces with an associated one-dimensional scale parameter t. This integral operator is defined as

$$A(\alpha(s), t) = \int_\Gamma \phi(s, u, t)\alpha(u)du, \tag{1}$$

where in the 2D case, Γ is a smooth 2D curve given as a function $\alpha(s)$ of distance s. Using a normalized Gaussian kernel function

$$\phi(s, u, t) = (2\pi t^2)^{-\frac{1}{2}} \exp\left(-\frac{(s-u)^2}{2t^2}\right) \tag{2}$$

and applying second order Taylor expansion to the 2D curve, they obtain that the operator A approximately displaces the point x on the curve in direction of the normal $n(x)$ to the curve and in proportion to its curvature $\kappa(x)$ in x, i.e.,

$$A(x, t) \approx x + \kappa(x)n(x)\frac{t^2}{2} \tag{3}$$

This fact can be transfered to 3D surfaces by considering the 2D normal curves Γ_θ in a point x, which result from intersecting the surface with normal planes Π_θ that are orthogonal to the tangent plane at an angle θ to some reference normal plane. Let Γ_θ be parametrized by function $\alpha_\theta(s)$. Using eq. (3) it follows that

$$\begin{aligned} A(x, t) &\approx x + \frac{1}{2\pi}\int \kappa(x, \theta)\ n(x)\ \frac{t^2}{2}\ d\theta \\ &= x + H(x)\ n(x)\ \frac{t^2}{2}. \end{aligned} \tag{4}$$

Similar to the 2D case, A displaces the point x normal to the 3D surface in proportion to the mean curvature $H(x)$ at x.

B. Invariance to Irregular Sampling

In the derivation of the operator A, we assume uniform sampling of the surface. Real depth sensing devices, however, sample surfaces irregularly. Unnikrishnan et al. propose to account for this sampling irregularity by normalizing the kernel ϕ and, thus, the operator A for the sampling distribution:

$$\begin{aligned} \tilde{A}(\alpha(s), t) &= \frac{1}{\tilde{d}(s,t)} \int_\Gamma \tilde{\phi}(s,u,t)\alpha(u)du, \\ \tilde{d}(s,t) &= \int_{-\infty}^{+\infty} \tilde{\phi}(s,u,t)p_t(u)du, \\ \tilde{\phi}(s,u,t) &= \frac{\phi(s,u,t)}{p_t(s)p_t(u)}, \end{aligned} \tag{5}$$

where $p_t(s)$ estimates the sampling distribution at s through kernel density estimation with bandwidth t,

$$p_t(s) = \int \phi(s,u,t)p(u)du. \tag{6}$$

C. Scale Selection

Unnikrishnan et al. motivate the characteristic scale of an interest point to be the radius of curvature in the case of a perfect sphere. In order to detect interest points on their characteristic scale, they define the scalar function

$$B(x,t) = \frac{2\left\|x - \tilde{A}(x,t)\right\|}{t} \exp\left(-\frac{2\left\|x - \tilde{A}(x,t)\right\|}{t}\right), \tag{7}$$

which attains local extrema in scale-space at $t_{max}(x) = \frac{1}{H(x)}$.

IV. Efficient Detection of Multi-Scale Interest Points in Depth Images

Unnikrishnan et al. determine the geodesic distance between points on a surface in a time-consuming preprocessing stage. They also consider the complete point neighborhood in the highest available resolution on all scales. In order to gain computational efficiency, we make two significant approximations to this algorithm. First, we approximate geodesic distance with Euclidean distance between points. Second, we tailor the approach to depth images and allow to specify an image neighborhood in which the interest operator is evaluated. Since the sampling density varies throughout the image with distance and view angle onto the viewed surface, we apply the operator on appropriate sampling rates of the depth image. We propose a mechanism to select a lowest image resolution at each image location to approximate the interest operator well.

In depth images, occlusions constitute a further difficulty for interest point detection. When surfaces are partially occluded, virtual geometric structure appears in the background at the edges of shadows. Furthermore, surfaces artificially appear highly curved at depth discontinuities. We present methods to recognize false detections of interest points in virtual structure.

Our third enhancement to the approach of Unnikrishnan et al. is a detector for sharp corners on characteristic scales. The

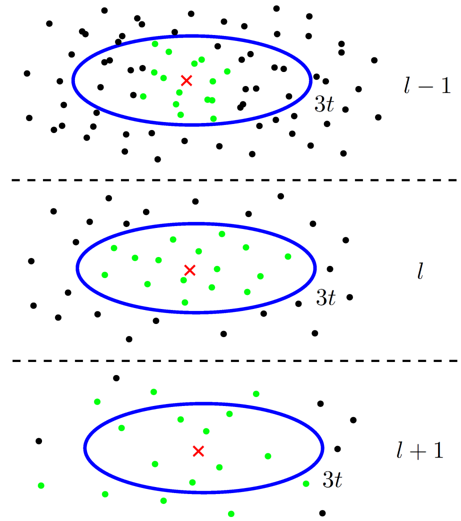

Fig. 2. For each scale t, we determine the layer l with lowest resolution at which the interest operator is sufficiently well approximated by the samples within a specified $k \times k$ image neighborhood (green/gray dots). The blue circles indicate the $3t$ range around the query point (red cross). Due to the fixed image neighborhood, oversampling (layer $l-1$) and undersampling (layer $l+1$) reduce approximation quality.

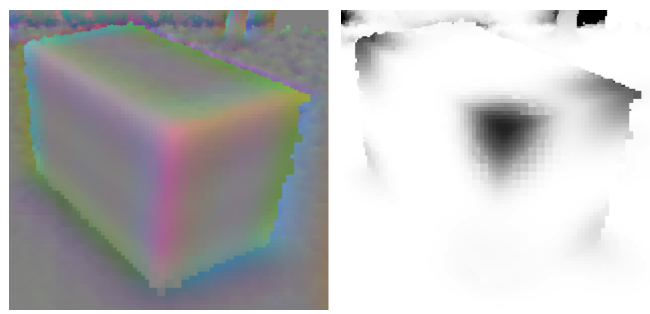

Fig. 3. Mean curvature scaled normals (left, RGB encodes 3D direction, saturation in proportion to mean curvature) and corner response (right).

interest operator measures local surface variation. We extract corners on scales that correspond to the largest extent of the feature.

A. Depth Image Pyramid

We seek to transfer concepts from the 2D image processing domain to achieve computational efficiency. In multi-scale representations, images are successively low-pass filtered. Since low-frequency structure in images can be approximated sufficiently well with adequately small sampling densities, subsampling is applied to reduce the computational cost with increasing scale. Such efficient pyramid representations are core to many efficient interest point detection algorithms in 2D images like SIFT [9], for example.

Our interest operator $\tilde{A}$ convolves surfaces with a Gaussian kernel that measures distance in the coordinates of points that are irregularly sampled in the depth image. This contrasts with typical convolution operators in intensity images, where distance is measured in regularly sampled image coordinates. Moreover, depth images sample surfaces with varying density throughout the image. This sampling density not only depends on measurement characteristics of the sensor like angular resolution, but also on distance to and impact angle on the reflecting surface.

For this reason, we estimate for each 2D image location (i, j) and scale t the subsampling layer $\hat{l}(i, j, t)$ which provides the best surface representation within a specified $k \times k$ image neighborhood. Smaller image neighborhoods achieve faster computation but crude approximations, while larger image neighborhoods trade better approximations for lower frame rates.

1) Multi-Scale Pyramid Representation: We generate a multi-scale representation ranging over S discrete scales $\{t_s\}_{s=1}^{S}$. On each scale, we apply the interest operator $\tilde{A}(x, t)$ to the 3D points represented in the depth image $\mathcal{I}^d$. As basis for the computation on various sampling densities, we construct a pyramid representation $\mathcal{I}_l^{\text{3D}}$ of the 3D point image consisting of L layers. We successively subsample the layers by a factor of 2 where we average over nearby subsumed points.

Since the sampling layer varies with image location and scale, we represent each scale in a pyramid of L images $\mathcal{I}_{t,l}^{\tilde{A}}$ for each scale t.

Let $x(i, j, l) := \mathcal{I}_l^{\text{3D}}(i, j)$ be the 3D point at image location (i, j) at subsampling layer l. We evaluate the interest operator $\tilde{A}$ at image location (i, j), scale t, and subsampling layer l by considering neighbors $\mathcal{N}(i, j, l, \hat{k})$ within a $\hat{k} \times \hat{k}$ image neighborhood

$$\mathcal{I}_{t,l}^{\tilde{A}}(i, j) = \frac{\sum_{x' \in \mathcal{N}(i,j,l,\hat{k})} \tilde{\phi}(x, x', t)\ x'}{\sum_{x' \in \mathcal{N}(i,j,l,\hat{k})} \tilde{\phi}(x, x', t)}, \tag{8}$$

where $x := x(i, j, l)$ and $\hat{k} \geq k$. The density normalized kernel $\tilde{\phi}$ involves the estimation of the local sampling density p_t for the image locations and layers at which we evaluate $\tilde{A}$. On each scale, the interest operator is only evaluated once for each location in the original resolution at its optimal sampling layer $\hat{l}(i, j, t)$.

2) Estimation of the Optimal Sampling Layer: Selecting the optimal sampling layer for the evaluation of the interest operator is crucial to obtain good approximations within the specified image neighborhood (s. Fig. 2). If the resolution is chosen too high, the samples within the image neighborhood may only contain a fraction of the neighborhood within scale range. The approximation error is then dominated by the neglected samples with comparatively high weight $\tilde{\phi}$. When the local surface properties are approximated too sparsely, important structure at higher frequencies may be omitted.

We define the expected local point density $\rho_t(i, j, l)$ to be the expected density of the points in the original resolution that we represent by the $k \times k$ image neighborhood $\mathcal{N}(i, j, l, k)$ on layer l:

$$\rho_t(i, j, l) := E(p_t(i, j, l)) = \sum_{x' \in \mathcal{N}(i,j,l,k)} \phi(x, x', t)\ c(x') \tag{9}$$

where $x := x(i, j, l)$ and $c(x')$ is the number of points subsumed on the layer l by the image location (i', j') that corresponds to x'. Note, that the number of points may be

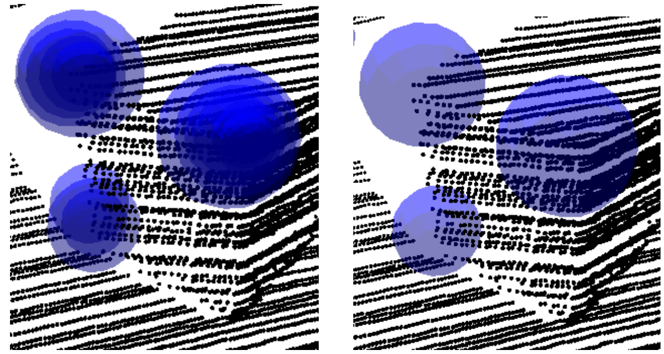

Fig. 4. Corners without scale-selection (left) and with scale-selection (right).

less than square the subsampling rate if the depth image contains invalid measurements. We determine the expected density from k^2 point samples on layer l. This avoids the inefficient calculation of the actual density over all points within the original resolution.

At coarse image resolution, the estimated density approximates the true density well, when there are enough samples with diverse Gaussian weight within the image neighborhood. We therefore require the density of the points within the image neighborhood to achieve a minimum density $\rho_{\min}$. When the image resolution is too fine-grained, only parts of the surrounding samples with high weight are contained in the image neighborhood. At such resolutions, the expected density yields only a fraction of the true density.

For each image location and scale, we select the first sampling layer $\hat{l}(i, j, t)$ at which the next higher resolution drops significantly in expected density. By this, we find the layer with lowest resolution at which the interest operator is sufficiently well approximated within the $k \times k$ image neighborhood.

B. Multi-Scale Blob Interest Points

Following the approach in Unnikrishnan et al., we detect multi-scale interest points in scale-space maxima of the function $B(x, t)$. We compare each location (i, j, t) in the scale-space with the neighbors within a range of t on the same scale and on the neighboring scales. Alluding to the notions of interest points in 2D images, we use the term ”blob” as a shorthand for this type of interest point.

C. Multi-Scale Corner Interest Points

We seek to detect interest points where the surface exhibits strong local variation in normal orientation. For each corner, we select a scale which describes its largest spatial extent.

1) Interest Point Detection: In regions of significant local mean curvature, the interest operator $\tilde{A}(x, t)$ yields an approximation of the surface normal at x. We measure the local variation of the surface normals on a scale t with the

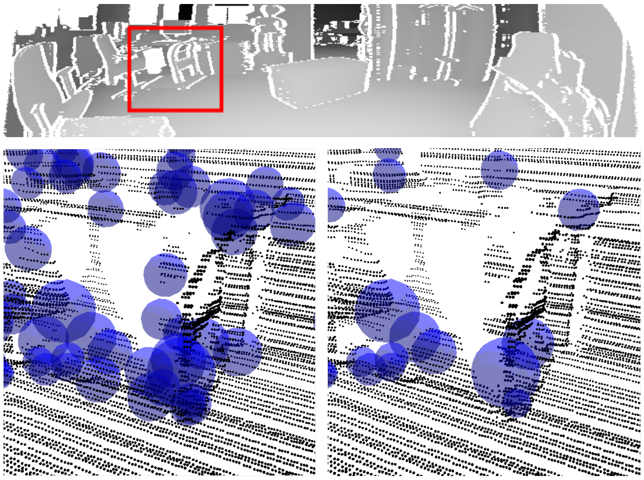

Fig. 5. Bottom: close view on detected corners without (left) and with border ownership constraint (right). Top: depth image of the scene (red rectangle marks viewed volume).

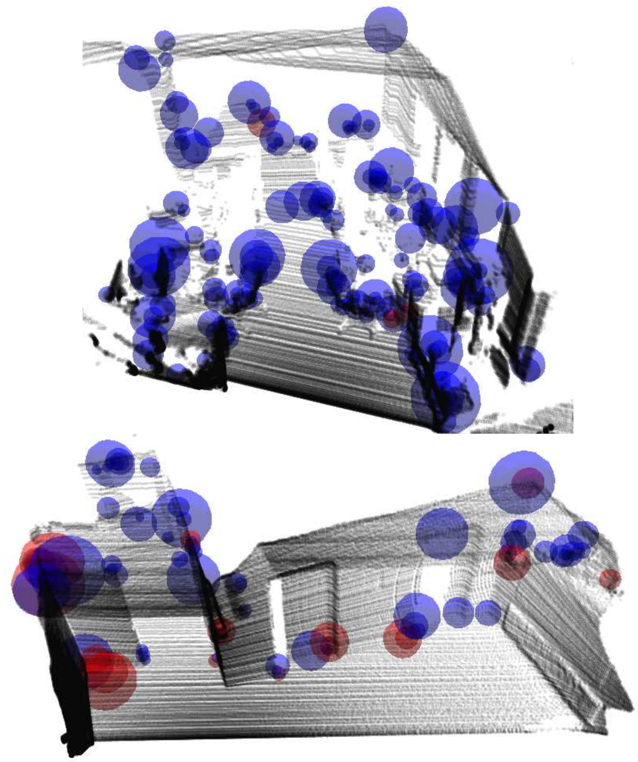

Fig. 6. Corners (blue) and blobs (red) detected in two indoor scenes (top: office, bottom: corridor). Best viewed in color.

3D structure tensor

$$S(x,t) = \frac{1}{\sum_{x' \in \mathcal{N}} \phi(x,x',t)} \cdot \sum_{x' \in \mathcal{N}} \phi(x,x',t) \begin{pmatrix} \tilde{n}_x^2 & \tilde{n}_x\tilde{n}_y & \tilde{n}_x\tilde{n}_z \\ \tilde{n}_x\tilde{n}_y & \tilde{n}_y^2 & \tilde{n}_y\tilde{n}_z \\ \tilde{n}_x\tilde{n}_z & \tilde{n}_y\tilde{n}_z & \tilde{n}_z^2 \end{pmatrix}, \quad (10)$$

where $\mathcal{N} := \mathcal{N}(i,j,l,\hat{k})$ is the $\hat{k} \times \hat{k}$ image neighborhood of x and $\tilde{n}(x) := \frac{2\left\|x - \tilde{A}(x,t)\right\|}{t}\, n(x)$. By scaling the normal at x with the scale-normalized displacement of the point, the measure emphasizes regions of high curvature and is comparable across scales.

The eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$ and the corresponding eigenvectors e_1, e_2, e_3 of the structure tensor summarize the local normal distribution at x. Regions with strong local surface variation are indicated by high eigenvalues in all three spatial directions. We therefore detect corners at the spatial extrema of the determinant of the structure tensor $C(x,t) := \det(S(x,t)) = \lambda_1\lambda_2\lambda_3$.

2) Scale Selection: We determine the scale of a corner as the highest scale at which the corner exerts a spatial extremum. After detecting corners on each scale t_s separately, we prune corners on the next lower scale t_{s-1} that are positioned within a constant factor $\tau = 2/3$ within the scale range of the corner.

D. Rejection of Unstable Interest Points

Noisy depth measurements may cause bad localized extrema along edges and ridges. Occlusions give rise to depth discontinuities, where the interest operator $\tilde{A}$ estimate is strongly influenced by border effects. Also, when parts of the background are occluded virtual geometric structure appears. We remove interest points caused by such effects through the following means.

1) Significant Interest Points: We discard spurious extrema at low blob and corner responses with thresholds on the blob response function B and the corner response function C, respectively.

2) Edge Responses: Surface ridges and edges appear highly curved. Noisy depth measurements causes small fluctuations in curvature and, hence, our method detects blobs along such structure. We reject these interest points, since they are not well localized. In order to identify blobs in elongated structure, we measure cornerness of the mean curvature. Similar to the Harris corner measure in intensity images, we compare the eigenvalues $\lambda_{min}, \lambda_{max}$ of the 2D structure tensor in the local image neighborhood of each interest point. We reject interest points, when the ratio $\lambda_{min}/\lambda_{max}$ of the eigenvalues is below a threshold (set to 0.15).

3) Depth Discontinuities and Border Ownership: At depth discontinuities, the interest operator $\tilde{A}$ does not measure the actual mean curvature of the partially viewed surface. We thus reject blobs close to depth discontinuities.

Occlusions cause virtual geometric structure in the background. Interest points in such structure can be identified where depth discontinuities towards the foreground occur. We filter corners at depth discontinuities that do not possess border ownership. We test for the proximity of an interest point to depth discontinuities within its support region by tracking depth jumps in the image. It suffices to search in the horizontal and the vertical directions.

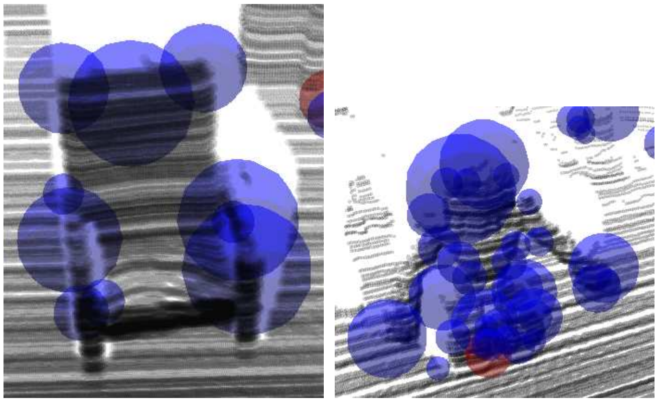

Fig. 7. Corners (blue) and blobs (red) detected on a chair (left) and a humanoid robot Nao (right). Best viewed in color.

	blobs		corners	
$k/\hat{k}$	$7/13$	$21/41$	$7/13$	$21/41$
office	0.20 (0.47)	0.16 (0.33)	0.37 (0.40)	0.38 (0.42)
corridor	0.24 (0.42)	0.24 (0.38)	0.39 (0.43)	0.42 (0.50)
box (d=0.6m)	0.09 (0.33)	0.31 (0.47)	0.43 (0.44)	0.50 (0.51)
box (d=1.0m)	0.02 (0.14)	0.45 (0.50)	0.46 (0.46)	0.46 (0.48)
box (d=1.4m)	n/a	0.28 (0.45)	0.38 (0.40)	0.41 (0.43)
nao (d=0.6m)	0.22 (0.25)	0.43 (0.45)	0.41 (0.41)	0.42 (0.43)
nao (d=1.0m)	0.23 (0.26)	0.25 (0.36)	0.37 (0.37)	0.37 (0.37)
nao (d=1.4m)	0.24 (0.48)	0.10 (0.33)	0.33 (0.33)	0.34 (0.34)
chair (d=0.6m)	n/a	0.16 (0.65)	0.49 (0.49)	0.50 (0.51)
chair (d=1.0m)	n/a	n/a	0.43 (0.43)	0.46 (0.46)
chair (d=1.4m)	n/a	n/a	0.38 (0.39)	0.39 (0.43)

TABLE I

BLOB AND CORNER OVERLAP (IN BRACKETS: OVERLAP FOR MATCHABLE INTEREST POINTS) FOR VARYING IMAGE NEIGHBORHOODS k AND $\hat{k}$ AND OBJECT DISTANCES (d)

$k/\hat{k}$	$7/13$	$11/21$	$21/41$
office, avg. image size 1040×176.3	5.27	8.48	48.28
corridor, avg. image size 1040×222.4	4.78	8.61	61.59

TABLE II

COMPUTATION TIME IN SECONDS FOR VARYING IMAGE NEIGHBORHOODS k AND $\hat{k}$.

V. EXPERIMENTS

We evaluate our approach on depth images obtained with a 3D laser-range finder (LRF) and a Microsoft Kinect camera. The 3D LRF consists of a Hokuyo UTM-30LX mounted on a pitch actuator. We preprocess the laser-range measurements by removing interpolation effects at depth discontinuities. Then, we apply a median filter of width 11. We smooth the 3D point images with a Gaussian kernel with bandwidth $\sigma = 0.01m$ in the $\hat{k} \times \hat{k}$ image neighborhood.

A. Qualitative Assessment

In the following, we use the settings $k = 21$ and $\hat{k} = 41$. The motivating example in Fig. 1 shows interest points detected by our approach on ideal corner and blob-like shapes. As can be seen, our approach detects interest points at the correct spatial position and selects a scale which reflects the spatial extent of the shapes.

Fig. 6 demonstrates the behavior of our interest point detector on scans of typical indoor scenes. In the cluttered office scene (top), corners and blobs detect salient geometric features of foreground objects (for example, the back and legs of chairs or corners of tables and monitors). The corridor scene (bottom) is less cluttered. Here, the interest points are detected in salient features of the building structure like wall corners and doors. Salient geometric structure is also discovered on objects such as a chair or a humanoid robot Nao (s. Fig. 7).

B. Quantitative Evaluation

We use two measures of repeatability to evaluate our approach. Matching repeatability measures the frequency with which corresponding interest points can be found between images. Interest point overlap reflects the stability of the interest points and the accuracy in spatial location and scale between images.

Following the definition in [1], we compute the overlap repeatability as the average overlap of corresponding interest points. For this measure, correspondence is established between interest points with largest non-zero overlap. We determine the overlap as the ratio of the intersection to the union of the support regions of the interest points. We define the 3D sphere with scale radius around the interest point as a point's support region. As noted by Unnikrishnan et al., an overlap of approx. 0.35 can be handled with a well designed descriptor.

1) Efficiency-Accuracy Trade-off: We measure overlap repeatability for several indoor scenes. For each scene we captured 10 images with the 3D LRF. For the box and nao scenes, we extract interest points from 8 scales ranging from $t_{min} = 0.04\,\text{m}$ to $t_{max} \approx 0.135\,\text{m}$. We evaluate the other scenes with 8 scales from $t_{min} = 0.1\,\text{m}$ to $t_{max} \approx 0.336\,\text{m}$. Table I and II depict, how the image neighborhoods k and $\hat{k}$ influence computational efficiency and accuracy. While computation time decreases with the neighborhood size, repeatability drops on our test datasets. The experiments have been carried out on an HP Pavilion dv6 notebook with an Intel Core i7 Q720 processor. Note, that our implementation may still be tuned for lower computation times.

2) Noise Effects: Fig. 8 shows that overlap repeatability increases with scale in the corridor dataset, especially for corners. This is due to the fact that larger scales are more robust against the noise inherent in the depth measurements. On smaller scales, noise artifacts and low sampling density induce spurious detections.

3) View Point Change: We further evaluate the repeatability of our interest points with regard to view point changes in experiments with ideal shapes for corners and blobs. We recorded RGB-D images of a half sphere on a plane and a corner on a cube with a Kinect camera from various distances and view angles in the horizontal plane. For these experiments, we choose $k = 21$ and $\hat{k} = 41$. Using visual

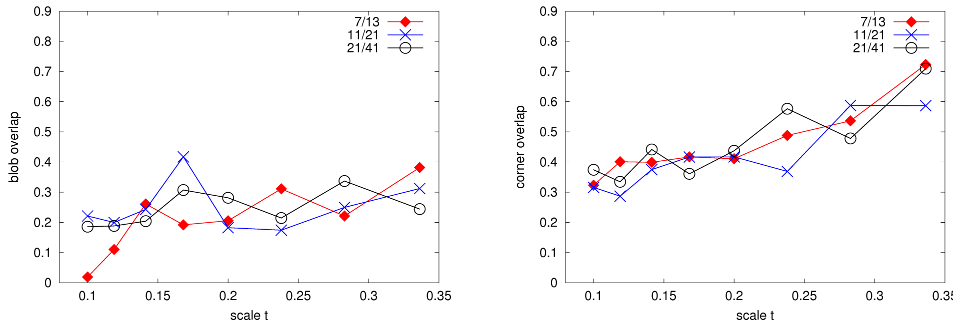

Fig. 8. Blob and corner overlap repeatability depending on scale on the corridor dataset for varying image neighborhoods k / $\hat{k}$.

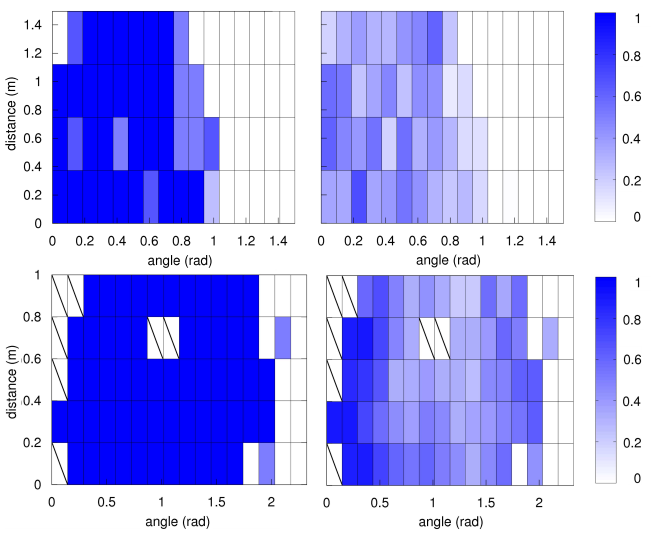

Fig. 9. Blob (top) and corner (bottom) repeatability (left: matching, right: overlap) wrt. view angle and distance change for a half sphere on a plane and a corner on a cube.

markers and point cloud registration of the depth images, we determine the view point change between the images. We measure matching and overlap repeatability of the first image interest point towards images from a wide range of view points. As shown in Fig. 9, the matching repeatability is close to one for large viewpoint changes. For corners, the average overlap measure drops from high values of approx. 0.7 to values of approx. 0.35 within 0.7 rad view angle change. Blobs are less well localized and thus the overlap measure is on average lower than for corners. Also, corners can be seen within a larger view angle range than blobs.

VI. Conclusions

In this paper, we presented a new method to extract multi-scale surface normal interest points (SNIPs) from depth images. Our method detects interest points in regions with high curvature, like corners and blob-like regions, on their characteristic scale. Our interest point detection method copes with irregular sampling density across the image and occlusion effects. We achieve computational efficiency by estimating for each image location and scale a lowest resolution at which sufficient point samples are available. The efficiency-accuracy trade-off is adjustable through the size of the image neighborhood k that is used to determine the optimal resolution for computation.

In experiments, we demonstrate that our approach is capable of extracting salient points in geometric structure. We evaluated the repeatability of our interest points for various image neighborhood sizes and viewpoint changes. The experiments indicate that our interest points are appropriate for description and matching.

In on-going work, we further improve the run-time of our interest point detector. Since all operations can be mapped to a parallel processing architecture, we will implement our detector on GPU, for instance using the CUV library [10]. We will also develop efficient descriptors for our interest points. Finally, our approach should find application for mapping, scene understanding, and object recognition purposes.

Acknowledgment

This work has been supported partially by grant BE 2556/2-3 of German Research Foundation (DFG).

References

[1] R. Unnikrishnan and M. Hebert. Multi-scale interest regions from unorganized point clouds. In *Workshop on Search in 3D, IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[2] M. Pauly, R. Keiser, and M. Gross. Multi-scale feature extraction on point-sampled surfaces. In *Eurographics*, 2003.

[3] J. Novatnack and K. Nishino. Scale-dependent 3D geometric features. In *Proc. of the IEEE Int. Conf. on Computer Vision (ICCV)*, 2007.

[4] T.-W. R. Lo and J. P. Siebert. Local feature extraction and matching on range images: 2.5D SIFT. *Computer Vision and Image Understanding*, 113(12), 2009.

[5] J. Novatnack and K. Nishino. Scale-dependent/invariant local 3D shape descriptors for fully automatic registration of multiple sets of range images. In *Proc. of the 10th European Conference on Computer Vision (ECCV)*, 2008.

[6] A. Flint, A. Dick, and A. van den Hengel. Thrift: Local 3D structure recognition. In *Proc. of the 9th Biennial Conf. of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications*, 2007.

[7] B. Steder, G. Grisetti, and W. Burgard. Robust place recognition for 3D range data based on point features. In *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2010.

[8] B. Steder, R. B. Rusu, K. Konolige, and W. Burgard. NARF: 3D range image features for object recognition. In *Workshop on Defining and Solving Realistic Perception Problems in Personal Robotics at the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2010.

[9] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, (2):91, 2004.

[10] CUV library, February 2011. http://github.com/deeplearningais/CUV.