In Proceedings of Spatial Cognition 2012, Seeon, Germany, September 2012.

SURE: Surface Entropy for Distinctive 3D Features

Torsten Fiolka[1], Jörg Stückler[2], Dominik A. Klein[3],
Dirk Schulz[1], and Sven Behnke[2]

[1] Fraunhofer Institute for Communication, Information Processing and Ergonomics FKIE, Wachtberg, Germany
torsten.fiolka@fkie.fraunhofer.de, dirk.schulz@fkie.fraunhofer.de
[2] Autonomous Intelligent Systems Group, University of Bonn, Germany
stueckler@ais.uni-bonn.de, behnke@cs.uni-bonn.de
[3] Intelligent Vision Systems Group, University of Bonn, Germany
kleind@iai.uni-bonn.de

Abstract. In this paper, we present SURE features – a novel combination of interest point detector and descriptor for 3D point clouds and depth images. We propose an entropy-based interest operator that selects distinctive points on surfaces. It measures the variation in surface orientation from surface normals in the local vicinity of a point. We complement our approach by the design of a view-pose-invariant descriptor that captures local surface curvature properties, and we propose optional means to incorporate colorful texture information seamlessly. In experiments, we compare our approach to a state-of-the-art feature detector in depth images (NARF) and demonstrate similar repeatability of our detector. Our novel pair of detector and descriptor achieves superior results for matching interest points between images and also requires lower computation time.

Keywords: Depth image interest points, local shape-texture descriptor

1 Introduction

Interest points paired with a descriptor of local image context provide a compact representation of image content. They can be used in various applications such as image registration [15, 13, 21], robot simultaneous localization and mapping (SLAM) [27], panorama stitching [3], photo tourism [7], as well as place [29] and object recognition [5, 18, 31].

Many applications require that a detector repeatedly finds interest points across images taken from various view poses and under differing lighting conditions. Since the scale of surface regions in the image depends on the distance of the sensor from the observed surface, the detector must also retrieve a repeatable scale if distance is not directly measured. This scale can then be used to normalize the size of the image region in which local image context is described.

Descriptors, on the other hand, are designed to distinguish well between different shapes and textures. They are often judged in terms of precision-recall

relations [17]. However, one must admit that descriptor distinctiveness depends clearly on the variety of shapes and textures that appear at the selected interest points. Thus, a detector is preferable that finds interest points in various structures and highly expressive regions.

In this paper, we propose a new approach for extracting shape features at surface points through a measure of surface entropy (SURE). Our features combine a novel pair of interest point detector and local context description. Our approach can be applied to depth images as well as unorganized 3D point clouds. An entropy-based interest measure selects points on surfaces that exhibit strong local variation in surface orientation. We complement our approach by the design of a descriptor that captures local surface curvature properties. We also propose means to incorporate color and texture cues into the descriptor when RGB information is available for the points. We implement both detector and descriptor to process point clouds efficiently on a CPU. Our approach extracts features at a frame rate of about 5 Hz from RGB-D images at VGA resolution.

In experiments, we measure the repeatability of our interest points under view pose changes for several scenes and objects. We compare our approach with state-of-the-art detectors and demonstrate the advantages of our approach. We also assess the distinctiveness of our descriptor and point out differences to state-of-the-art methods.

2 Related Work

2.1 Interest Point Detection

Feature detection and description has been a very active area of research for decades. The computer vision community extensively studies detectors in intensity images. Nowadays, interest point detection algorithms are designed to be invariant against moderate scale and viewpoint changes [35]. There is not a single method that is always best in every application, but some noteworthy stick out from the bulk: The Harris-Affine [16] detector that recognizes corner structures based on the second moment matrix, the MSER [15] detector that identifies groups of pixels that are best separable from their surrounding, and the well known SIFT [14] or optimized SURF [1] detectors that are based on intensity blobs found by a difference of Gaussians filter. One recent example is the SFOP [6] detector for combination of corners, junctions, and blob-like features from a spiral model.

Most related to our method, also the entropy measure based on image intensities has been investigated for interest point detection [10, 11]. It has been successfully applied to object recognition [5] due to the high informativeness of maximum entropy regions. Lee and Chen [12] picked up this idea of features based on histogram distributions and extended it to intensity gradients and color. They used the Bhattacharyya coefficient to identify local distributions that distinguish themselves most from the surrounding. Both approaches are not capable of real-time processing. In our approach, we adopted the entropy measure for 3D

normal orientations in order to get stable-placed features determined by multiple surfaces.

However, those methods purely based on intensity image data suffer problems emerging from projective reduction to 2D space. Moreels and Perona [18] evaluated affine detectors for recognition of conspicuously shaped 3D objects and found out that none "performs well with viewpoint changes of more than 25-30°".

With the steadily increasing availability of depth measuring sensors, recently various methods have been developed to extract interest points from dense, full-view point clouds. The notion of scale has a different interpretation in 3D data. It now depicts the 3D extent of a structure which has been only intrinsic to the scale in 2D images. In depth images, the 2D projection of a structure at a specific 3D scale still varies with distance to the sensor. Few approaches have been proposed that detect interest points at multiple 3D scales and that automatically select a scale for which an interest point is maximally stable w.r.t. repeatability and localization.

Pauly et al. [22], for example, measure surface variation at a point by considering the eigenvalues of the local sample covariance. Novatnack et al. [20] extract multi-scale geometric interest points from dense point clouds with an associated triangular connectivity mesh. They build a scale-space of surface normals and derive edge and corner detection methods with automatic scale selection. For depth images [20], they approximate geodesic distances by computing shortest distances between points through the image lattice. Surface normals are computed by triangulating the range image. Our approach does not require connectivity information given by a mesh. Unnikrishnan et al. [37] derive an interest operator and a scale selection scheme for unorganized point clouds. They extract geodesic distances between points using disjoint minimum spanning trees in a time-consuming pre-processing stage. They present experimental results on full-view point clouds of objects without holes. In [32], this approach has been applied to depth images and an interest detector for corners with scale selection has been proposed. Steder et al. [29] extract interest points from depth images without scale selection, based on a measure of principal curvature which they extent to depth discontinuities. However, our approach is not restricted to depth images and can be readily employed for full-view point clouds.

2.2 Local Descriptors

The SIFT-descriptor [14] has been successfully used in computer vision applications. It describes the local gradient pattern in spatial histograms of gradient magnitudes and orientations. It is made rotation-invariant by aligning the histograms to the dominant gradient orientation at the interest point.

Several improvements to the SIFT descriptor have been proposed. SURF [1] sums Haar wavelet responses as a representation of the local gradient pattern. Recently, Calonder et al. [4] and Rublee et al. [24] demonstrated that binarized pixel comparisons at randomly distributed sample points yield a robust and highly efficient descriptor that outperforms SIFT or SURF.

Other approaches do not solely focus on gradient descriptions of texture. Shape Contexts [2], for instance, build a histogram of contour points in the local neighborhood of a point. Tuzel et al. [36] propose to use covariance of feature values in local image regions as a descriptor.

Johnson and Hebert [9] introduce spin-images to describe local shape context in 3D point clouds. In this approach, cylindrical coordinates of the local point distribution are described in a 2D image-like histogram. The surface normal at an interest point is chosen as the cylindrical axis, and the polar angle is neglected to project the points into 2D.

Shape Context [8, 19] has been extended to 3D in order to describe the distribution of points in log-polar histograms. Tombari et al. [34] extract a local reference frame at a point and extract histograms of normals. In [33] they extend their approach to also capture the distribution of color. However, this method strongly depends on the stability of the reference frame.

Rusu et al. [26] quantify local surface curvature in rotation-invariant Fast Point Feature Histograms (FPFH). They demonstrate that the histograms can well distinguish between shapes such as corners, spheres, and edges.

Steder et al. [29] proposed the NARF descriptor for depth images. They determine a dominant orientation from depth gradients in a local image patch and extract radial depth gradient histograms. In conjunction with the NARF detector, Steder et al. [30] applied this descriptor for place recognition.

3 Entropy-based Interest Points in 3D Point Clouds

3.1 Interest Points of Local Surface Entropy

Our detector is based on statistics about the distribution of local surface normals. We are interested in regions of maximal diversely oriented normals, since they show promise to be stably located at transitions of multiple surfaces or capture entire (sub-)structures that stick out of the surroundings. To identify such regions, we measure the entropy

$$H(X_{\mathcal{E}}) = -\sum_{x \in X_{\mathcal{E}}} p(x) \log p(x), \tag{1}$$

where $X_{\mathcal{E}}$ is a random variable characterizing the distribution of surface normal orientations occurring within a region of interest $\mathcal{E} \subseteq \mathbb{R}^3$. We extract interest points where this entropy measure achieves local maxima, i.e. where $X_{\mathcal{E}}$ is most balanced.

Entropy Computation from Point Clouds Depth sensors usually measure surfaces by a set of discrete sample points $Q = \{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n\}, \boldsymbol{q}_k \in \mathbb{R}^3$. We approximate the surface normal at a sample point $n(\boldsymbol{q}_k)$ looking at the subset of neighboring points $\mathcal{N}_k = \{\boldsymbol{q}_l \in Q | \, \|\boldsymbol{q}_k - \boldsymbol{q}_l\|_1 < r\}$ within a given support range r. Then, $\hat{n}_r(\boldsymbol{q}_k)$ equals the eigenvector corresponding to the smallest eigenvalue of the sample covariance matrix $\mathrm{cov}(\mathcal{N}_k)$.

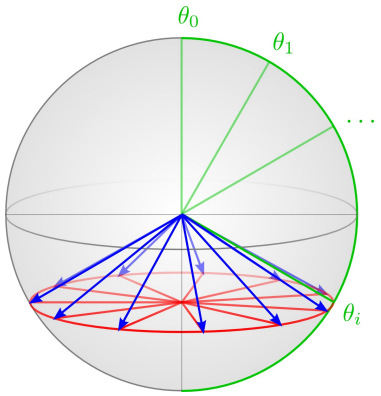

Fig. 1: Construction of an approx. uniform sphere partition. Green: equidistant inclination angles; red: sphere to cone section $C(\theta_i)$ and a_{θ_i} equidistant azimuth angles; blue: resulting orientation vectors on inclination level θ_i.

We discretize the surface normal distribution $X_{\mathcal{E}}$ by use of an orientation histogram in which we count the occurrences of surface normal orientations for a spherical surface partition. We follow the approach by Shah [28], subdividing the spherical surface into approximately equally sized patches. Those are specified by their centrical azimuth and inclination angles. To achieve an uniform decomposition of the sphere, we first choose t equidistant inclination angles $\theta_i = \frac{\pi i}{t}, i \in \{0, \dots, t-1\}$. Then, for each of these inclination angles, we calculate a number of

$$a_{\theta_i} := \lfloor 2\ t \sin(\theta_i) + 1 \rfloor \propto C(\theta_i) \tag{2}$$

equidistant azimuth angles. This way, the sample density in azimuth is proportional to the circumference $C(\theta_i)$ of the section of the sphere with a cone of inclination θ_i. Transforming from spherical into Cartesian coordinates, we obtain a set of normalized vectors $\boldsymbol{v}_{i,j}$ pointing to the centers of histogram bins. Figure 1 depicts the construction of these vectors.

Each estimated surface normal at a point $\boldsymbol{q}_m \in Q \cap \mathcal{E}$ contributes to the histogram bin $x_{i,j}$ with a weight

$$w_{i,j} = \begin{cases} 0 & \text{, if } \hat{n}_r(\boldsymbol{q}_m) \cdot \boldsymbol{v}_{i,j} < \cos\alpha \\ \frac{\hat{n}_r(\boldsymbol{q}_m) \cdot \boldsymbol{v}_{i,j} - \cos\alpha}{1 - \cos\alpha} & \text{, else} \end{cases}, \tag{3}$$

where α denotes the maximal angular range of influence. Finally, we normalize the histogram before calculating the surface normal entropy according to Equation 1.

3.2 Efficient Implementation using an Octree

For efficient data access and well-ordered computation, we set up an octree structure containing the 3D point cloud inferred from the RGB-D image given by

the sensor. In order to measure local surface entropy, our octree enables uniform sampling in 3D space. Furthermore, we exploit the multi-resolution architecture of the octree for fast volume queries of point statistics.

An octree organizes points from a 3D space into cubic subvolumes that are connected in a tree. The root node of the tree spans a volume chosen to fit the extent of the data. Edges between parent and child nodes represent a subset-relation. Each parent node branches into eight children constituting a partition of the parent's volume into eight equally sized octants. This branching is repeated iteratively until a predefined resolution, that equals a maximum depth of the tree, is reached.

The multi-scale structure of the octree allows for efficient bottom-up integration of data, facilitating the calculation of histograms, as well as search queries for local maxima in arbitrary volumes. In each node, we store histogram, integral and maximum statistics for different attributes of all points that are located within the volume of the node. These values can be computed efficiently by passing the attributes of points on a path from leave nodes to the root of the tree. This direction, every parent node accumulates and merges data received from its child nodes.

When querying for statistics inside an arbitrary 3D volume, we recursively descend the tree: if a node is fully inside the queried volume, its statistics are integrated into the response; if it is completely outside, this branch is discontinued; otherwise its child nodes are examined the same way. This is valid since each node already integrates the data of all leaves below in its own statistics. An easily understood example for data statistics is the average position of points within a certain volume $\mathcal{V}$. By integrating over the homogeneous coordinates of points $\boldsymbol{s} = (x, y, z, w)^T = \sum_{\boldsymbol{q}_i \in \mathcal{V}} (x_i, y_i, z_i, 1)^T$, one retains the mean via normalization $\bar{\boldsymbol{q}} = \frac{1}{w}\boldsymbol{s}$.

3.3 Interest Point Detection

The surface normal entropy function depends on two scale parameters: one is the radius r of vicinity $\mathcal{N}$ for the estimation of a surface normal orientation; the other is the extend of a region of interest $\mathcal{E}$, where the distribution of normals and thus the local surface entropy is gathered. These volumes are chosen to be cubic and appropriate to fit the intrinsic octree resolutions. The maximal depth ($\widehat{=}$resolution) of the octree is usually determined by the normal sampling interval at the finest scale that is specified to be a common multiple of the other dimensions. This way, range queries are processed most efficiently. Usually, sampling interval sizes of surface normals as well as normal orientation histograms are set to be at least half of the diameter of their respective local support volume.

All these parameters have to be chosen carefully. The histogram scale $\mathcal{E}$ corresponds directly to the size of the interest points, at which local structures become salient. Its sampling interval is a trade-off between preciseness and speed. According to the Nyquist-Shannon sampling theorem, a minimal sampling frequency of twice the region size is needed to reconstruct the surface entropy

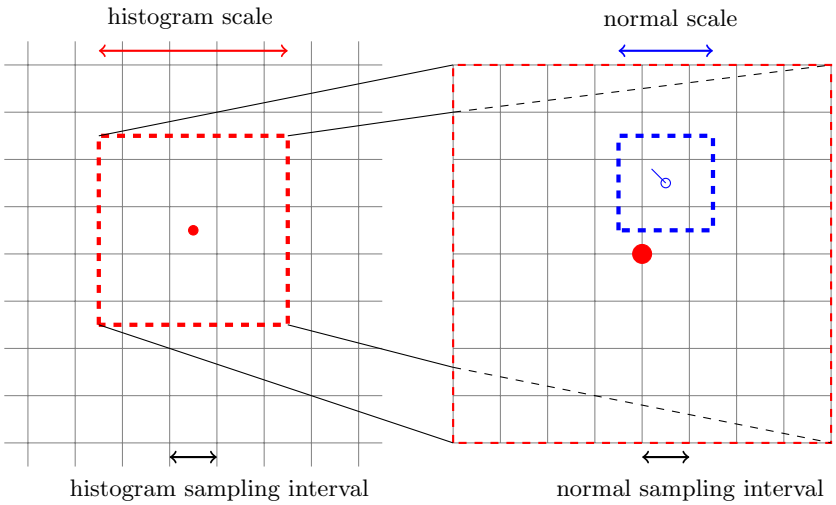

Fig. 2: Scheme of the different parameters for calculating normals and entropy

function, i.e. not to miss the occurrence of a local maximum. We choose the normal scale r to a constant fraction of the histogram scale. Accordingly, the sampling interval for normals must also obey the sampling theorem. Reproducing the effect of a lowpass filter for removal of artifacts, we consider an entropy sample to be an interest point candidate, if it exceeds all its spatial neighbors within a dominance region. In addition, the candidate is only kept if it exceeds a global entropy threshold $H_{\min}$. The latter is checked, because noisy sensor data, image borders, and depth jumps occasionally induce interest point candidates on planar surfaces.

While surface entropy along an ideal ridge would be constant in theory, sensor noise and discretization artifacts will induce spurious measurements at these structures and thus cause local maxima of surface entropy. Such interest point candidates should be filtered out by inspection of the local prominence, since their position is loose in one dimension. Inspired by cornerness measures from image based interest point operators, we test for a considerable variance of surface entropy in all directions. First, we compute the local center of surface entropy mass within the region $\mathcal{E}_{\boldsymbol{q}}$ around a sample point $\boldsymbol{q}$

$$\mu_H(\mathcal{E}_{\boldsymbol{q}}) := \frac{1}{\sum_{\boldsymbol{q}_i \in \mathcal{E}_{\boldsymbol{q}}} H(X_{\mathcal{E}_{\boldsymbol{q}_i}})} \sum_{\boldsymbol{q}_i \in \mathcal{E}_{\boldsymbol{q}}} H(X_{\mathcal{E}_{\boldsymbol{q}_i}})\, \boldsymbol{q}_i. \tag{4}$$

Then, the sample covariance matrix of local surface entropy mass equals to

$$\mathrm{cov}_H(\mathcal{E}_{\boldsymbol{q}}) := \frac{1}{\sum_{\boldsymbol{q}_i \in \mathcal{E}_{\boldsymbol{q}}} H(X_{\mathcal{E}_{\boldsymbol{q}_i}})} \sum_{\boldsymbol{q}_i \in \mathcal{E}_{\boldsymbol{q}}} H(X_{\mathcal{E}_{\boldsymbol{q}_i}}) \left((\boldsymbol{q}_i - \mu_H(\mathcal{E}_{\boldsymbol{q}}))(\boldsymbol{q}_i - \mu_H(\mathcal{E}_{\boldsymbol{q}}))^T\right). \tag{5}$$

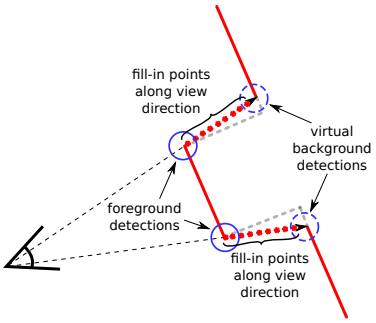

Fig. 3: Occlusion handling. In depth images, structure may be occluded (dashed gray). At depth discontinuities, we therefore add artificial measurements (red dots) from foreground towards the background. Any "virtual background" detections are discarded, since they are not stable w.r.t. view point changes.

By decomposition of $\mathrm{cov}_H(\mathcal{E}_{\boldsymbol{q}})$ we derive the eigenvalues λ_1, λ_2, and λ_3 sorted by value in ascending order. Finally, our local prominence check is defined

$$P(\mathcal{E}_{\boldsymbol{q}}) = \frac{\lambda_1}{\lambda_3} \geq P_{\min}, \tag{6}$$

where we used $P_{\min} = 0.15$ in our experiments.

Improved Localization After identification of interest point candidates, the true maximum location has to be recovered from the discretized surface entropy function. Starting from a candidate's location, we apply the mean-shift mode searching approach: We integrate surrounding surface entropy samples via a Gaussian window in order to estimate the gradient of the surface entropy density. Then, the position of the candidate is shifted along this gradient direction. This procedure is repeated up to three times.

Occlusion Handling in Depth Images Surface entropy is supposed to be high where multiple different layers join together. In depth images, however, one cannot always measure all joining surfaces explicitly due to occlusions, resulting in a reduced entropy. This peculiarity of the measuring system should be compensated. Therefore, we detect jump edges in the depth image. Since we know that there must exist another hidden surface behind each foreground edge, we approximate it by adding artificial measurements in viewing direction up to a distance that meets the biggest used local entropy scale (cf. Fig. 3). While we use such points for the detection of interest points, we do not include this artificial information into the descriptor. We also discard detected interest points in the background at occlusions, since they are not stable w.r.t. view point changes.

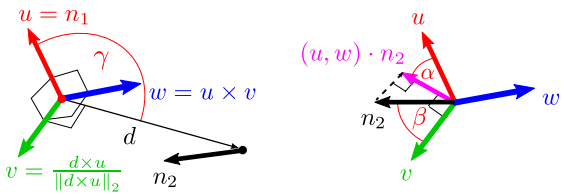

Fig. 4: Surfel pair relations describe rotation-invariant relative orientations and distances between two surfels.

4 Local Shape-Texture Descriptor

Since our surface entropy measure detects interest points at location where the surface exhibits strong local variation, we design a shape descriptor that captures local surface curvature. When RGB information is available, we also describe the local texture at an interest point. We aim at a rotation-invariant description of the interest points in order to match features despite of view pose changes. For each individual cue, we select a reasonable distance metric and combine them in a distance measure for the complete feature.

4.1 Shape

Surfel pair relations (see Fig. 4) have been demonstrated to be a powerful feature for describing local surface curvature [38, 26]. Given two surfels (q_1, n_1) and (q_2, n_2) at points q_1 and q_2 with surface normals n_1 and n_2, we first define a reference frame (u, v, w) between the surfels through

$$\begin{aligned} u &:= n_1, \\ v &:= \frac{d \times u}{\|d \times u\|_2}, \text{ and} \\ w &:= u \times v, \end{aligned} \tag{7}$$

where $d := q_2 - q_1$. In this frame, we measure relative angles and distances between the surfels by

$$\begin{aligned} \alpha &:= \arctan2\left(w \cdot n_2, u \cdot n_2\right), \\ \beta &:= v \cdot n_2, \\ \gamma &:= u \cdot \frac{d}{\|d\|_2}, \text{ and} \\ \delta &:= \|d\|_2 . \end{aligned} \tag{8}$$

By construction, surfel pair relations are rotation-invariant and, hence, they can be used for a view-pose invariant description of local shapes.

In order to describe curvature in the local vicinity of an interest points, we build histograms of surfel pair relations from neighboring surfels (see Fig. 5).

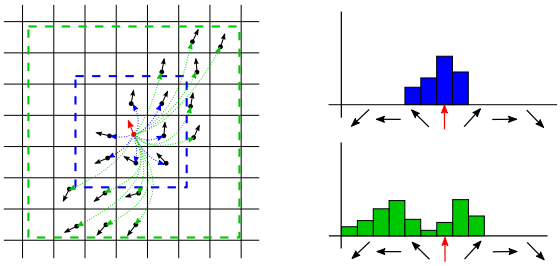

Fig. 5: Shape descriptor in a simplified 2D example. We build histograms of surfel pair relations from the surfels in a local neighborhood at an interest point. We relate surfels to the central surfel at the interest point. Histograms of inner and outer volumes capture distance-dependent curvature changes.

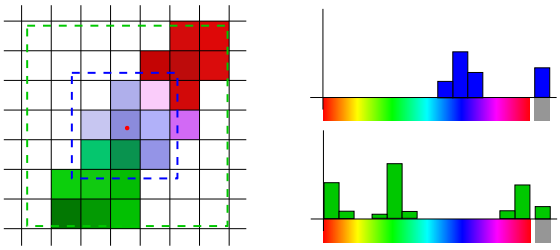

Fig. 6: Color descriptor. We extract hue and saturation histograms in an inner and outer local volume at an interest point.

Each surfel is related to the surfel at the interest point being the reference surfel (p_1, n_1). We discretize the angular features into 11 bins each, while we use 2 distance bins to describe curvature in inner and outer volumes. We choose the support size of the descriptor in proportion to the histogram scale.

4.2 Color

A good color descriptor should allow interest points to be matched despite illumination changes. We choose the HSL color space and build histograms over hue and saturation in the local context of an interest point (see Fig. 6). Our histograms contain 24 bins for hue and one bin for unsaturated, i.e., "gray", colors. Each entry to a hue bin is weighted with the saturation s of the color. The gray bin receives a value of $1 - s$. In this way, our histograms also capture information on colorless regions.

Similar to the shape descriptor, we divide the descriptor into 2 histograms over inner and outer volumes at the interest point. In this way, we measure the spatial distribution of color but still retain rotation-invariance.

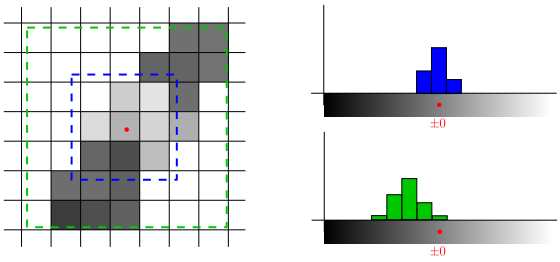

Fig. 7: Luminance descriptor. We describe luminance differences towards the interest point in histograms over local inner and outer volumes.

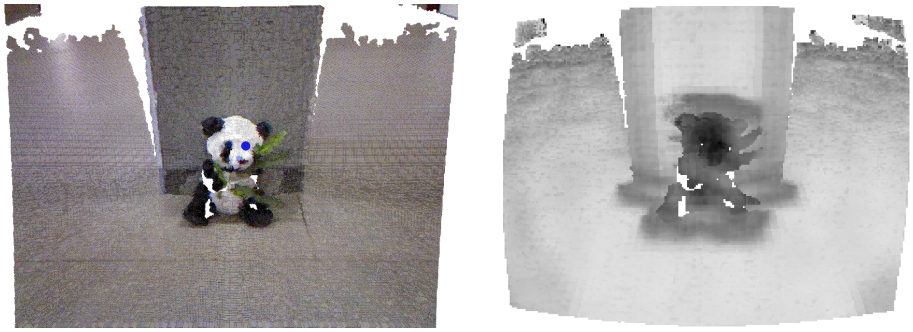

Fig. 8: Shape similarity w.r.t. the marked point (blue dot) measured using the Euclidean distance on our shape descriptors.

4.3 Luminance

Since the color descriptor cannot distinguish between black and white, we propose to quantify the relative luminance change towards the color at the interest point (see Fig. 7). By this, our luminance descriptor is still invariant to ambient illumination. We use 10 bins for the relative luminance and, again, extract 2 histograms in inner and outer volumes.

4.4 Measuring Descriptor Distance

The character of the individual components of our descriptor suggests different kinds of distance metrics. We combine the distances $d_s(q_1, q_2)$, $d_c(q_1, q_2)$, and $d_l(q_1, q_2)$ between two points q_1 and q_2 using the arithmetic mean

$$d(q_1, q_2) := \frac{1}{3} \sum_{i \in \{s,c,l\}} d_i(q_1, q_2). \tag{9}$$

Shape Distance: For the shape descriptor, we use the Euclidean distance as proposed for FPFH features in [26]. We measure the arithmetic mean of the

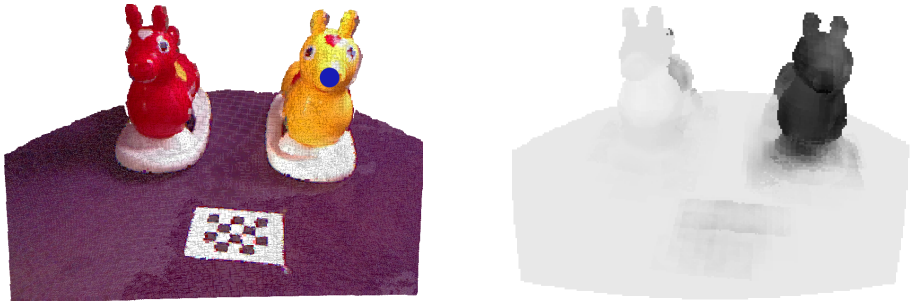

Fig. 9: Color similarity w.r.t. the marked point (blue dot) measured using the saturated Earth Mover's Distance ($\widehat{\text{EMD}}$) on our color descriptors.

Euclidean distance of the angular histograms in the inner and outer volumes. Fig. 8 illustrates this distance measure in an example scene.

Color Distance: Since the HSL color space is only approximately illumination invariant, the domains of our color histograms may shift and may slightly be misaligned between frames. Hence, the Euclidean distance is not suitable. Instead, we apply an efficient variant of the Earth Mover's Distance (EMD, [25]) which has been shown to be a robust distance measure on color histograms.

The EMD between two histograms P and Q measures the minimum amount of mass in a histogram that needs to be "moved" between the histograms to equalize them. Formally, the EMD is defined as

$$\text{EMD}(P,Q) = \frac{\min_{f_{ij}} \sum_{i,j} f_{ij} d_{ij}}{\sum_{ij} f_{ij}}, \tag{10}$$

where f_{ij} is the flow and d_{ij} is the ground distance between the bins P_i and Q_j. Pele and Werman [23] propose $\widehat{\text{EMD}}$, a modified EMD with saturated ground distance that is applicable to unnormalized histograms. They demonstrate that the $\widehat{\text{EMD}}$ can be implemented several magnitudes faster than the standard EMD but still retains its benefits. In our application, we saturate the ground distances at a distance of two bins. Fig. 9 illustrates our color distance in an example.

Luminance Distance: We also use the saturated $\widehat{\text{EMD}}$ to compare luminance histograms. See Fig. 10 for an example of our distance measure.

5 Experiments

5.1 Experiment Setup

We evaluate our approach on RGB-D images from a Microsoft Kinect and compare it with the NARF interest point detector and descriptor. We recorded

Fig. 10: Luminance similarity w.r.t. the marked point (blue dot) measured using the saturated Earth Mover's Distance ($\widehat{\text{EMD}}$) on our luminance descriptors.

dataset	SURE 640x480	NARF 640x480	NARF 320x240	NARF 160x120
box	0.19	160.18	1.95	0.27
rocking horses	0.2	133.36	3.25	0.36
teddy	0.2	164.43	2.09	0.26
clutter	0.2	179.20	3.24	0.27

Table 1: Average run-time in seconds per frame for SURE and NARF detection and feature extraction.

4 scenes, 3 containing objects of various size, shape, and color, and one cluttered scene with many objects in front of a wall. The objects are a box (ca. 50x25x25 cm), toy rocking horses (height ca. 1 m), and a teddy bear (height ca. 20 cm). Image sequences with 80 to 140 VGA images (640×480 resolution) have been obtained by moving the camera around the objects. We estimate the ground truth pose of the camera using checkerboard patterns laid out in the scenes. Furthermore, we evaluate the NARF descriptor on three resolutions of the datasets, at the original 640×480 and downsampled 320×240 and 160×120 resolutions. In each image of a sequence, we extract interest points on 3 histogram scales (SURE) or support sizes (NARF). We chose the scales 12, 24, and 48 cm.

5.2 Repeatability of the Detector

We assess the quality of our interest point detector by measuring its repeatability across view-point changes. We distinguish here between "simple repeatability" and "unique repeatability". Table 2 shows the average number of interest points found by the detectors. SURE finds a similar amount of features like NARF on 160×120 resolution.

We associate interest points between each image pair in the sequence using the ground truth transform. Each interest point can only be associated once to

dataset	SURE 640x480	NARF 640x480	NARF 320x240	NARF 160x120
box	11.8	32.5	18.2	14.9
rocking horses	35.2	121.6	72.4	44.6
teddy	6.8	43.0	26.9	15.3
clutter	47.5	93.4	48.4	26.5

Table 2: Average number of interest points for the SURE and NARF detectors.

an interest point in the other image. We establish the mutually best correspondences according to the Euclidean distance between the interest points. Valid associations must have a distance below the histogram scale (SURE) or support size (NARF) of the interest point. "Unique repeatability" only accepts an association between interest points, if the match is unambiguous. This means, that the matched interest points must be the only possible match within the support size/histogram scale, otherwise the association is discarded.

From Fig. 11 we see that SURE and NARF yield similar repeatability on the box and the teddy datasets. The NARF detector shows here a better performance in the smaller resolutions, while performing worse in full resolution. On the rocking horses and the cluttered scene, SURE performs worse than NARF. However, about 50% resp. 25% of the interest points are still matchable across 90° view angle change. In Fig. 13 SURE performs better than NARF in terms of "unique repeatability". The NARF detector allows several interest points being "close" to each other, i.e., in a distance smaller than their respective support sizes. A SURE interest point will be discarded if it lies within the histogram scale of another interest point and its entropy is lower compared to its neighbor. In that way, we ensure that a SURE interest point sticks out of its environment and can be uniquely matched by descriptor.

In Fig. 12 we also demonstrate the effect of our occlusion handling mechanism. If no artificial points are added along depth discontinuities, repeatability drops earlier with view angle change which is naturally expected.

5.3 Matching Score of the Descriptor

We also evaluate the capability of the detector-descriptor pair for establishing correct matches between images. We define the matching score as the fraction of interest points that can be correctly matched between images by the descriptor.

The results in Fig. 14 clearly demonstrate that SURE performs better than NARF in matching individual interest points. Its descriptor does not seem to be distinctive enough to reliably find correct matches. SURE, however, focuses on prominent local structure that is well distinguishable with our descriptor.

We also evaluate the matching score of the individual descriptor components of SURE in Fig. 15. In the teddy scene, very little color is present and the shape descriptor dominates color and lumincance. The clutter scene shows that the combination of these three descriptors performs considerably better than each of the descriptors alone.

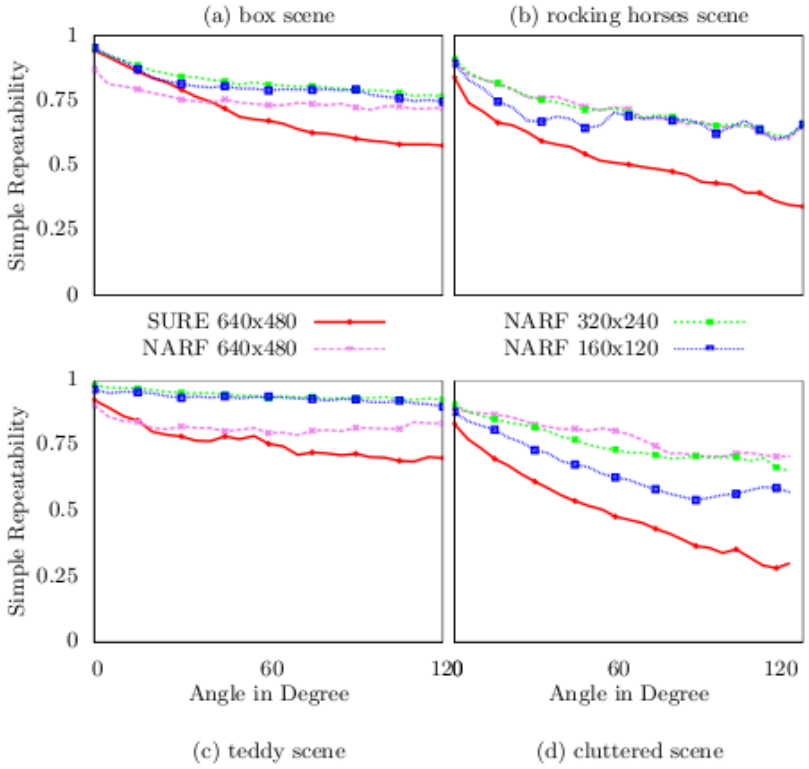

Fig. 11: Simple Repeatability in four different scenes comparing the SURE detector and the NARF detector. The NARF detector was applied in three different resolutions.

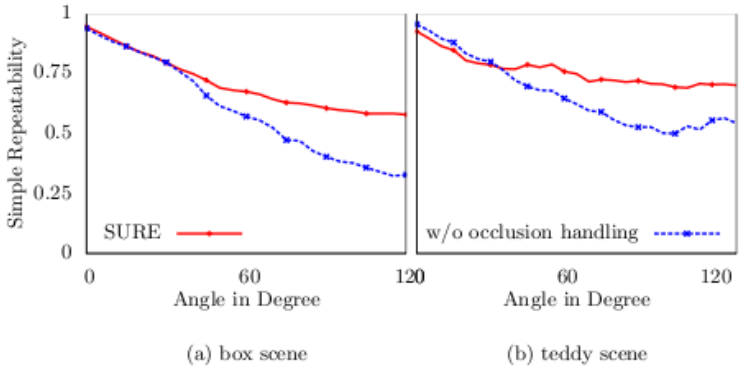

Fig. 12: Effect of occlusion handling on the repeatability of SURE.

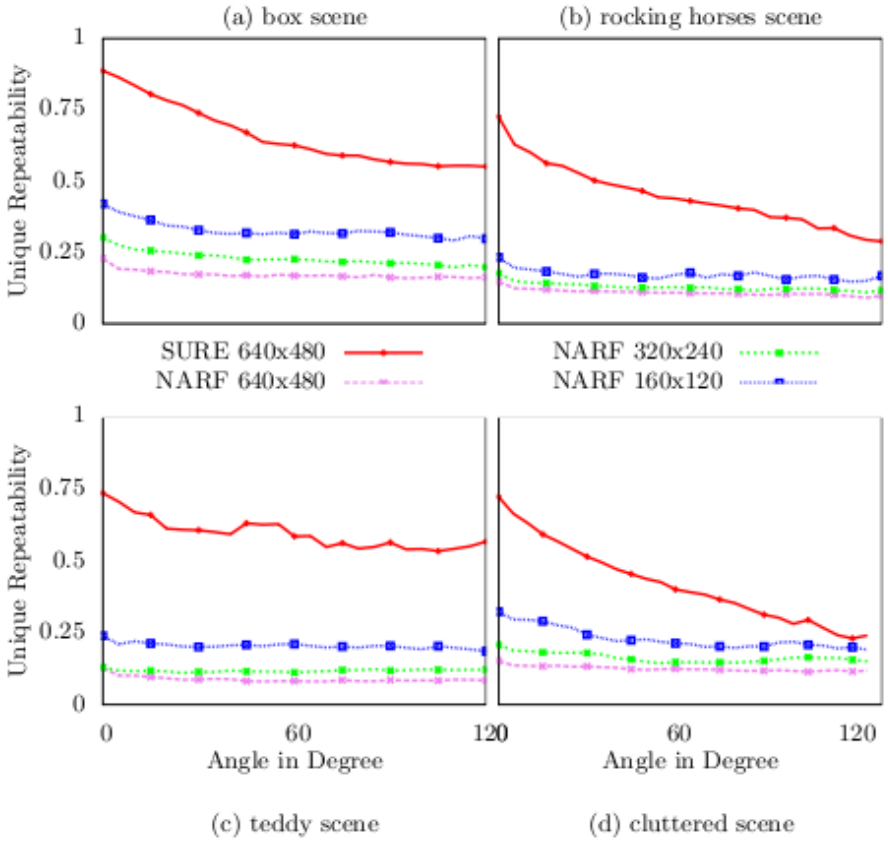

Fig. 13: Unique repeatability in four different scenes comparing the SURE detector and the NARF detector. Unique repeatability only accepts an association between interest points, if the match is unambiguous. This means, that the matched interest points must be the only possible match within the support size/histogram scale, otherwise the association is discarded.

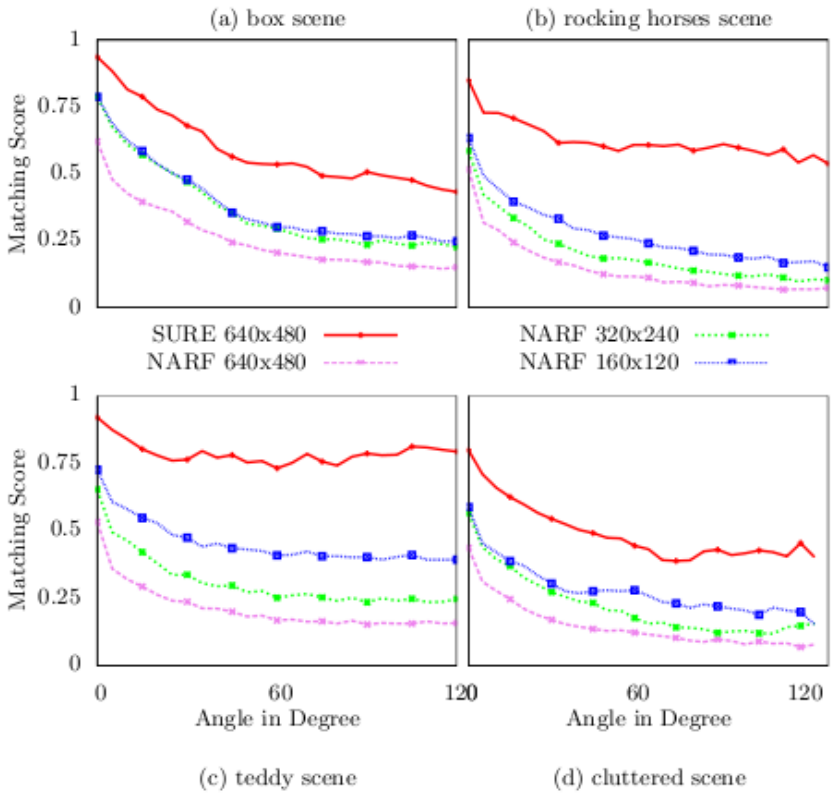

Fig. 14: Matching Score comparing SURE Feature Descriptor with the NARF Descriptor on four datasets.

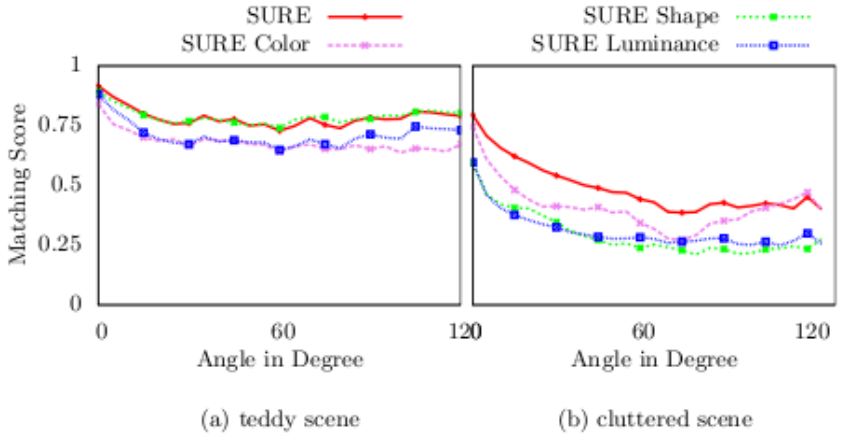

Fig. 15: Matching Score comparing the different SURE Descriptors.

5.4 Run-Time

Table 1 shows the run-time of NARF and SURE (detection and feature extraction). SURE outperforms NARF clearly on any of the processing resolutions of NARF, while SURE makes full use of the available data.

6 Conclusions

We proposed SURE, a novel pair of interest point detector and descriptor for 3D point clouds and depth images. Our interest point detector is based on a measure of surface entropy on normals that selects points with strong local surface variation. We designed a view-pose-invariant descriptor that quantifies this local surface curvature using surfel pair relations. When RGB information is available in the data, we also incorporate colorful texture information into the SURE descriptor. We describe color and luminance in the HSL space and measure distance using a fast variant of the Earth Mover's Distance to gain an illumination-invariant description at the interest point.

In experiments, we could demonstrate that the SURE detector achieves similar repeatability like the NARF descriptor. When matching features by descriptor, our SURE features outperform NARF regarding matching score. SURE also performs faster than NARF on 640×480 images.

In future work, we will further improve the run-time of SURE on depth and RGB-D images by exploiting the connectivity information in the image. We will also investigate automatic scale selection to further improve the repeatability and localization of the interest points.

References

1. Bay, H., Tuytelaars, T., Gool, L.V.: Surf: Speeded up robust features. In: Proc. of European Conference of Computer Vision (ECCV) (2006)
2. Belongie, S., Malik, J., Puzicha, J.: Shape context: A new descriptor for shape matching and object recognition. In: In NIPS. pp. 831–837 (2000)
3. Brown, M., Lowe, D.: Automatic panoramic image stitching using invariant features. Int'l Journal of Computer Vision 74, 59–73 (2007)
4. Calonder, M., Lepetit, V., Strecha, C., Fua, P.: BRIEF: Binary Robust Independent Elementary Features. In: European Conference on Computer Vision (2010)
5. Fergus, R., Perona, P., Zisserman, A.: Object class recognition by unsupervised scale-invariant learning (2003)
6. Förstner, W., Dickscheid, T., Schindler, F.: Detecting interpretable and accurate scale-invariant keypoints. In: 12th IEEE International Conference on Computer Vision (ICCV'09). Kyoto, Japan (2009)
7. Frahm, J.M., Fite-Georgel, P., Gallup, D., Johnson, T., Raguram, R., Wu, C., Jen, Y.H., Dunn, E., Clipp, B., Lazebnik, S., Pollefeys, M.: Building rome on a cloudless day. In: Proc. of European Conference on Computer Vision (ECCV) (2010)
8. Frome, A., Huber, D., Kolluri, R., Blow, T., Malik, J.: Recognizing objects in range data using regional point descriptors. In: Proc. of European Conference on Computer Vision (ECCV) (2004)

9. Johnson, A., Hebert, M.: Using spin images for efficient object recognition in cluttered 3d scenes. Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 21(5), 433 –449 (may 1999)
10. Kadir, T., Brady, M.: Saliency, scale and image description. Int'l J. of Computer Vision 45(2), 83–105 (2001)
11. Kadir, T., Zisserman, A., Brady, M.: An affine invariant salient region detector. In: Proc. of European Conf. of Computer Vision (ECCV) (2004)
12. Lee, W.T., Chen, H.T.: Histogram-based interest point detectors. In: Int'l Conf. on Computer Vision and Pattern Recognition (CVPR) (2009)
13. Lo, T.W.R., Siebert, J.P.: Local feature extraction and matching on range images: 2.5D SIFT. Computer Vision and Image Understanding 113(12) (2009)
14. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision (2), 91 (2004)
15. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. In: Proc. of the British Machine Vision Conference (2002)
16. Mikolajczyk, K., Schmid, C.: Scale & affine invariant interest point detectors. Int'l Journal of Computer Vision (2004)
17. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transactions of Pattern Analysis and Machine Intelligence 27(10) (2005)
18. Moreels, P., Perona, P.: Evaluation of feature detectors and descriptors based on 3d objects. Int'l Journal of Computer Vision 73, 263–284 (2007)
19. Mori, G., Belongie, S., Malik, J.: Efficient shape matching using shape contexts. Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 27(11), 1832 –1837 (nov 2005)
20. Novatnack, J., Nishino, K.: Scale-dependent 3D geometric features. In: Proc. of the IEEE Int. Conf. on Computer Vision (ICCV) (2007)
21. Novatnack, J., Nishino, K.: Scale-dependent/invariant local 3D shape descriptors for fully automatic registration of multiple sets of range images. In: Proc. of European Conference on Computer Vision (ECCV) (2008)
22. Pauly, M., Keiser, R., Gross, M.: Multi-scale feature extraction on point-sampled surfaces. In: Eurographics (2003)
23. Pele, O., Werman, M.: Fast and robust earth mover's distances. In: Proc. of the Int. Conference on Computer Vision (ICCV) (2009)
24. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: An efficient alternative to SIFT or SURF. In: International Conference on Computer Vision (2011)
25. Rubner, Y., Tomasi, C., Guibas, L.J.: The earth mover's distance as a metric for image retrieval. Int. J. of Computer Vision 40, 99–121 (2000)
26. Rusu, R.B., Blodow, N., Beetz, M.: Fast point feature histograms (fpfh) for 3d registration. In: The IEEE Int. Conf. on Robotics and Automation (ICRA) (2009)
27. Se, S., Lowe, D., Little, J.: Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. Int'l Journal of Robotics Research 21(8), 735–758 (August 2002)
28. Shah, T.R.: Automatic reconstruction of industrial installations using point clouds and images. Ph.D. thesis, TU Delft (2006)
29. Steder, B., Grisetti, G., Burgard, W.: Robust place recognition for 3D range data based on point features. In: Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA) (2010)

30. Steder, B., Ruhnke, M., Grzonka, S., Burgard, W.: Place recognition in 3D scans using a combination of bag of words and point feature based relative pose estimation. In: Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS) (2011)
31. Steder, B., Rusu, R.B., Konolige, K., Burgard, W.: 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)
32. Stückler, J., Behnke, S.: Interest point detection in depth images through scale-space surface analysis. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA) (2011)
33. Tombari, F., Salti, S., di Stefano, L.: A combined texture-shape descriptor for enhanced 3d feature matching. In: Proc. of the IEEE International Conference on Image Processing (ICIP) (2011)
34. Tombari, F., Salti, S., Stefano, L.D.: Unique signatures of histograms for local surface description. In: Proc. of the 11th European Conference on Computer Vision (ECCV) (2010)
35. Tuytelaars, T., Mikolajczyk, K.: Local invariant feature detectors: A survey. Foundations and Trends in Computer Graphics and Vision 3(3), 177–280 (2007)
36. Tuzel, O., Porikli, F., Meer, P.: Region covariance: A fast descriptor for detection and classification. In: Proc. of European Conference on Computer Vision (ECCV). vol. 3952, pp. 589–600 (2006)
37. Unnikrishnan, R., Hebert, M.: 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)
38. Wahl, E., Hillenbrand, G., Hirzinger, G.: Surflet-pair-relation histograms: a statistical 3D-shape representation for rapid classification. In: Proc. of the Int. Conf. on 3-D Digital Imaging and Modeling (2003)