In Proceedings of 6th European Conference on Mobile Robots (ECMR), Barcelona, Spain, 2013.

Distinctive 3D Surface Entropy Features for Place Recognition

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

***Abstract*— In this paper, we present a variant of SURE, an interest point detector and descriptor for 3D point clouds and depth images and use it for recognizing semantically distinct places in indoor environments. The SURE interest operator selects distinctive points on surfaces by measuring the variation in surface orientation based on surface normals in the local vicinity of a point. Furthermore SURE includes a view-pose-invariant descriptor that captures local surface properties and incorporates colored texture information. In experiments, we compare our approach to a state-of-the-art feature detector in depth images (NARF). Finally, we evaluate the use of SURE features for recognizing places and demonstrate its advantages.**

***Index Terms*— surface interest points, local shape-texture descriptor, place recognition**

I. Introduction

Interest points paired with a descriptor of local image context provide a compact representation of image content. Applications such as place or object recognition require that a detector repeatably finds interest points across images taken from various view poses and under differing lighting conditions. Descriptors, on the other hand, are designed to distinguish well between different shapes and textures. 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 will be preferable, if it selects interest points in various structures and highly expressive regions.

We introduced SURE in [1] and propose now a variant and its use for place recognition. SURE uses an entropy-based interest measure to select points on surfaces that exhibit strong local variation in surface orientation. Its descriptor captures local surface curvature properties as well as color and texture cues in case RGB information is available for the points.

In experiments, we measure repeatability of our interest points under view pose changes for several scenes and objects and compare our approach with a state-of-the-art detector and descriptor to demonstrate advantages of our approach. We show that SURE is capable of correctly recognizing the semantic label of scenes with a bag-of-words approach. The top row in Fig. 1 gives a short idea how the interest point detection works while the bottom row outlines the place recognition application with SURE features.

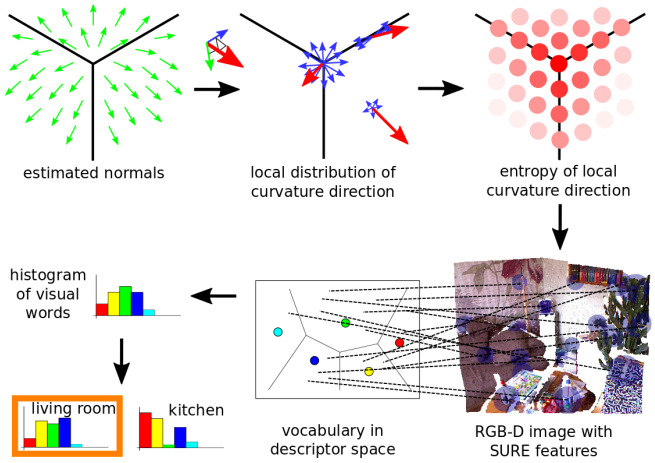


Fig. 1: We detect SURE features in depth images at locations with locally prominent surface curvature. Our interest operator measures the entropy of the distribution of curvature directions at a point in a local neighborhood (top row). The curvature direction (blue arrow) is obtained by the cross product between the estimated surface normal at the point of interest (red arrows) and at neighboring points (green arrows). We propose a descriptor that captures local shape and colored texture at interest points. We recognize places using a bag-of-words approach using SURE features (bottom row).

II. Related Work

A. Interest Point Detection

Feature detection and description has been a very active area of research since decades. Most related to our method, also the entropy measure based on image intensities has been investigated for interest point detection [2], [3], [4]. It has been successfully applied to object recognition [5] due to the high informativeness of maximum entropy regions.

However, those methods purely based on intensity image data suffer problems emerging from projective reduction to 2D space [6]. Recently, various methods have been developed to extract interest points from dense, full-view point clouds.

Novatnack et al. [7] extract multi-scale geometric interest points from dense point clouds with an associated triangular connectivity mesh. Our approach does not require connectivity information given by a mesh. Unnikrishnan et al. [8] derive an interest operator and a scale selection scheme for

[1] Fraunhofer Institute for Communication, Information Processing and Ergonomics FKIE, Wachtberg, Germany torsten.fiolka at fkie.fraunhofer.de, dirk.schulz at fkie.fraunhofer.de

[2] Autonomous Intelligent Systems, Computer Science Institute VI, University of Bonn, Germany stueckler at ais.uni-bonn.de, behnke at cs.uni-bonn.de

[3] Intelligent Vision Systems, Computer Science Institute III, University of Bonn, Germany kleind at iai.uni-bonn.de

unorganized point clouds. They extract geodesic distances between points using disjoint minimum spanning trees in a time-consuming pre-processing stage. In [9], this approach has been applied to depth images and an interest detector for corners with scale selection has been proposed. Steder et al. [10] 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.

B. Local Descriptors

Several point descriptors have been proposed for 3D point clouds and depth images. Prominent examples are spin-images [11], shape context [12], [13], and (C-)SHOT [14]. Steder et al. [10] 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. We directly compare our approach to NARF and demonstrate that SURE finds more distinct features. Rusu et al. [15] 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. We also base our descriptor on surfel-pair relations to describe shape. We complement this descriptor with colored texture cues.

C. Place Recognition

Bag-of-words approaches are frequently used for place recognition purposes [16], [17]. The image content is first compressed into a set of features that are then quantized using a vocabulary of visual words. The words represent clusters in descriptor space that are learned from training images. The histogram of word occurrences in an image is then compared to the training images in order to retrieve the place category. We also follow a bag-of-words approach and demonstrate the use of SURE features for place recognition.

III. Entropy-based Interest Points in 3D Point Clouds

A. Interest Points of Local Surface Entropy

Our detector is based on statistics about the distribution of local surface orientations. We are interested in regions with maximal diversely oriented surfaces, 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 discrete random variable characterizing the distribution of surface orientations occurring within a region of interest $\mathcal{E} \subseteq \mathbb{R}^3$ and $p(x)$ is a probability mass function on $X_{\mathcal{E}}$. We extract interest points where this entropy measure achieves local maxima, i.e. where $X_{\mathcal{E}}$ is most balanced.

B. Estimation of Surface Normals and Curvature Directions

As we make use of surface normals to estimate surface orientations, we shortly introduce our approach to estimate them. Depth sensors usually measure surfaces by a set of discrete sample points $Q = \{\vec{q}_1, \ldots, \vec{q}_n\}, \vec{q}_k \in \mathbb{R}^3$. We approximate the surface normal at a sample point $n(\vec{q}_k)$ from the subset of neighboring points $\mathcal{N}_k = \{\vec{q}_l \in Q | \, \|\vec{q}_k - \vec{q}_l\|_1 < r\}$ within a given support range r. Then, $\hat{n}_r(\vec{q}_k)$ equals the eigenvector corresponding to the smallest eigenvalue of the sample covariance matrix $\mathrm{cov}(\mathcal{N}_k)$.

We discretize the domain of weighted curvature directions into a fixed number of orientation bins and a center bin for parallel surfel-pairs. For the discretization of the orientation distribution in a given region of the surface we estimate an orientation histogram. We use the approach by Shah [18] for subdividing a spherical surface into approximately equally sized patches. Every patch is specified by its central azimuth and inclination angle. The number of azimuth angles is determined by the inclination angle so that it is proportional to the circumference of the section of the sphere. We transform the angles from spherical into Cartesian coordinates and obtain a set of normalized vectors $\vec{v}_{i,j}$ pointing to the centers of histogram bins.

C. Entropy Calculation

Our original approach used normals directly to estimate surface entropy. Each estimated surface normal at a point $\vec{q}_m \in Q \cap \mathcal{E}$ contributes to the histogram bin $x_{i,j}$ with a weight

$$w_{i,j} = \begin{cases} 0 & \text{, if } \hat{n}_r(\vec{q}_m) \cdot \vec{v}_{i,j} < \cos\alpha \\ \frac{\hat{n}_r(\vec{q}_m) \cdot \vec{v}_{i,j} - \cos\alpha}{1 - \cos\alpha} & \text{, else} \end{cases}, \tag{2}$$

where α denotes the maximal angular range of influence.

When directly measuring surface variation from the orientation of surface normals, measurement noise may cause spurious detections at ridges and edges. Such false detections were filtered out using heuristic post-processing [1]. Now we propose to measure the variation in surface curvature by the cross-product of pairs of surface normals. This second-order statistics exhibit strong orientation peaks and therefore low entropy for ridge- and edge-like structures. In contrast, at structures with diverse surface normal orientations such as corners, curvature directions will also point in diverse directions and entropy will be high. On planar surfaces, curvature direction is strongly affected by noise and, hence, the entropy is meaningless. Nevertheless, the planarity of the surface is indicated by the similarity of the normal orientations. We thus add a new class for parallel normal pairs and re-weight the curvature direction according to the scalar product of the normals.

For every sample point $\vec{q}_k$ we collect all estimated normals $\hat{N}_{\vec{q}_k}$ in the neighborhood $\mathcal{N}$ and calculate the main normal $\hat{n}(\mathcal{E})$, which incorporates all points in $\mathcal{E}$. We calculate curvature directions by the normalized cross products $c_k = (\hat{n}(\mathcal{E}) \times \hat{n}_r(\vec{q}_k)) \,/\, \|\hat{n}(\mathcal{E}) \times \hat{n}_r(\vec{q}_k)\|_2$ between the main normal and the neighboring normals. In order to handle planar

surfaces, we assign a weight

$$w_\times = (1 - \langle \hat{n}(\mathcal{E}), \hat{n}_r(\vec{q}_k) \rangle) \tag{3}$$

which indicates the likelihood of the normal pair belonging to a planar surface or not.

A normalized vector $\vec{v}$ contributes to the orientation histogram bin $h_{i,j}$ with a weight inversely proportional to its distance from the center of a histogram bin, i. e.,

$$w_{i,j} = \begin{cases} 0 & , \text{ if } \langle \vec{v}, \vec{v}_{i,j} \rangle < \cos\alpha \\ \frac{\langle \vec{v}, \vec{v}_{i,j} \rangle - \cos\alpha}{1 - \cos\alpha} & , \text{ otherwise.} \end{cases} \tag{4}$$

The maximal angular range of influence is bounded by α. In addition, we weight each entry in the orientation histogram with the unplanarity weight $w_\times$.

The center bin receives the weight $1 - w_\times$ for each normal-pair. Finally, we normalize the complete histogram before calculating the entropy according to Equation 1.

D. Efficient Implementation using Octrees

For efficient data access and well-ordered computation, we set up an octree structure containing the 3D point data from the depth image or point cloud. 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.

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 the root of the tree to the corresponding leaf node. 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 $\vec{s} = (x, y, z, w)^T = \sum_{\vec{q}_i \in \mathcal{V}} (x_i, y_i, z_i, 1)^T$, one retains the mean via normalization $\vec{q} = \frac{1}{w}\vec{s}$.

E. Interest Point Detection

The surface entropy function depends on two scale parameters: one is the radius r of vicinity $\mathcal{N}$ for the estimation of a surface normal orientation (called normal scale); 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 (called histogram scale) as seen in Fig. 2. With these parameters set and normals and their corresponding entropy calculated, we detect interest points by searching for local extrema in entropy. A lower threshold for entropy is used to suppress erroneous detections due to sensor noise. When using normals for entropy calculation, we also suppress detections on ridges by testing for surface variance in all directions.

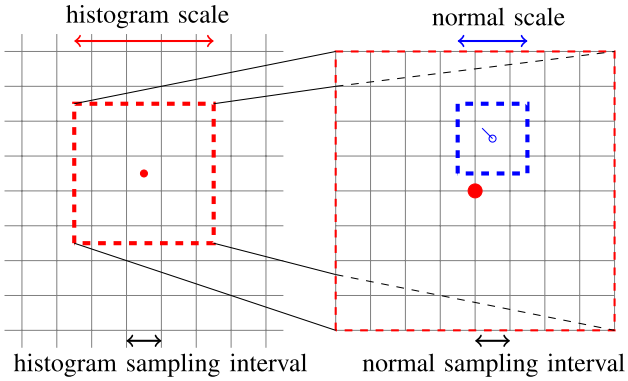


Fig. 2: We calculate normals and entropy histograms at equidistant sample points in fixed scale radii.

F. Improved Localization

Since the detector so far only considers a fixed discretization with an arbitrary global shift, the true maximum location in entropy has not yet been recovered. In order to improve localization, we apply mean-shift starting from a candidate's location: 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. We iterate this procedure up to three times.

G. Occlusion Handling in Depth Images

In depth images, one cannot always measure all joining surfaces explicitly due to occlusions, resulting in a reduced entropy. To compensate this 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 histogram scale. We exclude interest points at partial occlusions in the background, since one cannot make any assumptions on the occluded surfaces.

IV. Local Shape-Texture Descriptor

Since our detector finds interest points at locations where the surface exhibits strong local curvature variation, we design a shape descriptor to capture this distribution. 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.

A. Shape

Surfel-pair relations have been demonstrated to be a powerful feature for describing local surface curvature [19], [15]. In order to describe curvature in the local vicinity of an interest point, we build histograms of surfel-pair relations from neighboring surfels (see Fig. 3). Each surfel is related to the surfel at the interest point being the reference

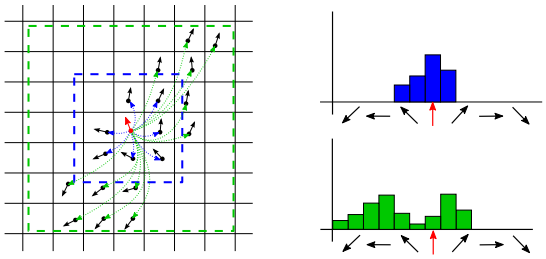


Fig. 3: Shape descriptor in a simplified 2D example. We build histograms of surfel-pair relations from the surfels in a local neighborhood at an interest point. We relate surfels to the central surfel at the interest point. Histograms of inner and outer volumes capture distance-dependent curvature changes.

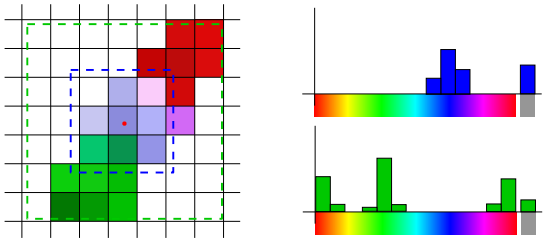


Fig. 4: Color descriptor. We extract hue and saturation histograms in an inner and outer local volume at an interest point.

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.

B. 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. 4). 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.

C. Luminance

Since the color descriptor cannot distinguish between black and white, we propose to quantify luminance contrasts of neighboring points towards the interest point. 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.

D. Measuring Descriptor Distance

The character of the individual components of our descriptor suggests different kinds of distance metrics. For the shape descriptor, we use the Euclidean distance as proposed for FPFH features in [15]. 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, [20]) from [21] which has been shown to be a robust distance measure on color histograms.

	box	rocking horses	teddy	clutter
SURE Normals	18	111.3	10.1	75.9
SURE Crossproducts	18.3	129.2	8.8	85.6
NARF 160x120	20.1	146.1	20.6	49.1
NARF 320x240	25.8	234.9	34.7	100.7
NARF 640x480	57.4	400.5	62	356.2

TABLE I: Average number of features per frame.

V. Place Recognition

We use SURE features for place recognition in indoor environments. Our approach has two stages: For training, we extract SURE features from each frame $f \in F_{train}$ in a training set of images. We apply the k-means algorithm on the descriptors of the SURE features to create a bag-of-words B (BoW) with $k = 400$ visual words. For each frame, we then build similarity histograms $f \in F_{train}$ using the BoW. In the recall stage, we again create similarity histograms for a frame t using the BoW. We determine the 20 best matching frames $M = \{m_1, ..., m_{20}\} \subset F_{train}$ with the training set by comparing histograms through histogram intersection. Afterwards, we directly compare the SURE features $S(t)$ in frame t with the features $S(m_i)$ in each of the 20 best matching frames $m_i \in M$. Our error metric sums up the minimal distance of pairings of each feature in $S(t)$ with all features in $S(m_i)$,

$$D(S(t), S(m_i)) = \sum_{s \in S(t)} \min_{s_M \in S(m_i)} d(s, s_M). \quad (5)$$

Finally, the semantic label of the best matching frame

$$m_{\text{best}} = \arg\min_{m_i \in M} D(S(t), S(m_i)) \quad (6)$$

is chosen as the classification of t.

VI. Experiments

We evaluate SURE features and our place recognition approach on RGB-D images from a Microsoft Kinect sensor and compare SURE with the NARF interest point detector and descriptor as implemented in PCL 1.6 (cf. [22]). In all experiments, we evaluate SURE at full resolution (640×480) while NARF is evaluated at full resolution and two downsampled ones (320×240 and 160×120).

A. Repeatability of the Detector

We assess the quality of our interest point detector by measuring its repeatability across view-point changes. We recorded 4 scenes, 3 containing objects of various size, shape, and color, and one cluttered scene with many objects in front

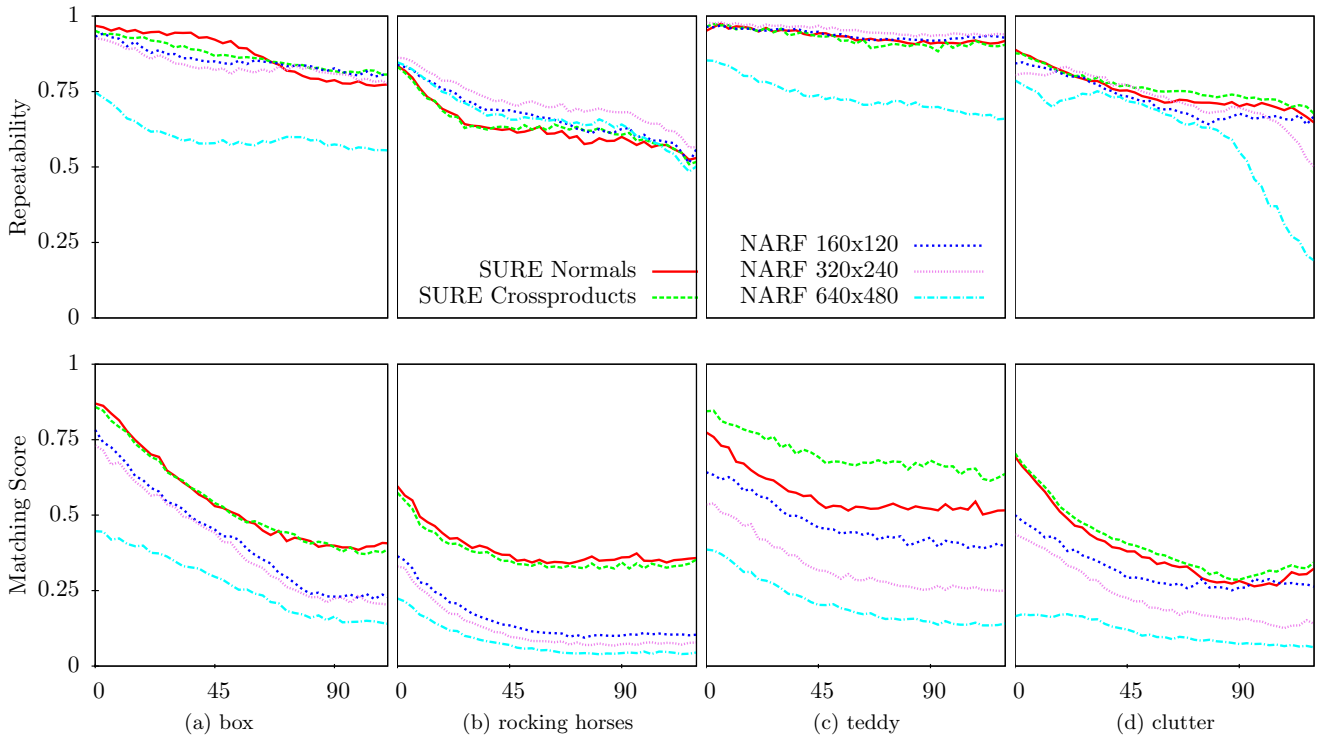
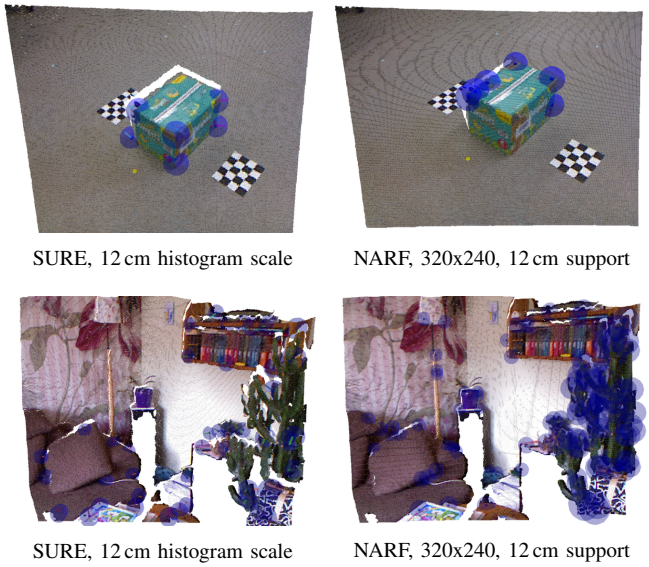


Fig. 6: Repeatability and matching score of SURE and NARF features in different scenes under view-point change (x-axis, in degrees). Top row: Repeatability where each interest points is mapped to the closest interest point in the second image. Bottom row: matching repeatability of the descriptors referring to the closest interest points in the second image.



SURE, 12 cm histogram scale

NARF, 320x240, 12 cm support

SURE, 12 cm histogram scale

NARF, 320x240, 12 cm support

Fig. 5: Examples of detected interest points on a box (top) and in a living room scene (bottom).

of a wall. The objects are a box (ca. 50x25x25 cm^3), 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. In each image of a sequence, we extract interest points on 4 histogram scales (SURE) or support sizes (NARF). We chose the scales 12, 24, 36, and 48 cm.

Since the original SURE detected very few features we lowered the influence radius of each feature to half its radius, matching to the default value for NARF. Table I depicts the average number of features for each detector and scene.

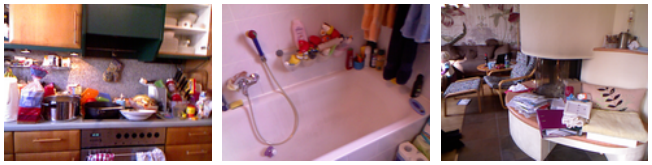
We then associate interest points between each image pair in the sequence using the ground truth transform. Each interest point can only be associated once to an interest point in the other image. We establish the mutually best correspondences according to the Euclidean distance between the interest points. Valid associations must have a distance below the histogram scale (SURE) or support size (NARF) of the interest point.

The repeatability measure (Fig. 6, top row) shows comparable perfomance of both SURE versions and NARF, except for NARF in full resolution. The SURE Crossproducts approach gains a little advantage over SURE Normals with increasing view-point change.

B. 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 descriptor.

The results in the bottom row of Fig. 6 clearly demonstrate that SURE performs better than NARF in matching individual interest points. The NARF descriptor which only incor-



(a) kitchen scene (b) bathroom scene (c) living room scene

Fig. 9: Example images of the second location (3 rooms).

dataset	method	correct	avg. run-time
3 rooms	NARF 160x120	63,9%	1.6 sec
	NARF 320x240	63,7%	16.4 sec
	NARF 640x480	36,0%	523.9 sec
	SURE Normals	92,9%	0.9 sec
	SURE Crossproducts	92,1%	5.5 sec
5 rooms	NARF 160x120	42,8%	1.1 sec
	NARF 320x240	32,5%	14.5 sec
	NARF 640x480	21,9%	677.9 secs
	SURE Normals	79,3%	0.8 sec
	SURE Crossproducts	84,9%	5.1 sec

TABLE II: Place recognition results.The average run-time includes the time needed for creating the features and the place recognition task averaged per frame.

porates shape information does not seem to be distinctive enough to reliably find correct matches. SURE focuses on prominent local structure that is well distinguishable with our descriptor, and it can take advantage of color and luminance information.

C. Place Recognition Results

For the evaluation of our place recognition approach, we recorded two locations with multiple scans of different rooms with the Kinect. The first location contains five different rooms (a living room, bathroom, kitchen, bedroom and corridor) with approx. 200 frames overall, the second location contains three different rooms with approx. 500 frames (see Fig. 9). For every location two sets were recorded to gain independent data for training and testing.

The results are displayed in Table II. Our approach classifies at least 80% of frames correctly, clearly outperforming NARF in recognition rate and run-time. The SURE Crossproducts approach is clearly better in recognition rate than the other approaches in the second dataset. While it is still faster than NARF 320x240 and 640x480, it demands more computational time than SURE Normals.

VII. Conclusions

We proposed a variant for SURE, a pair of interest point detector and descriptor for 3D point clouds, and applied it for place recognition. SURE is based on a measure of surface entropy on normals that selects points with strong local surface variation. Its view-pose-invariant descriptor quantifies this local surface curvature using surfel-pair relations and also incorporates colored texture, if available.

In experiments, we could demonstrate that the SURE descriptor outperforms NARF in matching corresponding features. SURE also performs faster than NARF on 640×480 images. We could demonstrate that SURE features are well suited for place recognition using a bag-of-words approach.

In future work, we will investigate the application of our place recognition approach for visual loop-closure detection in SLAM.

References

[1] T. Fiolka, J. Stückler, D. A. Klein, D. Schulz, and S. Behnke, "Sure: Surface entropy for distinctive 3d features," in *Spatial Cognition*, 2012, pp. 74–93.

[2] T. Kadir and M. Brady, "Saliency, scale and image description," *International Journal of Computer Vision*, vol. Volume 45 Issue 2, pp. 83–105, 2001.

[3] T. Kadir, A. Zisserman, and M. Brady, "An affine invariant salient region detector," in *8th European Conference on Computer Vision*. Springer-Verlag, 2004, pp. 228–241.

[4] W.-T. Lee and H.-T. Chen, "Histogram-based interest point detectors," in *Proc. of the Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[5] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in *Proc. of the Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2003.

[6] P. Moreels and P. Perona, "Evaluation of feature detectors and descriptors based on 3D objects," *Int. Journal of Computer Vision*, 2007.

[7] J. Novatnack and K. Nishino, "Scale-dependent 3d geometric features," in *International Conference of Computer Vision*, 2007.

[8] R. Unnikrishnan and M. Hebert, "Multi-scale interest regions from unorganized point clouds," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[9] J. Stückler and S. Behnke, "Interest point detection in depth images through scale-space surface analysis," in *Proc. of the IEEE Int. Conference on Robotics and Automation (ICRA)*, 2011.

[10] B. Steder, R. B. Rusu, K. Konolige, and W. Burgard, "Narf: 3d range image features for object recognition," in *International Conference on Intelligent Robots and Systems*, 2010.

[11] A. E. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3d scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, pp. 433–449, May 1999.

[12] A. Frome, D. Huber, R. Kolluri, T. Blow, and J. Malik, "Recognizing objects in range data using regional point descriptors," in *Proc. of European Conference on Computer Vision (ECCV)*, 2004.

[13] G. Mori, S. Belongie, and J. Malik, "Efficient shape matching using shape contexts," *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 27, no. 11, pp. 1832 –1837, nov. 2005.

[14] F. Tombari, S. Salti, and L. di Stefano, "A combined texture-shape descriptor for enhanced 3D feature matching." in *Proc. of the IEEE Int. Conference on Image Processing (ICIP)*, 2011.

[15] Rusu, Blodow, and Beetz, "Fast point feature histograms (fpfh) for 3d registration," in *In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, May 2009.

[16] M. Cummins and P. Newman, "Appearance-only SLAM at large scale with FAB-MAP 2.0," *Int. Journal of Robotics Research*, 2010.

[17] B. Steder, M. Ruhnke, S. Grzonka, and W. Burgard, "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.

[18] T. R. Shah, "Automatic reconstruction of industrial installations using point clouds and images," *Publications on Geodesy*, vol. 62, 2006.

[19] E. Wahl, U. Hillenbrand, and G. Hirzinger, "Surflet-pair-relation histograms: A statistical 3d-shape representation for rapid classification," in *Proceedings of Forth International Conference on 3-D Digital Imaging and Modeling (3DIM 2003)*, 2003.

[20] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth movers distance as a metric for image retrieval," *International Journal of Computer Vision*, vol. 40, p. 99121, 2000.

[21] O. Pele and M. Werman, "Fast and robust earth mover's distances," in *ICCV*, 2009.

[22] R. B. Rusu and S. Cousins, "3d is here: Point cloud library (pcl)," in *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, May 9-13 2011.