All-Season 3D Object Recognition Challenges

Levente Tamas[1] and Bjoern Jensen[2]

***Abstract*— Both the appearance based place and object recognition is heavily influenced by the weather conditions. This is valid for data acquired with 3D sensors such as stereo cameras or time-of-flight sensors. Both sensors are affected by the changes of the scene lightning conditions as well as the other weather factors such as rain or snow. The real benefits-pitfalls of the object classification algorithms can be evaluated on the data recorded in different illumination or weather conditions. The focus on this paper is on the benchmarking of the different 3D feature-descriptors in a object recognition pipeline in order to highlight advantages of currently implemented algorithms. The datasets were recorded in various weather conditions including cloudy daytime, night time with moonlight illumination, nigh time with snow and daytime with sun. These objects were captured with simultaneously from a similar viewpoint with a stereo camera and a time-of flight sensor.**

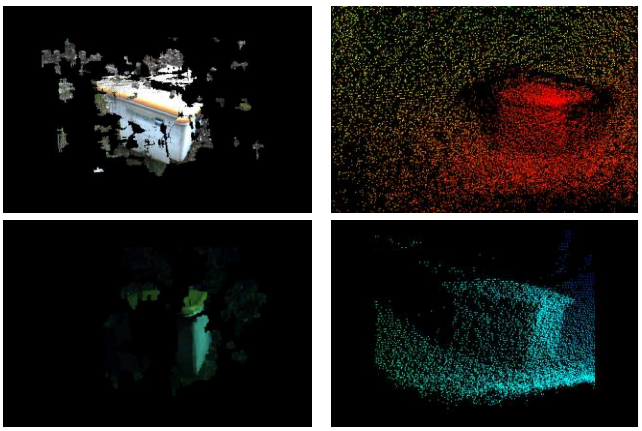
I. INTRODUCTION

The 3D visual perception in an autonomous robotics application is a crucial in order to acquire real-time dense enough information from the environment. A key component of the perception is the scene understanding and the object recognition which servers as the basis for several mapping and navigation algorithms [1]. The use of 3D perception sensors in the mobile robotics application became popular during the last few years mainly due to the affordable camera solutions. According to this idea the focus on the object recognition based on feature descriptors got into the focus [2], [3], [4], [5].

While in the 2D perception the problem of robustness and invariance to the weather changes is already tackled in several papers including [6], [7], [8], [9], in the 3D perception field only a limited number of publications tackle similar problems focusing mainly on indoor data [10]. Similar concepts to the 2D perception such as the popular bag-of-words approach [1] is applied for RGB-D data in the work of [11] and the popular SIFT [12] keypoint counterparts for RGB-D data is available the challenging problems due to the variations in the scene illumination or weather conditions are not generally solved. Further more, the different sensors react in a different way to the changing weather conditions, as this is shown in Figure 1. This example contains the same box shaped object captured with stereo camera and mesa time of flight (ToF) camera in sunny and moonlight illumination conditions. These differences are due to the way that the physical signals are received with the two sensors: while the strong lighting is an important disturbance factor for the ToF sensor, this is essential for the good quality stereo images. These differences are also valid in case of extreme weather conditions such as the night-snow or day-rain conditions tested during the data acquisition phase.

In this paper we propose a thoughtful analyses for the already implemented feature descriptors in the PCL [13] library available with a BSD type license for different RGB-D and XYZ-I data recorded in different weather conditions for object recognition purposes [14], [15], [16]. This analyze is focused on the data acquired with a Swiss Ranger 4000 (SR4K) time-of-flight camera and a Bumblebee2 (BB2) stereo camera for outdoor objects with size ranging from $0.5m-3m$ and $1m-5m$ in depth. For performance metrics the receiver operating characteristic (ROC) [17] curve is considered based on the true positives and false positives during the recognition tests.

In the first part of the paper we present the data preprocessing as well as object recognition pipeline. The object recognition test-bench proposed in this paper contains two phases: an off-line training for extracting and storing the characteristics of the object, and an on-line testing in which the extracted features of an object is searched within the database constructed in the previous phase. In the next section the descriptors considered during the analyses phase are shortly described. The robustness test results are summarized in the section IV. Finally, the paper is concluded with the overall test results performed on different recorded datasets.



(a) Stereo camera image (b) Mesa ToF image

Fig. 1: Test object observed with stereo camera (a) and ToF camera (b) at daytime and night time(best viewed in color)

*This work was supported by Sciex-NMS project nr. 12.239

[1] Technical University of Cluj-Napoca, Robotics Research Group, Dorobantilor st. 71-73, 400609, RO; Levente.Tamas@aut.utcluj.ro

[2] Bern University of Applied Sciences, Institute for Human Centered Engineering, roboticsLab, Quellgasse 21, Biel/Bienne, CH-2502; bjoern.jensen@bfh.ch

II. OBJECT RECOGNITION PIPELINE

In this section the data acquisition and preprocessing steps are presented including the raw data filtering and the object segmentation are necessary as a preliminary step for the object recognition pipeline. Then the details regarding the object recognition based on various types of feature descriptors are presented as well as two different classification algorithms.

A. Object filtering

The first step of the object recognition pipeline was the filtering of the raw data. The main role of this step was to reduce the outliers with a statistical filtering and to get a compact representation of the data with a voxel-grid type filter. Also a pass-through filter was considered in order to cancel out the false readings from the SRK4 sensor which are often present like far shadows for the objects and the elimination of the points from the BB2 outside the disparity range. All these filters are part of the PCL library and were used with standard setups.

In our benchmarks the tuned parameters for the filters were as follows: for the voxel grid we considered a grid size of $0.01m$, which is close to the actual resolution of the sensor, while for the statistical outliers we considered 50 points and with the standard deviation threshold of 0.8.

B. Object segmentation

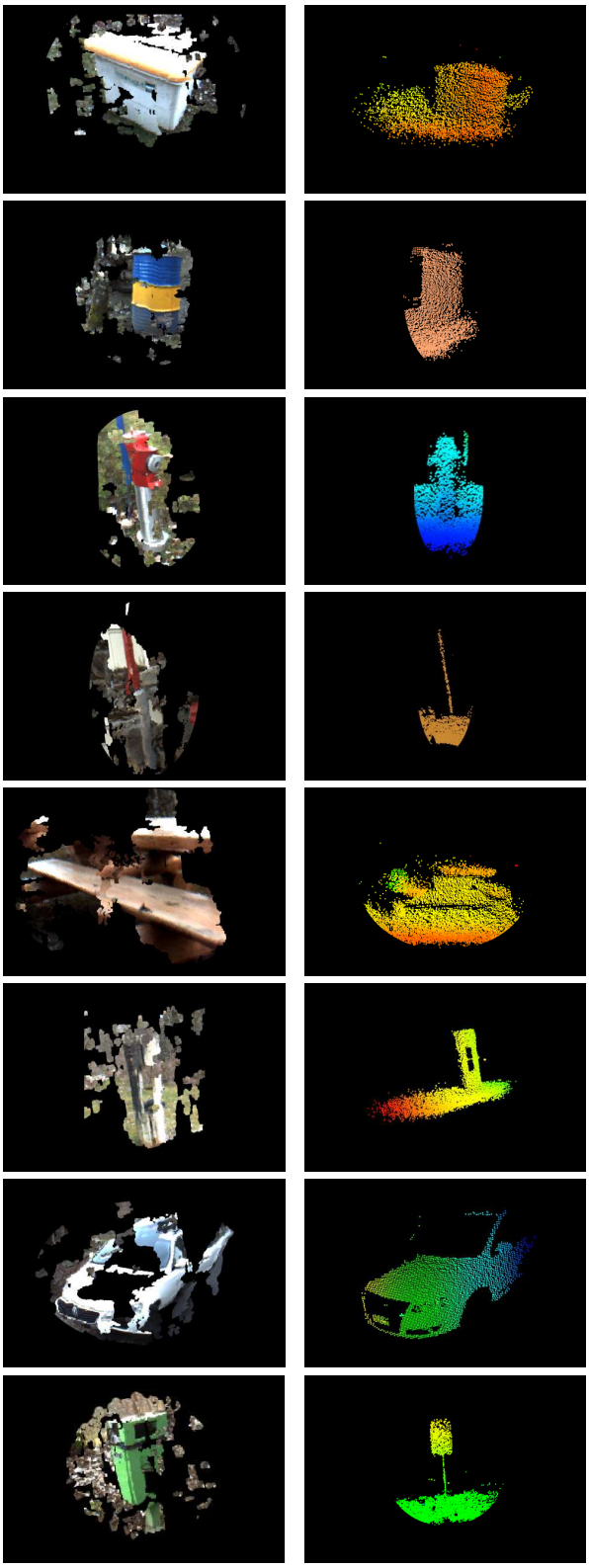
In the next step of the data preprocessing the major planes are segmented out in order to get the objects from the scene. This is an important step in order to end up with the data containing only the objects from a scenario. In order to achieve this, planar models are fitted to the acquired data, and the largest horizontal ground planes up to a tuned percentage are removed. The plane fitting is performed in a standard sampling consensus (SAC) approach, and for the plane removal parameter we set to 30 percentage of the original data.

C. Object recognition with support vector machines

Another alternative for the object feature descriptor classification is the support vector machines (SVM) introduced in [18] with the basic intuition of using hyperplanes to separate the training data in different sets. The general problem formulation makes use of a set of instance-pair labels denoted with $(\mathbf{x}_i, y_i)$ with $i = 1, \ldots, l$ where $\mathbf{x}_i \in R^n$ and $\mathbf{y} \in \{1, -1\}^l$ used to solve the following optimization problem:

$$\min_{\mathbf{w},b,\xi} \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{l}\xi_i \tag{1}$$

subject to the constrains $y_i\left(\mathbf{w}^T\phi(\mathbf{x}_i) + b\right) \geq 1 - \xi_i$. The function ϕ maps the training vector $\mathbf{x}$ to a higher dimensional space, while the C is an error penalty parameter, $\mathbf{w}$ is the normal of the vector to the separating hyperplane and ξ expresses the degree of misclassified points. Further more, for the ϕ function the kernel $K(x_i, x_j) \equiv \phi(x_i)^T\phi(x_j)$ is defined, for which several implementation variants exist. In



(a) (b)

Fig. 2: Test same object capture with stereo camera (a) and time-of-flight camera (b). The objects(in top-down order): box, cylinder, hydrant, stick, table, trunk, car, trash-bin (best viewed in color)

our approach we used the radial basis function defined as $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$ with the tuning parameter γ. Thus, in total two tuning parameters had to be chosen for the implementation: C and γ [19].

This tuning we performed in the training phase using a adaptive grid search technique, i.e. a course grid for the potential optimal value search and a fine grid around this point for further refinement. Before preforming the parameter selection is essential to scale the input data in a common scale in order to get good classification results. Further on, in the prediction phase, the same range scaling has to be done in order to match the train and test feature descriptor data scale. A typical output for the feature data training using the grid search approach is presented in Figure 3 with the logarithmic scale representation for the parameters C and γ.

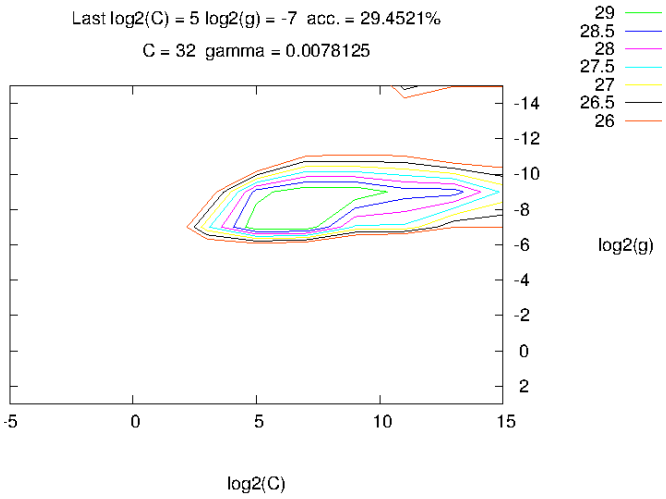


Fig. 3: A typical output of the grid based search space for the parameters C and γ

III. 3D FEATURE DESCRIPTORS

In this section we present briefly the 3D feature descriptors considered for the object recognition pipeline. The depth feature descriptors as well as the image descriptors are compact representation of data. Beside the fact that the data is represented in a compact form, the features tend to contain rich information extracted from the data. Another important general characteristics is the invariance with respect to certain transformations and disturbances. This invariance is essential in order to use them as discriminators in a recognition type applications.

Two main categories of descriptors are distinguished based on the range of data on which are computed: local and global ones [20]. Global descriptors are meant to be used to describe a larger set of data containing objects of interest, and they capture the geometric properties of these objects. Local descriptors in contrast are usually computed on a small subset of representative data (keypoints) for already segmented scenes, and are capturing properties around these keypoints.

A. Rotation invariant feature transform

The rotation invariant feature transform (RIFT) can be applied to 3D data containing intensity information too. Originally it was proposed in the work of [21] as an extension to the SIFT image descriptors [22]. The algorithm iterates over each point P_i in the input point cloud and and within a sphere of radius r all the points are considered to belong to a subset P_{ik}. An imaginary circle with 4 bins (rings) are considered perpendicular to the normal at the point P_i. All neighbors of the selected point are assigned to a ring based on the relative distance based on gradient histogram computed with 8 bins using a thresholding. Thus a total number of 32 histograms are computed with this technique, which describe circular features of a point P_i.

B. Intensity Gradient Estimator

The intensity gradient estimator (IGE) uses as input depth data with intensity information. In the first step the normals of the 3D points are computed, and for each point the analytical intensity gradient is computed along the surface direction of the considered region. This technique is similar to the 2D image intensity computation [23]. In the final step the computed gradient is projected onto the surface, and the output of the estimator is the projection gradient vector containing the intensity information computed for each input point.

C. Intensity Spin Estimator

This type of descriptor is based on the work of [21], however the idea of using intensity information as a descriptor was already present in the earlier work [24]. In contrast to the IGE type of descriptor in this case there is no need for explicit normal pre-computation at the input point cloud, which gives a considerable speed-up for this algorithm. As tuning parameters the point distance and intensity distance bins can be set, having the same meaning as in the case of the RIFT descriptor.

D. Spin Image Estimator

The original idea for the spin image (SI) estimation is presented in the work [25] and can be applied to depth data with pre-computed normals. The algorithm computes two types of distances: the distance of the normals computed at a point and the source normal n and the between the from the considered point along n. The distances larger than a tuning threshold are rejected. From the remaining distance pairs a histogram is built, which represents the occurrence of the discrete distance pairs.

E. Point Feature Histogram

The local point feature histogram (PHF) [26] descriptor extends the original surflet-pair relation histograms suggested in the work of [27]. The input for this feature descriptor is a pointcloud with normals. In the first step for each point P_i the neighbors within a search radius are computed, denoted with sets P_{ik}. Within these sets point pairs are considered denoted with P_s and P_t with the meaning source and target.

For these pairs, the difference of normals are computed and, described with 3 angles around the axis and a distance. As the distance if varying with viewpoint, this can be left out. Finally, these angles are considered to be sorted in the 125 binned histogram, which is the output of the algorithm for each point.

F. Fast Point Feature Histogram

The fast point feature histogram (FPFH) [28] is an extension of the PFH yielding to computationally less expensive variant of the PFH. The major difference between PFH and FPFH is that while in the case of PFH all pairs of points are considered in the subsets P_{ik}, in this case only the point pairs between P_i and the rest of the point within P_{ik} are considered. Thus the computation cost drops from $O(nk^2)$ to $O(nk)$. The three angles in this case are binned into a 11 bin histogram, the total length of the obtained histogram is 33 for each point.

G. Viewpoint Feature Histogram

The global viewpoint feature histogram (VFH) [29] describes the pointcloud P as containing two components: a component representing the viewpoint of the scene and one containing the FPFH features. In this case the FPFH features are binned into a 45 bin histogram, and the distance between the points is also taken into account, thus a total number of 4 extended FPFH features are stored. The additional view point feature is computed by taking the centroid of the pointcloud denoted with P_c and computing the FPFH for each neighbors. The later histogram is represented using 128 bins, thus the total number of bins for a pointcloud is 308 for this descriptor for the entire pointcloud.

H. Clustered Viewpoint Feature Histogram

The clustered viewpoint feature histogram (CVFH) [30] is an extension of the VFH in order to handle occlusion or other types of sensor noise. This is mainly important for the VFH, as in case of an occluded view of the same object the histogram of the descriptor varies considerably. The basic idea of the CVFH is the construction of stable regions (clusters) S which step is done by computing compact regions using the region growing approach with thresholding on the normal values. Once these S regions are computed, the VFH for each of them is determined, and an additional shape distribution (SD) is computed as $SD = \frac{(c-p_i)^2}{sup(c-p_i^2)}$, where c is the centroid of the cluster S and the p_i represents the points from this region. This component is also stored in a binned histogram, the total number of descriptor histogram bins for a single point being equal to 308.

I. Ensemble of Shape Functions

The ensemble of shape functions (ESF) type descriptor was proposed in the work [31] which is based on the A3, D2, D3 shape descriptor functions [32] and extends the D2 type description presented in [33].

The algorithm starts with selecting a subset of 20000 points from the pointcloud, and samples three random points from this P_a, P_b, P_c. The D2 distance is based on the metric distance between the points P_a and P_b. In the next step is verified weather the line connecting the two points are on the surface (in), or out the surface (out) or both (mixed). The corresponding bin for the D2 distance is incremented at the the computed bin. This procedure is repeated for the remaining two point-pairs.

Another histogram, D2 ratio, captures the ratio between the parts of lines lying on the surface and outside the surface. The D3 histogram incorporates the area information for the triangle defined by the three randomly chosen points. The last histogram, A3 contains the angle information for the three points and categorized as in case of D2 into in, out and mixed categories based on the way that they are in space with respect to the surface.

Finally, at the end of each loop the global descriptor with 10 sub-histograms containing 64 bin each are added together, and the total of 640 bins are returned for the pointcloud.

J. Radius-based Surface Descriptor

The Radius-based Surface Descriptor (RSD) [34] has as input an oriented pointcloud, i.e. with normals and describes the local geometry of a point with respect to its neighbors in terms of radii. Each point pair is supposed to lie on a sphere, and the distance of the points d and the angle between the normals at the two points has the relation:

$$d(\alpha) = \sqrt{2r} \cdot \sqrt{1 - cos(\alpha)} \approx r\alpha + r\alpha^3/24 + O(\alpha^5) \quad (2)$$

The equation holds for $\alpha \in (0, \pi/2)$, while for an accurate estimation of the radii a linear regression is applied on the extremas of the (α, r) pairs. Also an intuitive geometric interpretation of the obtained radii makes it usable for surface categorization, i.e. the large radii denotes planar surfaces, while small radii is for cylindric objects. For the recognition test we used the radii as histograms computed for individual objects.

IV. OBJECT RECOGNITION BENCHMARKS

This section covers the object recognition benchmarking description as well as the results of the different classification benchmarks. During the training-test phase different combinations of recorded data were tested, i.e. training with data recorded daytime and testing against data captured at night or night with snow. The tuning parameters for the feature descriptors were chosen in such a way to be the most discriminative to the size of the objects considered as test objects in Figure 2. In order to apply the same type of feature descriptors for the color depth data from the stereo camera, this was transformed into depth-intensity data.

The output of the object classifier described in section II were benchmarked using a quantitative comparison the AC_d metric was used defined as [17]:

$$AC_d = 1 - \sqrt{W \cdot (1 - TP)^2 - (1 - W) \cdot FP^2} \quad (3)$$

where TP and FP denote the true positives respectively the false positives and W is a weighting factor giving an

Descr.	Svd.	Srns.	Srn	Srd
VFH	0.55	0.63	0.78	0.44
CVFH	0.61	0.61	0.69	0.51
PFH	0.62	0.56	0.71	0.37
FPFH	0.41	0.46	0.50	0.39
RSD	0.73	0.52	0.77	0.42
RIFT	0.59	0.54	0.69	0.33
SI	0.78	0.64	0.83	0.59
ISE	0.77	0.59	0.83	0.48
IGE	0.64	0.57	0.80	0.42
ESF	0.63	0.65	0.79	0.49

TABLE I: SVM classification summary for different feature-descriptor using stereo vision (Sv) data and Swiss ranger (Sr) camera at daytime (d), night (n) and night-snow (ns)

application specific weighting to TP in favor of FP. The over-season data matching results are shown in Table I: the data recorded with stereo camera (Sc) and Swiss ranger (Sr) at cloudy daytime as training data and compared against sunny daytime (d) recordings, night (n) time data and night time with snow (ns) fall. The night time images were only partially visible with the stereo camera due to the lighting limitations, thus the output for these test cases are with limited performance.

The best results in average was achieved with the SI type of feature descriptor and the most discriminative training-test data set pair was obtained with the ToF camera on night time data. The output for the SI type of feature descriptor is summarized in a confusion matrix form using the same daytime both at the training and testing phase for the SR camera.

TABLE II: Rounded confusion matrix for the test objects shown in Figure 2 using SVM and SI feature descriptors

	box	cyl.	hyd.	sti.	tab.	tru.	car.	t-b.
box	0.3	0.0	0.0	0.1	0.1	0.2	0.1	0.2
cyl.	0.0	0.5	0.0	0.0	0.0	0.1	0.4	0.0
hyd.	0.0	0.1	0.6	0.0	0.0	0.2	0.0	0.0
sti.	0.0	0.0	0.1	0.8	0.0	0.1	0.0	0.0
tab.	0.0	0.1	0.0	0.0	0.5	0.3	0.0	0.2
tru.	0.0	0.0	0.0	0.3	0.1	0.6	0.0	0.0
car.	0.0	0.0	0.1	0.0	0.1	0.3	0.5	0.0
t-b.	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.6

The next evaluation of the SVM classifier shows the performance in terms of the ROC curve of the two cameras for the best case scenario: the same period of day is considered both for the training and test phases, and the daytime is chosen for the stereo camera (results in Figure 6) and night-daytime time for the SR sensor (results in Figure 4, 5). The points on the curve represent the classifier output for the objects. While the range of the considered objects varies in a large range, the output for the different object classes also suffers from these variations.

As it can be seen on these figures the data from SV and ToF cameras have different discriminative characteristics, and also the best results according to these figure is achieved with the ToF camera during night-time.

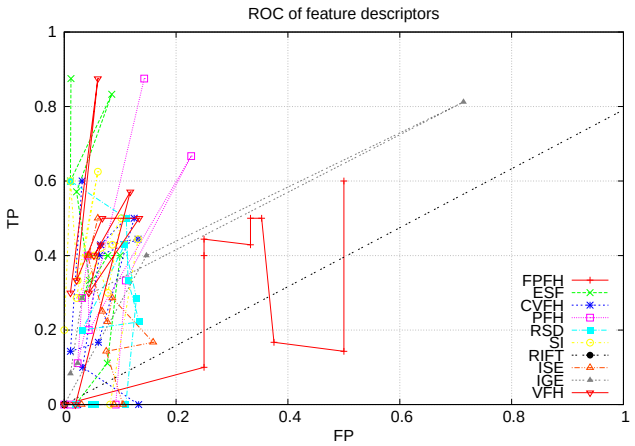


Fig. 4: Feature descriptor test results for the night-night training-testing dataset with ToF camera (best viewed in color)

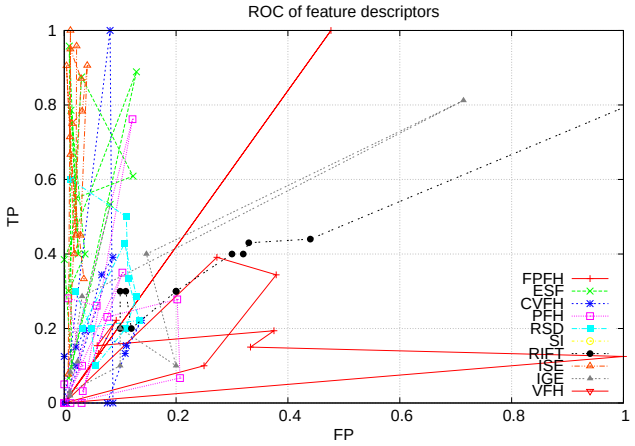


Fig. 5: Feature descriptor test results for the daytime-daytime training-testing dataset with ToF camera (best viewed in color)

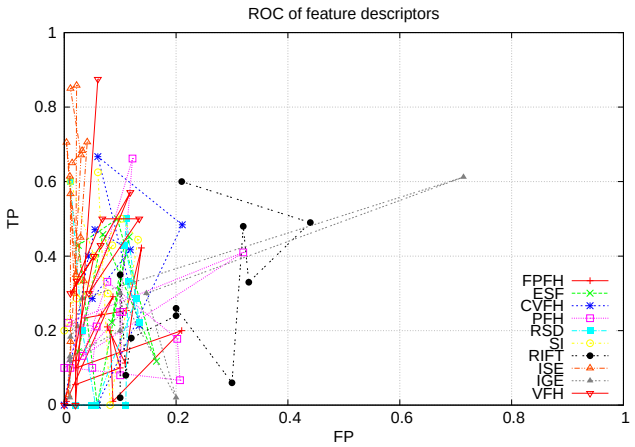


Fig. 6: Feature descriptor test results for the daytime-daytime training-testing dataset with stereo camera (best viewed in color)

V. CONCLUSIONS

In this paper the robustness of various 3D feature descriptors was analyzed for outdoor data under different weather conditions. The main scope of the analyses were to give a quantitative and qualitative evaluation based on the test dataset of the different feature descriptors for data acquired with stereo camera and time-of-flight sensor under different weather conditions including daytime, snow, nighttime and cloudy weather. The results were summarized for the different approaches using the ROC curves as well as the confusion matrix for the test objects.

In future the fusion of the output from different 2D-3D classifiers is proposed to be considered in an unsupervised learning framework in order to enhance the classification rate for the 3D outdoor classification under various weather conditions.

REFERENCES

[1] M. Cummins and P. Newman, "Fab-map: Probabilistic localization and mapping in the space of appearance," *Int. J. Rob. Res.*, vol. 27, no. 6, pp. 647–665, June 2008. [Online]. Available: http://dx.doi.org/10.1177/0278364908090961

[2] L. Alexandre, "3D descriptors for object and category recognition: a comparative evaluation," *Workshop on Color-Depth Camera Fusion in Robotics*, 2012.

[3] L. Tamas and A. Majdik, "Heterogeneous feature based correspondence estimation," in *IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems*, IEEE. Munich, Germany: IEEE, September 2012, pp. 89–94.

[4] A. Anand, H. S. Koppula, T. Joachims, and A. Saxena, "Contextually Guided Semantic Labeling and Search for 3D Point Clouds," *International Journal of Robotics Research*, Nov. 2013.

[5] H. Ali, F. Shafait, E. Giannakidou, A. Vakali, N. Figueroa, T. Varvadoukas, and N. Mavridis, "Contextual object category recognition for RGB-D scene labeling," *Robotics and Autonomous Systems*, vol. 62, no. 2, pp. 241–256, Feb. 2014.

[6] G. Dudek and D. Jugessur, "Robust place recognition using local appearance based methods," in *Robotics and Automation, 2000. Proceedings. ICRA '00. IEEE International Conference on*, vol. 2, 2000, pp. 1030–1035 vol.2.

[7] N. Sünderhauf, P. Neubert, and P. Protzel, "Are we there yet? challenging seqslam on a 3000 km journey across all four seasons," in *Workshop on Long-Term Autonomy, IEEE International Conference on Robotics and Automation (ICRA)*, 2013.

[8] M. Milford, "Vision-based place recognition: how low can you go?" *The International Journal of Robotics Research*, vol. 32, no. 7, pp. 766–789, 2013. [Online]. Available: http://ijr.sagepub.com/content/32/7/766.abstract

[9] E. Johns and G.-Z. Yang, "Feature co-occurrence maps: Appearance-based localisation throughout the day," in *ICRA*. IEEE, 2013, pp. 3212–3218.

[10] 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)*, Taipei, Taiwan, 2010.

[11] L. Bo, X. Ren, and D. Fox, "Unsupervised feature learning for rgb-d based object recognition," *ISER, June*, pp. 1–15, 2012. [Online]. Available: http://homes.cs.washington.edu/ lfb/paper/iser12.pdf

[12] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, November 2004.

[13] R. Rusu and S. Cousins, "3d is here: Point cloud library (pcl)," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, May 2011, pp. 1 –4.

[14] P. Arbelaez, B. Hariharan, S. Gupta, L. Bourdev, and J. Malik, "Semantic segmentation using regions and parts," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, June 2012, pp. 3378–3385.

[15] W. Choi, Y.-W. Chao, C. Pantofaru, and S. Savarese, "Understanding Indoor Scenes Using 3D Geometric Phrases," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, June 2013, pp. 33–40.

[16] R. Paul, R. Triebel, D. Rus, and P. Newman, "Semantic categorization of outdoor scenes with uncertainty estimates using multi-class gaussian process classification," *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2404–2410, Oct. 2012.

[17] T. Fawcett, "An introduction to roc analysis," *Pattern Recogn. Lett.*, vol. 27, no. 8, pp. 861–874, June 2006. [Online]. Available: http://dx.doi.org/10.1016/j.patrec.2005.10.010

[18] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, ser. COLT '92. New York, NY, USA: ACM, 1992, pp. 144–152. [Online]. Available: http://doi.acm.org/10.1145/130385.130401

[19] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, May 2011. [Online]. Available: http://doi.acm.org/10.1145/1961189.1961199

[20] A. Aldoma, Z.-C. Marton, F. Tombari, W. Wohlkinger, C. Potthast, B. Zeisl, R. B. Rusu, S. Gedikli, and M. Vincze, "Tutorial: Point cloud library: Three-dimensional object recognition and 6 dof pose estimation." *IEEE Robot. Automat. Mag.*, vol. 19, no. 3, pp. 80–91, 2012.

[21] S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using local affine regions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1265–1278, Aug. 2005.

[22] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[23] R. Shams, R. A. Kennedy, P. Sadeghi, and R. I. Hartley, "Gradient intensity-based registration of multi-modal images of the brain." in *IEEE 11th International Conference on Computer Vision, ICCV 2007, Rio de Janeiro, Brazil, October 14-20, 2007*. IEEE, 2007, pp. 1–8.

[24] A. E. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 433–449, May 1999.

[25] A. Johnson and M. Hebert, "Surface matching for object recognition in complex 3-D scenes," *Image and Vision Computing*, vol. 16, pp. 635–651, 1998.

[26] R. Rusu, N. Blodow, Z. Marton, and M. Beetz, "Aligning point cloud views using persistent feature histograms," in *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, Sept 2008, pp. 3384–3391.

[27] E. Wahl, U. Hillenbrand, and G. Hirzinger, "Surflet-pair-relation histograms: a statistical 3d-shape representation for rapid classification," in *3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings. Fourth International Conference on*, Oct 2003, pp. 474–481.

[28] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (fpfh) for 3d registration," in *Proceedings of the 2009 IEEE International Conference on Robotics and Automation*, ser. ICRA'09. Piscataway, NJ, USA: IEEE Press, 2009, pp. 1848–1853.

[29] R. B. Rusu, G. Bradski, R. Thibaux, and J. Hsu, "Fast 3D Recognition and Pose Using the Viewpoint Feature Histogram," in *Proceedings of the 23rd IEEE International Conference on Intelligent Robots and Systems (IROS)*, Taipei, Taiwan, October 2010.

[30] A. Aldoma, M. Vincze, N. Blodow, D. Gossow, S. Gedikli, R. Rusu, and G. Bradski, "Cad-model recognition and 6dof pose estimation using 3d cues," in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, Nov 2011, pp. 585–592.

[31] W. Wohlkinger and M. Vincze, "Ensemble of shape functions for 3d object classification," in *Robotics and Biomimetics (ROBIO), 2011 IEEE International Conference on*, Dec 2011, pp. 2987–2992.

[32] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, "Matching 3d models with shape distributions," in *Proceedings of the International Conference on Shape Modeling & Applications*, ser. SMI '01. Washington, DC, USA: IEEE Computer Society, 2001, pp. 154–160.

[33] C. Y. Ip, D. Lapadat, L. Sieger, and W. C. Regli, "Using shape distributions to compare solid models," in *Proceedings of the Seventh ACM Symposium on Solid Modeling and Applications*, ser. SMA '02. New York, NY, USA: ACM, 2002, pp. 273–280.

[34] Z.-C. Marton, D. Pangercic, N. Blodow, and M. Beetz, "Combined 2D-3D categorization and classification for multimodal perception systems," *The International Journal of Robotics Research*, vol. 30, no. 11, pp. 1378–1402, Aug. 2011.