2014 IEEE International Conference on Robotics & Automation (ICRA)
Hong Kong Convention and Exhibition Center
May 31 - June 7, 2014. Hong Kong, China

RGB-D Object Classification Using Covariance Descriptors

Duc Fehr, William J. Beksi, Dimitris Zermas and Nikolaos Papanikolopoulos
{fehr, beksi, dzermas, npapas}@cs.umn.edu
Department of Computer Science and Engineering
University of Minnesota
Minneapolis, MN 55455

Abstract— **In this paper, we introduce a new covariance based feature descriptor to be used on "colored" point clouds gathered by a mobile robot equipped with an RGB-D camera. Although many recent descriptors provide adequate results, there is not yet a clear consensus on how to best tackle "colored" point clouds. We present the notion of a covariance on RGB-D data. Covariances have not only been proven to be successful in image processing, but in other domains as well. Their main advantage is that they provide a compact and flexible description of point clouds. Our work is a first step towards demonstrating the usability of the concept of covariances in conjunction with RGB-D data. Experiments performed on an RGB-D database and compared to previous results show the increased performance of our method.**

I. Introduction

One of the most important properties of a mobile robot is the ability to process information from the surrounding environment. In order to perform its tasks, a robot needs to be able to perceive its environment. For instance, in navigation a robot needs to be able to recognize certain features that can be used as landmarks. If a robot needs to grasp an object, it first needs to find the object in the scene. For this task, 3D object detection and 3D object classification have to be performed. The recent introduction of affordable RGB-D cameras such as the Microsoft Kinect [1], has induced a great interest in the robotics and computer vision community towards developing efficient algorithms for point cloud processing. Previously, in order to capture a point cloud, expensive specialized sensors such as lasers or dedicated range imagers were needed; now range data is readily available from cheap sensors which provide point clouds that can easily be extracted from a depth map. In addition, these RGB-D cameras provide additional information in a "colored" point cloud. In image processing, descriptors such as SIFT [2] and SURF [3] have been shown to be very successful. Work has been done in the last few years on descriptors for RGB-D data that have proven to be successful. However, it is not yet clear how to robustly process RGB-D data from a mobile robot.

In this work, we introduce a covariance based 3D point descriptor which is compact (low dimensionality) and fast (processing speed) to match against one another. Its greatest benefit is the flexibility that it offers in including other discriminative features from the objects of interest for improved object retrieval. Also, the description is very compact; a single positive definite matrix describes an entire "colored" point cloud. In Section III-A we describe in detail which features were chosen as well as discuss the number of parameters used to compute the point cloud features. This paper attempts to open the door to exploring covariances on RGB-D data, as they have been successful in a range of other areas.

In the next section we point to those successful areas and discuss some of the previous descriptors for RGB-D data. Then we present the covariance descriptor and show the experimental results. Finally, we conclude and mention potential future work.

II. Related Work

One of the earliest descriptors developed for point clouds are the spin images introduced by Johnson *et al.* [4]. Spin images have been evolved and refined by many other works. For instance, Carmichael *et al.* [5] address spin images in the context of different depth and thus different sampling densities. Moving away from the idea of spin images, Frome *et al.* [6] export Belongie *et al.*'s work from image processing into point cloud processing. A survey on different point cloud descriptors has been published by Tangelder and Veltcamp [7]. In a more recent development, Rusu *et al.* [8] use a histogram based approach to find matching points in order to facilitate alignment and merging of point clouds. Tombari *et al.* [9] show that building signatures on top of histograms provides good results for feature detection.

The common denominator of these methods is that they have been developed for "simple" point clouds. Now that access to "colored" point clouds has been drastically simplified with the introduction of low cost RGB-D cameras, a lot of work has begun in the area of RGB-D data. Lai *et al.* [10] introduced an RGB-D image database of objects that have been pre-segmented with which different algorithms can be tested. The present paper uses this database for the experiments. Alongside the introduction of the benchmark, the authors introduce a descriptor that can classify and detect objects from RGB-D data. The same authors introduce a second approach to classification in [11] using sparsity techniques in order to classify objects. Their work also shows that shape, in conjunction with appearance information, provides better classification results than classifying for shape or appearance separately. Bo *et al.* [12] introduces kernel descriptors for object recognition. In [13], Bo *et al.* compute and evaluate descriptors at different patch sizes and thus

978-1-4799-3684-7/14/$31.00 ©2014 IEEE

build a hierarchical model of depth kernels. With a similar focus, Blum *et al.* [14] use dictionary learning in order to find relevant features for classification. In their work, Bo *et al.* [15] merge the previous ideas of hierarchical models with dictionary learning. Similar work has been performed by Tang *et al.* [16] on a dataset provided by Willow Garage for the 2011 ICRA Solutions in Perception Instance Recognition Challenge.

Point clouds in the Lai *et al.* database [10] are cleanly segmented. The assumption of pre-segmented objects facilitates the classification task, but moves the problem towards answering the question of how to perform the segmentation. Work by Mishra *et al.* [17] that segments objects from "colored" point clouds using oriented masks, tries to answer this question. Covariances have been introduced as descriptors in image processing by Tuzel *et al.* [18] and have been shown to produce excellent results for detection and classification. Porikli *et al.* [19] continue to build on this idea to track objects in videos.

III. Covariance Descriptor

A. Covariance

The various different application domains of covariance descriptors led us to test this paradigm for classification. A first try of the approach has been given by the authors in [20]. These point-wise comparisons were very time consuming due to the enormous number of descriptors that characterized each object and do not qualify well for the task of classification. They are rather suited for a different task such as registration. In this paper, a different approach is pursued. Instead of computing point-wise covariance descriptors, a covariance descriptor over an entire object is computed, resulting in a single positive definite matrix characterizing each object. Comparisons between different objects are thus drastically sped up and the number of saved descriptors is enormously reduced.

The descriptors are defined as follows:

Definition 1: Let $F_i \in \mathbb{R}^p$, for $i = 1, 2, \ldots, N$, be the feature vectors of the N points of an object, then the covariance descriptor of this object $K \in \mathcal{S}^p_{++}$ is defined as:

$$K = \frac{1}{N-1} \sum_{i=1}^{N} (F_i - \mu_F)(F_i - \mu_F)^T \tag{1}$$

where μ_F is the mean feature vector and $\mathcal{S}^p_{++}$ is the space of $p \times p$ Symmetric Positive Definite (SPD) matrices.

This allows for a compact description in each frame. Since the covariance is a positive symmetric matrix, we are effectively reducing the number of values that we need to save to the number of upper or lower triangular entries in the matrix. It is also quite easy to add additional features at each point as the matrix will only grow by a row and column. Compared to a histogram based approach where adding a feature adds an additional dimension to the histogram cube, the appeal of the covariance descriptor method is clearly apparent. The used features will be described in Section III-C.

B. Distance

One of the problems with using covariances as descriptors is that their space is not Euclidean, but they instead span a Riemannian manifold. Arsigny *et al.* [21] introduced the following metric that is used throughout this work:

$$d_{LE}(X, Y) = \|\log(X) - \log(Y)\|_F \tag{2}$$

where X and Y are two positive definite matrices, $\log(\cdot)$ designates the matrix logarithm, and $\|\cdot\|_F$ is the Froebinius norm.

The matrix logarithms can be precomputed and determining this distance becomes a vector operation. The additional advantage of using this distance is that we can use it as is, as the distance for the radial basis function (RBF) kernel for the SVM classification.

C. Features

The RGB-D dataset provides at each point the x, y, z Cartesian coordinate as well as the r, g, b color channel values. This allows for six different features.

This number is augmented by computing the normal (n_x, n_y, n_z) at each point to get three additional features. The normals are computed using the method introduced by Hoppe *et al.* [22]. At each point x, a neighborhood is taken and the directions along which the data is least scattered is determined. This direction is the normal estimate.

Once the normals are computed at each point, the curvature can be estimated which provides us with two values along the main curvature axes (c_1, c_2). These directions are estimated by projecting the normals in the neighborhood of the point onto the tangential plane at that point. The parameters of the ellipse fitting these points are then computed which provides us the main curvature axes values. Finally, the product of these curvatures provides a "total" curvature $C = c_1 \cdot c_2$ at the point. The use of the local curvature at a point has been inspired by work on non-rigid shapes by the Bronstein brothers [23], where they show that the curvatures provide a good feature for classifying non-rigid objects.

Since image derivatives have produced excellent results in image processing, they are computed here as well in addition to the different derivatives on the depth image: I is the intensity image computed from RGB. I_x and I_y correspond to the output of the Sobel operator [24], when applied to I along x and y respectively. This in effect produces the gradient along x and y at each pixel. For I_{xx}, I_{yy}, I_{xy}, the operator is applied a second time on the patch. $M = \sqrt{I_x^2 + I_y^2}$ corresponds to the magnitude of the gradient of the image patch. The same operations are performed on the depth image D to produce D_x, D_y and $D_M = \sqrt{D_x^2 + D_y^2}$.

Thus, a total of 22 features can be used in the computation of our covariance descriptor. The experiments are run with 15 different combinations of these features, a sample of which is as follows:

$$
\begin{aligned}
F_{PCN} &= [x, y, z, r, g, b, n_x, n_y, n_z] \\
F_{PCNK} &= [x, y, z, r, g, b, n_x, n_y, n_z, C] \\
F_{PCIM} &= [x, y, z, r, g, b, I_x, I_y, I_{xx}, I_{yy}, I_{xy}, M] \\
F_{PCIMIDNKk} &= [x, y, z, r, g, b, I, I_x, I_y, I_{xx}, I_{yy}, I_{xy}, \\
&\quad M, D_x, D_y, D_M, n_x, n_y, n_z, C, c_1, c_2] \,.
\end{aligned}
$$

Each object is represented by its covariance matrix. Since this matrix is symmetric, each object is characterized by the number of upper-triangular (or lower-triangular) values. The total number of parameters that describe a frame are then $\frac{(p+1)p}{2}$ (p is the number of features).

D. Scale Invariance

The descriptor is scale invariant in that the covariance captures the relationships between the different features. At different distances, the distribution of the points remains constant as well as the covariance. In order to show this scale invariance we performed the experiment described in Section IV-C.4.

IV. Evaluation

A. Setup

For our experiments, the database of RGB-D data introduced by Lai *et al.* [10] is used. This database comprises video-frames of approximately 300 objects divided into 51 categories. There are about 250,000 total frames in the database. Following the experimental procedure in [10] and [14], we sub-sample this database by taking every fifth frame in order to have around 45,000 frames on which we run classification. As described in the previous section, we compute the normal and curvature at each point and add these to our features. During the experimental runs, we follow [10]'s Section V. For category recognition, we randomly leave one object out from each category for testing and train the classifiers on all the views of the remaining objects. For instance recognition, we consider two scenarios:

- Alternating contiguous frames: Divide each video into 3 contiguous sequences of equal length. There are 3 heights (videos) for each object which gives 9 video sequences for each instance. We randomly select 7 of these for training and test on the remaining 2.
- Leave-sequence-out: Train on the video sequences of each object where the camera is mounted $30°$ and $60°$ above the horizon and evaluate on the $45°$ video sequence.

An additional test on category recognition is performed to test the correlation between the classification accuracy and the number of categories. In this experiment, for every chosen number $n \in [2, \ldots, 45]$, n categories are chosen randomly out of 51 and the same category recognition procedure is applied in which for every category one object is randomly removed from the entire set for testing purposes and the learning is done on the remaining objects.

Where frames are randomly chosen, the experiments are run 10 times and the average is reported.

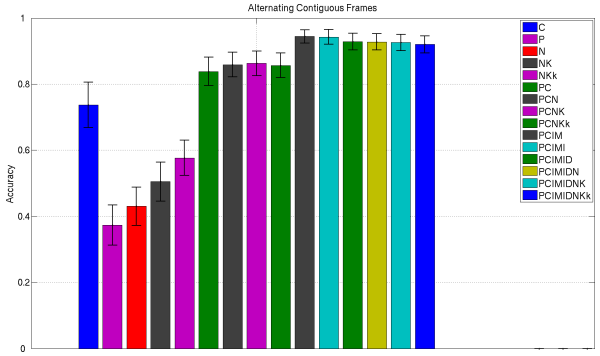


Fig. 1: This figure shows the classification accuracy results for the Alternating Contiguous Sequence experiment. The x-axis shows the different features used. The standard deviation over 10 runs is displayed in red.

B. Experiments

The experiments are run with the ten different combinations of feature vectors described in Section III-C. For each set of features, the covariance for each frame is computed as well as its matrix logarithm. Classification is performed using an RBF kernel SVM classifier [25] in which the log-Euclidean distance described in Section III-B is used.

C. Results

1) Instance Classification: Fig. 1 shows the results for the Alternating Contiguous Sequences run. The standard deviation is also given for 10 tries. It is interesting to note that for instance recognition, the classifier on color C (color channels rgb) performs better than the classifiers using only shape P, N, Kk (point coordinates xyz, normals N, and curvatures Kk respectively). The feature vectors containing more image cues perform best in these scenarios. It is reasonable to obtain such results as the differences in an instance of "similar" objects lie more in its appearance than in its shape. Using shape together with color information produces improved results as already shown in [10] and [11]. When classifying over color features as well as shape features, the results are similar for the combination of features. This indicates that the curvature information in addition to the normal information is not that useful. A similar case can be made against the information the depth image provides. More investigation is needed in order to provide a definite explanation of this phenomenon.

2) Category Classification: Fig. 2 shows the results over the category test. These results are interesting in that they show improved classification with the shape features compared to color features. This is somewhat expected as the color pattern over different categories might be similar whereas the shape pattern differs more. The best result is achieved when using all the shape features available to us.

The best results over the different feature combinations are reported in Table I for comparison purposes with the other methods.

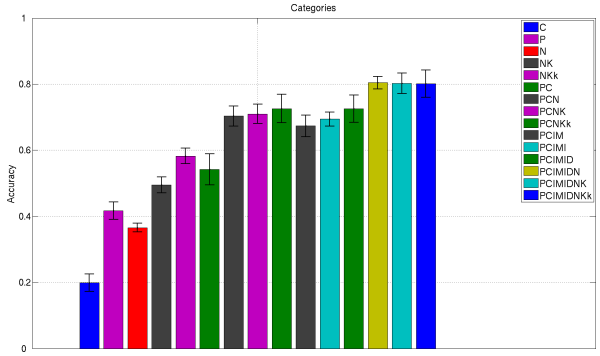


Fig. 2: This figure shows the classification accuracy results for the categories experiment. The x-axis shows the different features used. The standard deviation over 10 runs is displayed in red.

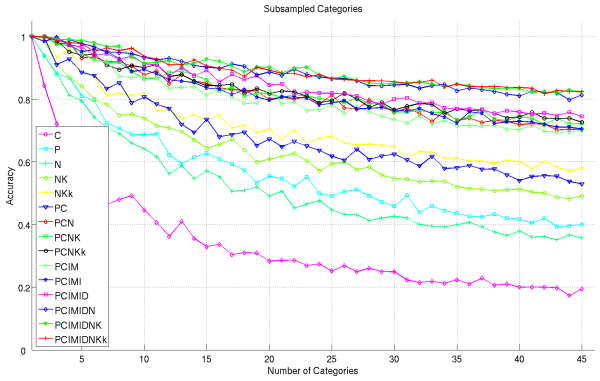


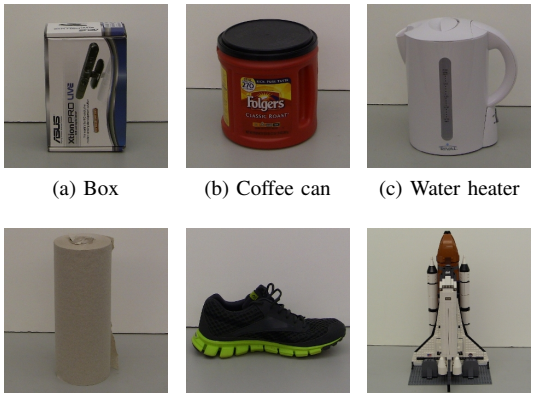
Fig. 3: This figure shows the classification accuracy results for the sub-sampled categories experimental run. The x-axis shows the number of sub-sampled categories for each run.

TABLE I: Classification Accuracy. (a) Leave-Sequence-Out (b) Alternating Contiguous Frames. Accuracies Are Averaged Over 10 Trials. Dimensionality Comparison

Method	Instance		Category	Dim
	(a)	(b)		
Linear SVM [10]	73.9	90.2 ± 0.6	81.9 ± 2.8	4203
Nonlinear SVM [10]	74.8	90.6 ± 0.6	83.8 ± 3.5	4203
Random Forest [10]	73.1	90.5 ± 0.4	79.6 ± 4.0	4203
IDL [11]	-	91.3 ± 0.3	85.4 ± 3.2	4203
HKDES [13]	82.4	-	84.1 ± 2.2	7000
Kernel Desc. [12]	84.5	-	86.2 ± 2.1	39000
CKM Desc. [14]	90.4	92.1 ± 0.4	86.4 ± 2.3	19200
upgraded HMP [15]	92.1	-	87.5 ± 2.9	188300
Cov Desc. (this work)	90.7	94.4 ± 2.0	80.4 ± 1.9	253

3) Sub-sampled Category Classification: In an effort to understand the relationship between the number of categories and the classification accuracy, the sub-sampled category classification was run. Fig. 3 provides the results for this experiment. The x-axis shows the number of categories used in the runs. For clarity purposes, only the results from the best performing descriptors are displayed. The decaying trend is somewhat expected. For each run, the categories have been redrawn. There exists 10 runs for each number of subsamples.

4) Scale Invariance: For the six objects given in Fig. 4, we performed an experiment where a point cloud of each object was captured at four increasing distances from the RGB-D camera corresponding to 2-8 *ft* (0.61-2.43 *m*). Fig. 5 shows one such set of images. At 8 feet, the object is near the far-end range of the camera's capabilities. For each combination of features, we computed the norm of the geodesic distance between each object and itself, and object to object at the various positions. The object's feature vector consisted of the following parameters: $F_{PCN} = [x, y, z, r, g, b, n_x, n_y, n_z]$. The average norm of the geodesic distance between each object and itself, at different scales, was found to be 0.104. The average norm between each object and all other objects was 0.204. This corresponds to a difference of 48.8% between the different objects, and shows a clear distinction in the geodesic distance regardless of the scale.

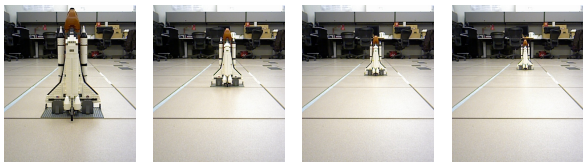


(a) Box (b) Coffee can (c) Water heater

(d) Paper roll (e) Shoe (f) Shuttle

Fig. 4: Objects used for the scale invariance testing.

A second experiment to test the classification performance of the scaled data set was run using the Microvision robot [26]. A circling motion control scheme [27] was used to capture a sequence of view points of each object. Next, the object point clouds were filtered and used to train an SVM. Finally, the data of the scaled object was passed to the SVM classifier to predict the label. In all cases, the prediction performed flawlessly, i.e. a misclassification rate of zero. It is worthy to note that at the smallest scale, the point cloud composition of the object is less than 1% of the total point cloud size, yet the classification was consistently performed correctly.



(a) 2 *ft* (b) 4 *ft* (c) 6 *ft* (d) 8 *ft*

Fig. 5: Shuttle at four different scales.

V. Discussion

In [10] and [11] the authors come to the conclusion that using shape and vision information separately yields low classification scores, whereas using the same information together provides better accuracies. Our experimental results confirm this conclusion.

A. Instance Classification

The results of the experiments for instance classification are interesting in that they give insight into which features are important. For example, one can see in Fig. 1 that the descriptors using solely the color (C) information, although not giving very good results, outperform the shape feature based descriptors (P, N, NK, NKk). This is expected as the objects in each different category usually have very similar shapes. For instance, there are various colored apples in the database and consequently they share the same (similar) shape but only differ by color. The different objects have a similar overall shape inside each category which makes it difficult to classify on these features. On the other hand, color can manage the classification better. It explains why the descriptors containing the most visual information perform best. This trend is visible in both scenarios, leave-sequence-out and alternating contiguous frames.

When comparing the results with other works (Table I), our method produces comparable results to other state of the art methods. However, it is important to note that for our method only the covariance over the different objects is computed. We did not compute and choose specific key feature points needed for comparison purposes. Finally, the classification is operated in a $\frac{(p+1)p}{2}$ dimensional space, where p is the number of features used and as such is relatively small. The amount of processing necessary to compute covariances is minimal. Our method not only performs as well as others, but can do so using less parameters.

B. Category Classification

Contrary to the previous runs, in the category classification test (Fig. 2), the shape features (P, N, NK, NKk) are more meaningful than the color descriptor (C). Intuitively, this again makes sense since different objects should have different shapes, but could have the same color. For instance, in the database there is a red apple and red toothpaste tube that may be confused when classifying only on color. Combining the classifiers provides better results. When we compare our classifier to other classifiers (Table I), the covariance descriptor produces results that are reasonable. From these results, it shows that the classification space cannot be too small, allowing the different classes to remain separable. An indication of this fact is that the best classification results come from the higher dimensional feature descriptors. Repeating the previous section, a strength but at the same time a potential weakness of our descriptors, is that they have a very small number of parameters that are used for classification.

C. Sub-sampled Category Classification

To understand where the category classification breaks down, another experiment was run in which an increasing number of categories was used. For each run, the used categories have been redrawn. The results are shown in Fig. 3.

It is noticeable that the addition of the curvature does not significantly increase the accuracy of the classification. This raises the point that some features might be more suitable than others for classification. In the scope of this experiment, $(PCIMIDN)$, $(PCIMIDNK)$, and $(PCIMIDNKk)$ all perform similarly well. This means that there is information in the normal distribution of the objects, but the curvatures do not add more separability. It is interesting to note that the descriptors with the highest dimension perform best. These additional dimensions may be necessary to make the various classes separable enough in the classification space.

Another possible approach is to completely switch to a different classification paradigm. Instead of using a "linear" classifier such as SVMs, a dictionary learning approach might provide better results. Initial work by Blum *et al.* [14] and Bo *et al.* [15] suggests that this approach is viable for processing of 3D point cloud data. There has been significant research done that uses classifiers based on dictionary learning, such as Ramirez *et al.* [28] or Wright *et al.* [29], in which reconstruction error is used for classification purposes.

D. Scale Invariance

The experiment run provides us with strong empirical evidence that the covariance descriptor is scale invariant. We are currently working on a formal proof.

VI. Conclusion and Future Work

This paper has presented a novel point cloud descriptor based on covariances. The main goal is to introduce the notion of covariances in conjunction with point clouds in general and RGB-D data ("colored" point clouds) in particular. There are two main advantages to using covariances:

1) **Compactness**. Only a small number of parameters is necessary. These parameters amount to the number of upper or lower triangular entries in the covariance matrix computed from the features. The advantages are two-fold. First, the descriptors can be computed very efficiently. Second, the storage space necessary for these descriptors is dramatically reduced. An entire object cloud is now saved with 253 parameters (for 22 features). In the same space, only 43 RGB-D points need to be saved.
2) **Flexibility**. It is easy to add new features for use in the covariance. Adding a feature amounts to adding a new row and a column to the descriptor. This is simpler than a histogram based approach in which adding a feature amounts to adding a new dimension to the histogram cube.

The experiments have shown the usability of such an approach in instance classification as well as category classification. They show that a simple covariance approach produces comparable results to other state of the art methods.

These same experiments have also uncovered some potential weaknesses that provide a direction for future work. The results have shown that the curvature used as a feature may not provide enough discriminative power in order to help in the classification. Other features that focus more on the visual appearance, have shown promise in instance classification. Combining these image cues with shape cues allows for a larger space in which classification performance of the SVM classifier could be increased. Future work includes using principle component analysis (PCA) to determine the compact features to use in the covariance descriptor. We are also interested in testing other classifiers such as random forests and implementing classification based on a dictionary learning approach.

There is much work left in the area of RGB-D data processing methods such as the one presented in this paper. The works of Lai *et al.* [10], Bo *et al.* [13] [15] and Blum *et al.* [14] only scratch the surface. Due to there being no de facto method for processing RGB-D data, interesting research opportunities persist in the area of processing point cloud data. This work has presented a new direction that has produced promising results and provides a new perspective in the exploration of RGB-D data processing.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation through grants #IIP-0934327, #IIP-1032018, #SMA-1028076, #CNS-1039741, #IIS-1017344, #CNS-1061489, #IIP-1127938, #IIP-1332133, and #CNS-1338042.

REFERENCES

[1] Kinect. [Online]. Available: http://www.xbox.com/en-us/kinect

[2] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60(2), pp. 91–110, 2004.

[3] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Computer Vision ECCV 2006*, ser. Lecture Notes in Computer Science, A. Leonardis, H. Bischof, and A. Pinz, Eds. Springer Berlin / Heidelberg, 2006, vol. 3951, pp. 404–417.

[4] A. Johnson, "Spin-images: A representation for 3D surface matching," Ph.D. dissertation, Carnegie Mellon University, 1997.

[5] O. Carmichael, D. Huber, and M. Hebert, "Large data sets and confusing scenes in 3D surface matching and recognition," in *Proceedings of the Second International Conference on 3D Digital Imaging and Modeling*, Oct. 1999, pp. 358–367.

[6] A. Frome, D. Huber, R. Kolluri, T. Blow, and J. Malik, "Recognizing objects in range data using regional point descriptors," in *Computer Vision - ECCV 2004*, vol. 3023, 2004, pp. 224–237.

[7] J. W. Tangelder and R. C. Veltkamp, "A survey of content based 3D shape retrieval methods," *International Conference on Shape Modeling and Applications*, pp. 145–156, 2004.

[8] 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 (ICRA)*, 2009, pp. 1848–1853.

[9] F. Tombari, S. Salti, and L. Di Stefano, "Unique signatures of histograms for local surface description," in *Proceedings of the 11th European Conference on Computer Vision: Part III*. Springer-Verlag, 2010, pp. 356–369.

[10] K. Lai, L. Bo, X. Ren, and D. Fox, "A large-scale hierarchical multi-view rgb-d object dataset," in *Proceedings of the 2011 IEEE International Conference on Robotics and Automation (ICRA)*, 2011, pp. 1817–1824.

[11] ——, "Sparse distance learning for object recognition combining rgb and depth information," in *Proceedings of the 2011 IEEE International Conference on Robotics and Automation (ICRA)*, 2011, pp. 4007 – 4013.

[12] L. Bo, X. Ren, and D. Fox, "Depth kernel descriptors for object recognition," in *Proceedings of the 2011 IEEE Conference on Intelligent Robots and Systems (IROS)*, 2011, pp. 821–826.

[13] L. Bo, K. Lai, X. Ren, and D. Fox, "Object recognition with hierarchical kernel descriptors," in *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 1729–1736.

[14] M. Blum, J. Springenberg, J. Wulfing, and M. Riedmiller, "A learned feature descriptor for object recognition in rgb-d data," in *Proceedings of the 2012 IEEE International Conference on Robotics and Automation (ICRA)*, 2012, pp. 1298–1303.

[15] L. Bo, X. Ren, and D. Fox, "Unsupervised feature learning for rgb-d based object recognition," *ISER*, 2012.

[16] J. Tang, S. Miller, A. Singh, and P. Abbeel, "A textured object recognition pipeline for color and depth image data," in *Proceedings of the 2012 IEEE International Conference on Robotics and Automation (ICRA)*, 2012, pp. 3467–3474.

[17] A. Mishra, A. Shrivastava, and Y. Aloimonos, "Segmenting simple objects using rgb-d," in *Proceedings of the 2012 IEEE International Conference on Robotics and Automation (ICRA)*, 2012, pp. 4406–4413.

[18] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: A fast descriptor for detection and classification," in *Computer Vision ECCV 2006*, ser. Lecture Notes in Computer Science, A. Leonardis, H. Bischof, and A. Pinz, Eds. Springer Berlin / Heidelberg, 2006, vol. 3952, pp. 589–600.

[19] F. Porikli, O. Tuzel, and P. Meer, "Covariance tracking using model update based on lie algebra," in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006, pp. 728–735.

[20] D. Fehr, A. Cherian, R. Sivalingam, S. Nickolay, V. Morellas, and N. Papanikolopoulos, "Compact covariance descriptors in 3d point clouds for object recognition," in *Proceedings of the 2012 IEEE International Conference on Robotics and Automation (ICRA)*, 2012, pp. 1793–1798.

[21] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Log-euclidean metrics for fast and simple calculus on diffusion tensors," *Magnetic Resonance in Medicine*, vol. 56, no. 2, pp. 411–421, 2006.

[22] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stuetzle, "Surface reconstruction from unorganized points," *Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH*, vol. 26, no. 2, pp. 71–78, July 1992.

[23] A. Bronstein, M. Bronstein, and R. Kimmel, *Numerical Geometry of Non-Rigid Shapes*, 1st ed. Springer Publishing Company, Incorporated, 2008.

[24] R. Gonzales and R. Woods, *Digital Image Processing (Third Edition)*. Pearson Education, Inc., 2008.

[25] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 27:1–27:27, 2011.

[26] W. J. Beksi, K. Choi, D. Canelon, and N. Papanikolopoulos, "The microvision robot and its capabilities," *submitted to the 2014 IEEE Conference on Intelligent Robots and Systems (IROS)*, 2014.

[27] D. Fehr, W. J. Beksi, D. Zermas, and N. Papanikolopoulos, "Occlusion alleviation through motion using a mobile robot," *Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA)*, 2014.

[28] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2010, pp. 3501–3508.

[29] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.