

## Object recognition in 3D scenes with occlusions and clutter by Hough voting

Federico Tombari  
 DEIS-ARCES  
 University of Bologna  
 Bologna, Italy  
 federico.tombari@unibo.it

Luigi Di Stefano  
 DEIS-ARCES  
 University of Bologna  
 Bologna, Italy  
 luigi.distefano@unibo.it

**Abstract**—In this work we propose a novel Hough voting approach for the detection of free-form shapes in a 3D space, to be used for object recognition tasks in 3D scenes with a significant degree of occlusion and clutter. The proposed method relies on matching 3D features to accumulate evidence of the presence of the objects being sought in a 3D Hough space. We validate our proposal by presenting a quantitative experimental comparison with state-of-the-art methods as well as by showing how our method enables 3D object recognition from real-time stereo data.

**Keywords**—Hough voting; 3D object recognition; surface matching;

### I. INTRODUCTION

Increasing availability of low-cost 3D sensors promotes research toward intelligent processing of 3D information. In this scenario, a major research topic concerns 3D object recognition, that aims at detecting the presence and estimating the pose of particular objects, represented in the form of 3D models, in 3D data acquired by a sensor such as a laser scanner, a Time-of-Flight camera, a stereo camera. Though a lot of effort has been devoted to the design of robust and discriminative 3D features aimed at reliably determining correspondences between 3D point sets [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], with scenes characterized by clutter and occlusions relatively few approaches are available for the task of detecting an object and estimating its pose based on feature correspondences. Indeed, many approaches for 3D object recognition are aimed at object retrieval in model databases and hence can not deal with clutter and occlusion. Besides, 3D object retrieval methods usually do not estimate the 3D pose of the object nor can deal with the presence of multiple instances of a given model. This is the case of Bag-of-3D Features methods [11], [12], [13], approaches based on the *Representative descriptor method* [8] and probabilistic techniques such as e.g. [14] (see [15] for a survey). On the other hand, the well-known Geometric Hashing technique can in principle be generalized seamlessly to handle 3D data [16], although it hardly withstands a significant degree of clutter [17], [18].

Instead, methods specifically designed for the task of object recognition in 3D scenes with clutter and occlusions should basically allow for discarding wrong feature

correspondences caused by these nuisance factors, so as to determine a subset that can be reliably deployed for detecting objects and estimating their poses. State-of-the-art approaches are mainly two. On one side, starting from a seed feature correspondence, correspondence grouping is carried out by iteratively aggregating those correspondences that satisfy geometric consistency constraints [19], [20]. The other main approach relies on clustering pose hypotheses in a 6-dimensional pose space, each correspondence providing a pose hypothesis (i.e. rotation and translation) based on the local Reference Frame (RF) associated with the two corresponding features [1], [3]. Once reliable feature correspondences have been selected, by either enforcement of geometric consistency or clustering, a final processing stage based on Absolute Orientation [21] and/or Iterative Closest Point (ICP) [22], can be performed to further validate the selected subset of correspondences and refine pose estimation.

In this work we propose a novel approach for object recognition in 3D scenes that can withstand clutter and occlusions and seamlessly allows for recognition of multiple instances of the model to be found. The proposed approach deploys 3D feature detection and description to compute a set of correspondences between the 3D model and the current scene. In addition, each feature point is associated with its relative position with respect to the centroid of the model, so that each corresponding scene feature can cast a vote in a 3D Hough space to accumulate evidence for possible centroid position(s) in the current scene. This enables simultaneous voting of all feature correspondences within a single tiny 3-dimensional Hough space. To correctly cast votes according to the actual pose(s) of the object(s) sought for, we rely on the local RFs associated with each pair of corresponding features. Next we briefly review the state-of-the-art concerning the Hough Transform, while in Section III we describe the proposed 3D Hough voting approach. Then, we present experiments on 3D scenes characterized by significant clutter and occlusions acquired by means of a laser scanner and two different stereo vision setups. The proposed experiments show how the proposed approach can be usefully deployed to perform 3D object recognition and allow to assess in quantitative terms how it

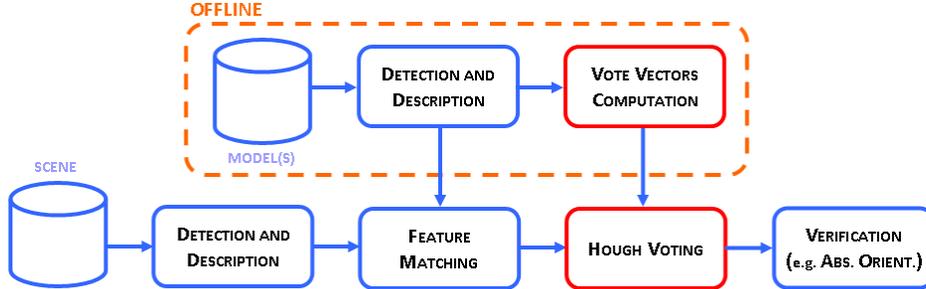


Figure 1. Use of the proposed Hough voting scheme (red blocks) in a 3D object recognition pipeline.

is significantly more discriminative and robust compared to the main existing methods.

## II. HOUGH VOTING

The Hough Transform (HT) [23] is a popular computer vision technique originally introduced to detect lines in 2D images. Successive modifications allowed the HT to detect analytical shapes such as circles and ellipses. Overall, the key idea is to perform a voting of the image *features* (such as edges and corners) in the parameter space of the shape to be detected. Votes are accumulated in an accumulator whose dimensionality equals the number of unknown parameters of the considered shape class. For this reason, although general in theory, this technique can not be applied in practice to shapes characterized by too many parameters, since this would cause a sparse, high-dimensional accumulator leading to poor performance and high memory requirements. By means of a matching threshold, peaks in the accumulator highlight the presence of a particular shape in the image. The Generalized Hough Transform (GHT) [24] extends the HT to detection of objects with arbitrary shapes, with each feature voting for a specific position, orientation and scale factor of the shape being sought. To reduce the complexity, the gradient direction is usually computed at each feature position to quickly index the accumulator.

The extension of the original HT formulation to 3D data is quite straightforward and allows detection of planes within 3D point clouds. Similarly to the 2D case, also the 3D HT has been modified to deal with additional 3D analytical shapes characterized by a small number of parameters, such as spheres [25] and cylinders [26]. A slightly more general class of objects, i.e. polyhedra, is considered in [27], with a Hough Voting method in two separate 3D spaces accounting for rotation and translation allowing to detect objects based on correspondences between vertex points established by matching straight edges. Unlike our proposal, though, this method cannot provide a unique pose hypothesis for each correspondences and, more importantly, cannot be applied to generic free-form objects.

More recently, an extension to the 3D domain of the GHT where gradient directions are substituted by point normals

has been proposed [28]. Yet, as pointed out in the paper, this technique has several disadvantages that hardly allow its direct application. In particular, to deal with generic rotations and translations in a 3D space the Hough space becomes 6-dimensional, leading to a high computational cost of the voting process (i.e.  $O(M \cdot N^3)$ ,  $M$  being the number of 3D points and  $N$  the number of quantization intervals) as well as to high memory requirements. Also, the resulting array would apparently be particularly sparse. Conversely, as described in the next section, with our approach the complexity of the voting process is  $O(M_f)$  ( $M_f$  being number of feature points) and the Hough space is 3-dimensional.

Another approach is represented by employment of Hough voting for the sake of hypothesis verification in 3D object recognition [29]. Unlike the previously discussed 3D extensions, this approach relies on feature correspondences established between the object model and the current scene. Correspondences are grouped in pairs and in triplets in order to vote, respectively, in two distinct 3-dimensional Hough spaces, one meant to parametrize rotation and the other to account for translation. Since otherwise the number of groups would grow prohibitively large, only a fraction of the feature correspondences is deployed in each of the two voting processes. Then, peaks in the Hough spaces indicate the presence of the sought object. Differently, with our approach only a single 3D Hough space is needed and, due to the deployment of the local RFs attached to features, each correspondence can cast its own vote, without any need for grouping correspondences. The latter difference renders our approach intrinsically more robust with respect to wrong correspondences caused by clutter and also allows for deployment of all of the available information (i.e. correspondences) within the voting process. Finally it is also worth pointing out that, due to the grouping stage, the method in [29] shares significant similarities with the geometric consistency approaches mentioned in previous section.

## III. THE PROPOSED 3D HOUGH VOTING ALGORITHM

Suppose we have an object model that we want to recognize in a scene, both in the form of 3D meshes. The flow of

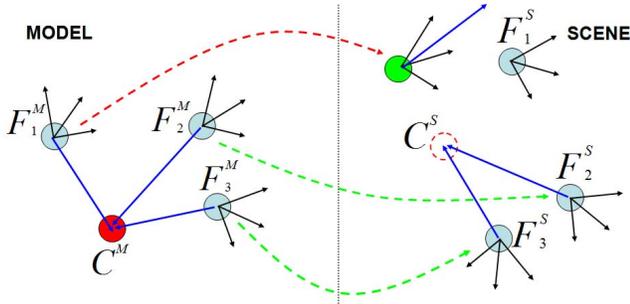


Figure 2. Example of 3D Hough Voting based on local RFs.

the proposed object recognition approach is sketched in Fig. 1. At first, interest points are extracted from both the model and the scene either by choosing them randomly or by means of a feature detector [1], [2], [3], [4]. They are represented as blue circles in the toy example shown in Fig. 2. Then, each feature point is enhanced with a piece of information representing a *description* of its local neighborhood, i.e. a 3D feature descriptor [3], [5], [6], [7], [8], [9], [10]. Typically, detecting and describing features of the model(s) can be performed once and for all off-line. Given a set of *described* features both in the model and in the scene, a set of feature correspondences (green arrows in Fig. 2) can be determined by thresholding, e.g., the euclidean distance between their descriptors. Due to the presence of nuisance factors such as noise, cluttered background and partial occlusions of the object being sought, typically this set includes also wrong correspondences (red arrow in Fig. 2).

The use of the proposed Hough voting scheme aims at accumulating evidence for the presence of the object being sought. If enough features vote for the presence of the object in a given position of the 3D space, then the object is detected and its pose is determined by means of the computed correspondences. In particular, at initialization time (i.e. offline) a unique *reference point*,  $C^M$ , is computed for the model (red circle in Fig. 2). In our experiments, we have selected the centroid of the model, though this particular choice does not affect the performance of the algorithm. Then, still at initialization, the vector between each feature and the centroid is computed and stored (blue arrows in Fig. 2). Since we want our method to be rotation and translation invariant, we can not store these vectors in the coordinates of the global RF (Reference Frame) since this would render them dependent on the specific RF of the current 3D mesh. Hence, and as sketched in Fig. 4, we need to compute an invariant RF for each feature extracted (i.e. a *local RF*) both in the model and in the scene. In particular, the local RF has to be efficiently computable (since we need to compute one RF for each feature) and very robust to disturbance factors (to hold on its invariance properties). Several proposals for local RF frames for 3D

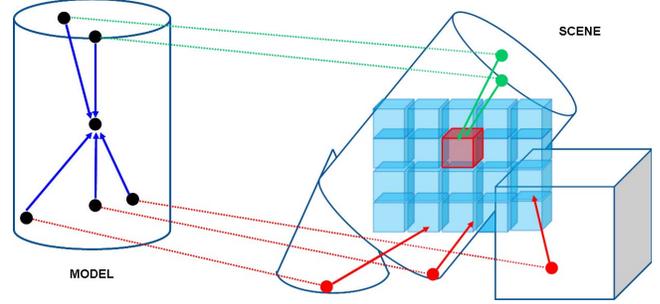


Figure 3. Toy example showing the proposed 3D Hough Voting scheme.

meshes are present in literature [3], [7], [5], [9], [10]. In our approach, we use the fully unambiguous local RF method proposed in [30]. This method selects as the local RF of a feature the three eigen vectors obtained by EVD of the distance-weighted covariance matrix of a local neighborhood of the feature. Then, a proper sign disambiguation procedure is applied to render the local RF unique and unambiguous.

Hence, in our approach, we perform an additional offline step (see Fig. 1), that represents the initialization of the Hough accumulator. Supposing that all point coordinates of the 3D model are given in the same global RF, for each model feature point  $F_i^M$  we compute first the vector between  $C^M$  and  $F_i^M$ :

$$V_{i,G}^M = C^M - F_i^M \quad (1)$$

Then, to render this representation rotation and translation invariant, each vector  $V_{i,G}^M$  has to be transformed in the coordinates given by the corresponding local RF (i.e. that computed on  $F_i^M$ , see Fig. 4) by means of the following transformation:

$$V_{i,L}^M = R_{GL}^M \cdot V_{i,G}^M \quad (2)$$

where  $\cdot$  represents the matrix product and  $R_{GL}^M$  is the rotation matrix where each line is a unit vector of the local RF of the feature  $F_i^M$ :

$$R_{GL}^M = [L_{i,x}^M L_{i,y}^M L_{i,z}^M]^T \quad (3)$$

The offline stage ends by associating to each feature  $F_i^M$  its vector  $V_{i,L}^M$ .

In the online stage, once correspondences between the model and the scene have been obtained as described above, each scene feature  $F_j^S$  for which a correspondence has been found ( $F_j^S \leftrightarrow F_i^M$ ) casts a vote for the position of the reference point in the scene. Since the computation of the local RF for  $F_j^S$  is invariant to rotations and translations, this allows to determine the transformation shown in Fig. 4 as  $R^M S_L$ , yielding  $V_{i,L}^S = V_{i,L}^M$ . Finally, we have to transform  $V_{i,L}^S$  into the global RF of the scene, by means of the following relationship:

$$V_{i,G}^S = R_{LG}^S \cdot V_{i,L}^S + F_j^S \quad (4)$$

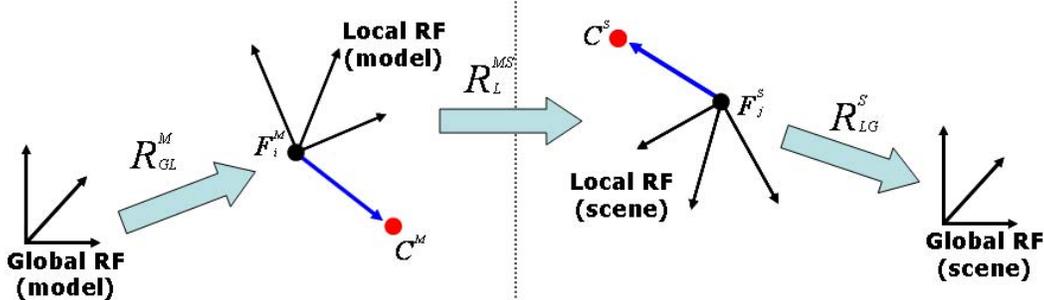


Figure 4. Transformations induced by the use of local RFs.

where  $R_{LG}^S$  is the rotation matrix obtained by lining up by columns the unit vectors of the local RF of the feature  $F_j^S$ :

$$R_{LG}^S = [L_{j,x}^S L_{j,y}^S L_{j,z}^S] \quad (5)$$

Thanks to these transformations, the feature  $F_j^S$  can cast a vote in a tiny 3D Hough space by means of vector  $V_{i,G}^S$  (see Fig. 3). Evidence for the presence of a particular object can then be evaluated by thresholding the peaks of the Hough space. Seamlessly, multiple peaks in the Hough space highlight the presence of multiple instances of the object being sought. In the specific case that only one instance of the object is sought for in the scene, we selected the bin in the Hough space having the maximum number of votes. Moreover, in order to render the peak selection process more robust to quantization effects, also the neighboring bins to those yielding a local maximum can be accounted for. In particular, we propose to threshold the Hough space peaks by adding to each bin the values of its 6 neighboring bins, since the presence of noise can cause correct correspondences to fall into neighboring bins. Then, when selecting the subset of correspondences to be used for successive stages (absolute orientation, pose estimation, ..) only those concerning the central bin of over-threshold maxima are accounted for, since this allows improved accuracy.

Hence, a subset of coherent correspondences providing evidence of the presence of the model is selected among all correspondences determined in the matching stage. As a successive step, the pose of the object can be determined by means of an Absolute Orientation algorithm [21]. As previously mentioned, this step can also be deployed as a further geometric verification of the presence of the model by thresholding the Root Mean Square Error (RMSE) between the subset of scene features and the transformed subset of model features. By repeating the same procedure for different models, the presence of different objects (i.e. those belonging to a reference library) can be evaluated.

#### IV. EXPERIMENTAL EVALUATION

This Section presents experimental results concerning the proposed Hough voting scheme for 3D object recognition.

In particular, we propose here two different Experiments. Experiment 1 concerns quantitative comparison between the proposed approach and the main existing methods. In particular, we compare our proposal to the algorithm presented in [1], as a representative of the approaches relying on clustering in the pose space, and to that described in [19], as a representative of methods based on geometric consistency. Hereinafter, we will refer to these two methods as, respectively, *Clustering* and *GC*. Given a set of models and a certain number of scenes (not containing all models but only a subset), all models are sought in each scene by means of each of the considered object recognition method. The outcome of an object recognition experiment can be a True Positive (TP), if the model sought for is present in the scene and correctly detected and localized, or a False Positive (FP), either if a model present in the image is detected but not correctly localized or the model is not present in the image but detected by the method. Hence, quantitative results are shown in terms of *Recall vs. 1-Precision* curves. To evaluate the localization, we first compute the RMSE between ground-truth feature positions (i.e. features mapped according to the ground-truth poses) and those obtained by applying the pose transformation estimated by the algorithm, then we assume a correct localization if the RMSE is lower than a fixed threshold (5 times the average model mesh resolution in our experiments).

We have performed our object recognition tests on two different dataset. The former (*Dataset 1*) is a publicly available 3D dataset built by Mian et al. [1] and made out of 5 models and 50 scenes, for a total of 250 object recognition instances. The second dataset (*Dataset 2*) has been acquired in our lab by means of the Spacetime Stereo (STS) technique [31], [32] and consists of 6 models and 12 scenes, for a total of 72 object recognition instances. Both datasets are quite challenging in terms of degree of clutter and occlusions (see Fig. 5). Since the focus of our experiment is on comparing the ability of the considered object recognition methods to filter out outliers due to clutter and occlusion, to attain a fair comparison we feed each of the 3 algorithms with exactly the same correspondences, which are determined as follows.

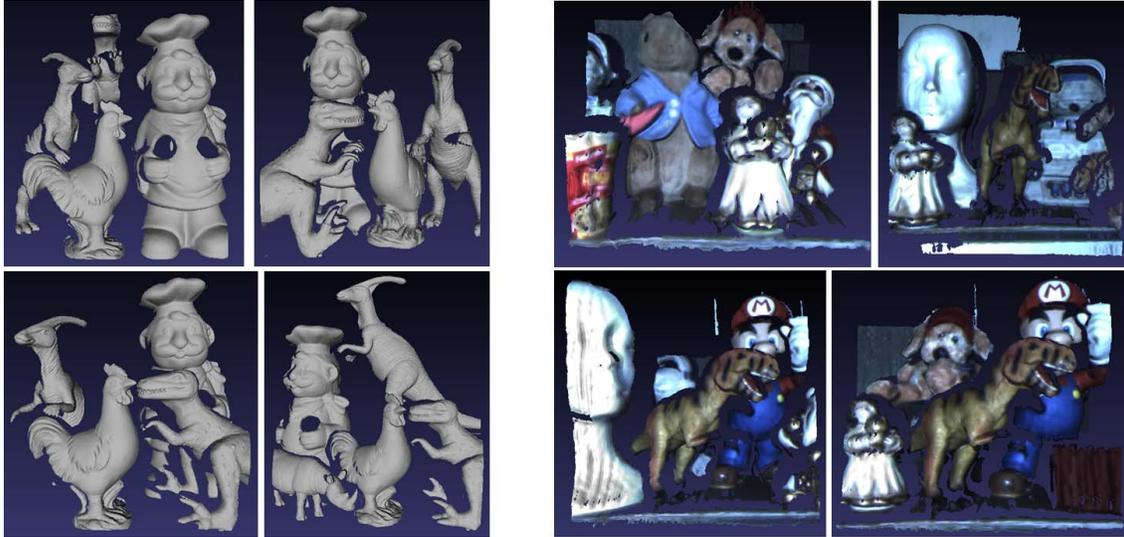


Figure 5. Sample scenes from *Dataset 1* (left) and *Dataset 2* (right).

As for detection of keypoints, a fixed number of 3D points is randomly extracted from each model (i.e. 1000) and each scene (i.e. 3000). We use a random detector for fairness of comparison, so as to avoid any possible bias towards some object recognition methods that might benefit of specific features found by specific 3D detectors. It is also worth pointing out that, overall, the use of a random detector increases the difficulty of the object recognition task. The descriptor and the feature matcher are also the same for all object recognition methods, and they are ran with the same parameter values. More precisely, we use the recent hybrid signature-histogram descriptor proposed in [30], while for feature matching we rely on the euclidean distance and deploy a well-know efficient indexing technique (i.e. *Kd-tree* [33]) to speed-up the computations. Instead, a specific tuning has been performed for the parameters of the object recognition techniques, i.e. the Hough space bin dimension for the proposed approach, the geometrical consistency threshold for *GC* and the number of k-means clusters for *Clustering*. The number of k-means iterations for *Cluster* has been set to 100 since we noted that the algorithm performance was not sensible to this parameter. Tuning has been performed over *Dataset 1*, then the same parameters have been used for the experiments on *Dataset 2*.

Figure 6 reports the results concerning Experiment 1. As it can be seen, on both datasets the proposed approach yields improved recognition capabilities with respect to *Clustering* and *GC*. In particular, on both datasets the proposed Hough-based recognition approach always yields higher Recall at the same level of Precision. Moreover, on *Dataset 2* a specific threshold value exists that allows our approach to yield 100% correct recognitions and no false detections. It is

also worth pointing out that in our experiments on *Dataset 1* the *GC* approach, unlike the other 2, turned out to be particularly sensitive to the threshold parameter. In particular, the selected correspondence subset is always characterized by a smaller cardinality compared to the other two approaches, this denoting a worse capability of consensus grouping for *GC*. As for efficiency, with these parameters and in our experiments the proposed approach and *GC* are overall much more efficient than *Clustering* (i.e. they run more than one order of magnitude faster).

As for Experiment 2, we evaluate qualitatively the performance of the proposed algorithm in an online object recognition framework based on stereo data. In particular, the stereo setup is based on two webcams and a stereo processing algorithm that computes range maps using the real-time OpenCV stereo matching algorithm [34]. To improve the performance of the stereo algorithm, and hence the quality of the retrieved 3D data, a binary pattern is projected on the scene using a standard projector. This helps stereo matching by enriching homogeneous regions with a certain degree of texture. Nevertheless, we wish to point out that such a real-time stereo setup is very challenging for the purpose of 3D object recognition, since the 3D data attainable for both models and scenes are still notably noisy and with missing parts (see Fig. 7). As for the experiment, we try to detect two models previously acquired with the same stereo setup within scenes characterized by significant clutter and occlusions (see again Fig. 7). In Figure 7, the blue bounding box shows the pose of the model estimated by our method. Such results shows that our proposal allows for carrying out object recognition in presence of noisy 3D data, as those attainable with a real-time stereo setup, and in scenes characterized by a significant amount of clutter

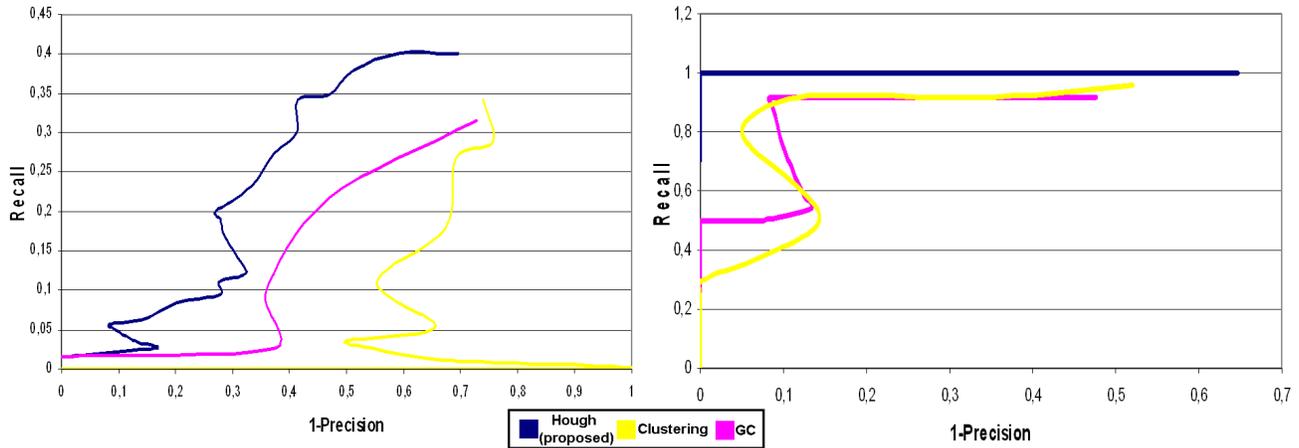


Figure 6. Experiment 1: Precision-recall curves of the 3 evaluated object recognition approaches. Left: 250 object tests on Dataset 1. Right: 72 tests on Dataset 2.

and occlusions. Additional results are provided as a video, included in the supplementary material, showing additional object recognition experiments performed within the stereo object recognition framework presented in this Section.

## V. CONCLUSIONS

We have proposed a novel approach based on 3D Hough voting for detection and localization of free-form objects from range images, such as those provided by laser scanners and stereo vision sensors. The experimental evaluation shows that our method outperforms clearly the algorithms chosen as representative of the two main existing approaches, i.e. those relying on geometric consistency and on pose space clustering. We have also provided results proving that our method is effective in detecting 3D objects from the very noisy 3D data attained through a real-time stereo setup, in particular in complex scenes characterized by significant clutter and occlusion.

## REFERENCES

- [1] A. Mian, M. Bennamoun, and R. Owens, "On the repeatability and quality of keypoints for local feature-based 3d object retrieval from cluttered scenes," *Int. J. Computer Vision*, p. to appear, 2009.
- [2] H. Chen and B. Bhanu, "3d free-form object recognition in range images using local surface patches," *Pattern Recognition Letters*, vol. 28, no. 10, pp. 1252–1262, 2007.
- [3] Y. Zhong, "Intrinsic shape signatures: A shape descriptor for 3d object recognition," in *Proc. 3DRR Workshop (in conj. with ICCV)*, 2009.
- [4] J. Novatnack and K. Nishino, "Scale-dependent 3d geometric features," in *Proc. Int. Conf. on Computer Vision*, 2007, pp. 1–8.
- [5] —, "Scale-dependent/invariant local 3d shape descriptors for fully automatic registration of multiple sets of range images," in *ECCV*, 2008, pp. 440–453.
- [6] A. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3d scenes," *PAMI*, vol. 21, no. 5, pp. 433–449, 1999.
- [7] C. S. Chua and R. Jarvis, "Point signatures: A new representation for 3d object recognition," *IJCV*, vol. 25, no. 1, pp. 63–85, 1997.
- [8] A. Frome, D. Huber, R. Kolluri, T. Bülow, and J. Malik, "Recognizing objects in range data using regional point descriptors," in *ECCV*, vol. 3, 2004, pp. 224–237.
- [9] F. Stein and G. Medioni, "Structural indexing: Efficient 3-d object recognition," *PAMI*, vol. 14, no. 2, pp. 125–145, 1992.
- [10] A. Mian, M. Bennamoun, and R. Owens, "A novel representation and feature matching algorithm for automatic pairwise registration of range images," *IJCV*, vol. 66, no. 1, pp. 19–40, 2006.
- [11] Y. Liu, H. Zha, and H. Qin, "Shape topics: A compact representation and new algorithms for 3d partial shape retrieval," in *Proc. CVPR*, 2006.
- [12] X. Li, A. Godil, and A. Wagan, "Spatially enhanced bags of words for 3d shape retrieval," in *Proc. ISVC*, 2008, pp. 349–358.
- [13] R. Ohbuchi, K. Osada, T. Furuya, and T. Banno, "Salient local visual features for shape-based 3d model retrieval," in *Proc. Int. Conf. on Shape Modeling and Applications*, 2008, pp. 93–102.
- [14] G. Hetzel, B. Leibe, P. Levi, and B. Schiele, "3d object recognition from range images using local feature histograms," in *Proc. CVPR*, 2001, pp. 394–399.

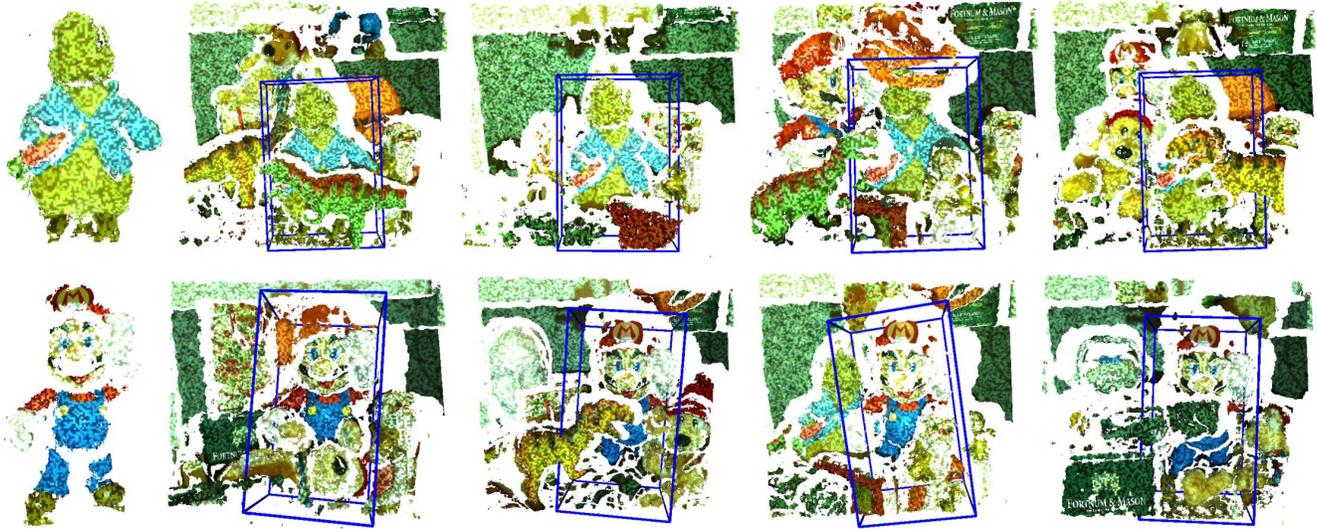


Figure 7. Results yielded by the proposed approach in Experiment 2.

- [15] J. W. H. Tangelder and R. C. Veltkamp, "A survey of content based 3d shape retrieval methods," in *Proc. Shape Modeling International*, 2004, pp. 145–156.
- [16] Y. Lamdan and H. J. Wolfson, "Geometric hashing: A general and efficient model-based recognition scheme," in *Proc. Int. Conf. on Computer Vision*, 1988, pp. 238–249.
- [17] W. Grimson and D. Huttenlocher, "On the sensitivity of geometric hashing," in *Proc. Int. Conf. on Computer Vision*, 1990, pp. 334–338.
- [18] Y. Lamdan and H. Wolfson, "On the error analysis of 'geometric hashing'," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1991, pp. 22–27.
- [19] A. E. Johnson and M. Hebert, "Surface matching for object recognition in complex 3-d scenes," *Image and Vision Computing*, vol. 16, pp. 635–651, 1998.
- [20] H. Chen and B. Bhanu, "3d free-form object recognition in range images using local surface patches," *J. Pattern Recognition Letters*, vol. 28, pp. 1252–1262, 2007.
- [21] B. Horn, "Closed-form solution of absolute orientation using unit quaternions," *J. Optical Society of America A*, vol. 4, no. 4, pp. 629–642, 1987.
- [22] Z. Zhang, "Iterative point matching for registration of free-form curves and surfaces," *Int. J. Computer Vision*, vol. 13, no. 2, pp. 119–152, 1994.
- [23] P. Hough, "Methods and means for recognizing complex patterns," US Patent 3069654, 1962.
- [24] D. Ballard, "Generalizing the hough transform to detect arbitrary shapes," *SPIE Proc. on Vision Geometry X*, vol. 13, no. 2, pp. 111–122, 1981.
- [25] G. Vosselman, B. Gorte, G. Sithole, and T. Rabbani, "Recognising structure in laser scanner point cloud," *Int. Arch. of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 46, pp. 33–38, 2004.
- [26] T. Rabbani and F. Van Den Heuvel, "Efficient hough transform for automatic detection of cylinders in point clouds," in *ISPRS Ws. Laser Scanning*, 2005, pp. 260–65.
- [27] H. Tsui and C. Chan, "Hough technique for 3d object recognition," *IEE Proceedings*, vol. 136, no. 6, pp. 565–568, 1989.
- [28] K. Khoshelham, "Extending generalized hough transform to detect 3d objects in laser range data," in *Proc. ISPRS Workshop on Laser Scanning*, 2007, pp. 206–210.
- [29] A. Ashbrook, R. Fisher, C. Robertson, and N. Werghi, "Finding surface correspondence for object recognition and registration using pairwise geometric histograms," in *Proc. European Conference on Computer Vision*, 1998, pp. 674–686.
- [30] Anonymous, "Anonymous submission," in *Accepted to ECCV 2010*, 2010.
- [31] J. Davis, D. Nehab, R. Ramamoorthi, and S. Rusinkiewicz, "Spacetime stereo : A unifying framework for depth from triangulation," *PAMI*, vol. 27, no. 2, pp. 1615–1630, 2005.
- [32] L. Zhang, B. Curless, and S. Seitz, "Spacetime stereo: Shape recovery for dynamic scenes," in *Proc. CVPR*, 2003.
- [33] J. Beis and D. Lowe, "Shape indexing using approximate nearest-neighbour search in high dimensional spaces," in *Proc. CVPR*, 1997, pp. 1000–1006.
- [34] <http://opencv.willowgarage.com>.