A Non-Cooperative Game for 3D Object Recognition in Cluttered Scenes

Andrea Albarelli, Emanuele Rodolà, Filippo Bergamasco, and Andrea Torsello
Dipartimento di Scienze Ambientali, Informatica e Statistica
Università Ca' Foscari Venezia
Venice, Italy
Email: albarelli@unive.it rodola@dsi.unive.it fbergama@dsi.unive.it torsello@dsi.unive.it

***Abstract*—During the last few years a wide range of algorithms and devices have been made available to easily acquire range images. To this extent, the increasing abundance of depth data boosts the need for reliable and unsupervised analysis techniques, spanning from part registration to automated segmentation. In this context, we focus on the recognition of known objects in cluttered and incomplete 3D scans. Fitting a model to a scene is a very important task in many scenarios such as industrial inspection, scene understanding and even gaming. For this reason, this problem has been extensively tackled in literature. Nevertheless, while many descriptor-based approaches have been proposed, a number of hurdles still hinder the use of global techniques. In this paper we try to offer a different perspective on the topic. Specifically, we adopt an evolutionary selection algorithm in order to extend the scope of local descriptors to satisfy global pairwise constraints. In addition, the very same technique is also used to shift from an initial sparse correspondence to a dense matching. This leads to a novel pipeline for 3D object recognition, which is validated with an extensive set of experiments and comparisons with recent well-known feature-based approaches.**

***Keywords*-Object Recognition; Rigid Alignment; Game Theory; Object in Clutter;**

I. Introduction

In the recent past, the acquisition of 3D data was only viable for research labs or professionals that could afford to invest in expensive and difficult to handle high-end hardware. However, due to both technological advances and increased market demand, this scenario has been altered significantly: Semi-professional range scanners can be found at the same price level of a standard workstation, widely available software stacks can be used to obtain reasonable results even with cheap webcams, and, finally, range imaging capabilities have been introduced even in very low-end devices such as game controllers. Given this trend, it is safe to forecast that range scans will be so easy to acquire that they will complement or even replace traditional intensity based imaging in many computer vision applications. The added benefit of depth information can indeed enhance the reliability of most inspection and recognition tasks, as well as providing robust cues for scene understanding or pose estimation. Many of these activities include fitting a known model to a scene as a fundamental step. For instance, a setup for in-line quality control within a production line, could need to locate the manufactured objects that are meant to be measured [1]. Moreover, a range-based SLAM system [2], can exploit the position of known 3D reference objects to achieve a more precise and robust robot localization. Finally, non-rigid fitting could be used to recognize hand or whole-body gestures in next generation interactive games or novel man-machine interfaces [3]. The matching problem in 3D scenes shares many aspects with object recognition and location in 2D images: The common goal is to find the relation between a model and its transformed instance (if any) in the scene. In both cases, transformations could include uniform and non-uniform scaling, differences in pose or partial modification of the shape. They also share common hurdles, such as measurement errors on intensities or point positions, and indirect changes in the appearance due to occlusion or the simultaneous presence in the scene of extraneous objects that can act as distractions. Feature-based approaches, both in 2D and in 3D, adopt descriptors that are associated to single points respectively on the image or on the object surface. In principle, each feature can be matched individually by comparing the descriptors, which of course decouples the effect of partial occlusion. In the 2D domain, intensity based descriptors such as SIFT [4] have proven to be very distinctive and be able to perform very well even with naive matching methods that do not include any global information [5]. However, the problem of balancing local and global robustness is more binding

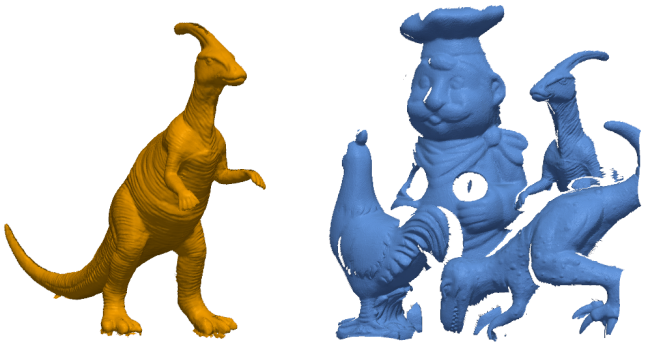


Figure 1. A typical 3D object recognition scenario. Clutter of the scene and occlusion due to the geometry of the ranging sensor seriously hinder the ability of both global and feature-based techniques to spot the model.

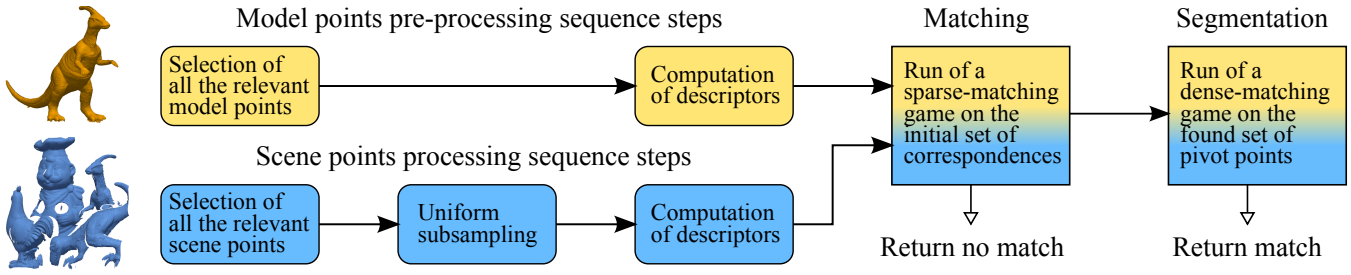


Figure 2. An overview of the object recognition pipeline presented (see text for description).

with 3D scenes than with images, as no natural scalar field is available on surfaces and thus feature descriptors tend to be less distinctive. In practice, global or semi-global inlier selection techniques are often used to avoid wrong correspondences. This, while making the whole process more robust to a moderate number of outliers, could introduce additional weaknesses. For instance, if a RANSAC-like inlier selection is applied, occlusion coupled with the presence of clutter (*i.e.*, unrelated objects in the scene) can easily lower the probability for the process to find the correct match. The limited distinctiveness of surface features can be tackled by introducing scalar quantities computed over the local surface area. This is the case, for instance, with values such as mean curvature, Gaussian curvature or shape index and curvedness, which can be constructed in order to classify surface patches into types such as pits, peaks or saddles [6]. Unfortunately, this kind of characterization has proven to be not very selective for matching purposes, since it is frequent to obtain similar values in many different locations. Another approach is to augment the point data with additional scalar values that can be obtained during the acquisition process. To this extent, the use of natural textures coming from the scanned object have shown to allow good performance since they show high variability and can be used to compute descriptors similar to those usually adopted in the 2D domain [7]. Still, textures cannot be obtained from all the surface digitizing techniques and, even when available, their usability for descriptor extraction strongly depends on the appearance of the scanned object. To overcome the limitations of scalar descriptors, methods that gather information from the whole neighborhood of each point to characterize have been introduced. Such methods can be roughly classified in approaches that define a full reference frame for each point (for instance, by using PCA) and techniques that only need a reference axis (usually some kind of normal direction for the point). When a full reference frame is available it is possible to build very discriminative descriptors [8], [9]. Unfortunately, noise and differences in the mesh could lead to instabilities in the reference frame, and thus to a brittle descriptor. By converse, methods that just require a reference axis (and are thus invariant to the rotation of the frame) trade some descriptiveness to gain greater robustness. These latter techniques almost invariably build histograms based on some properties of points falling in a cylindrical volume centered and aligned to the reference axis. The most popular histogram-based approach is certainly Spin Images [10], but many others have been proposed in literature [11], [12]. Lately, an approach that aims to retain the advantages of both full reference frames and histograms has been introduced [13]. Other recent contributions include scale invariant detectors [14], [15] and tensor-based descriptors [16]. Any of these interest point descriptors can be used to find correspondences between a model and a 3D scene that could possibly contain it. Most of the cited papers, in addition to introducing the descriptor itself, propose some matching technique. These span from very naive approaches, such as associating each point in the model with the point in the scene having the most similar descriptor, to more advanced techniques such as customized flavors of PROSAC and specialized keypoint matchers that exploit locally fitted surfaces for computing depth values to use as feature components [17].

In this paper we introduce a novel feature-based 3D object recognition pipeline crafted to deal in a robust manner with both strong occlusion and clutter. This happens by adopting a histogram-based local surface descriptor to find a set of matching candidates among a selection of relevant points on the model and the scene. Such candidates are then let to compete in a non-cooperative game where payoffs are proportional to the degree of Euclidean compatibility between them. This leads to a smaller set of sparse but reliable surviving matches which, in turn, will be used as the seeds for an additional game aimed at the selection of a denser population. While the use of Game Theory for matching has already been explored [18], the contribution of this paper is threefold. It introduces a novel pipeline that outperforms the state-of-the-art for 3D object recognition in clutter. Further, it suggests a simple but general rule for samples selection for the purpose of recognition. Finally, it defines a new kind of game for building a dense surface correspondence starting from a sparse set of pivot points, which can be useful also for other matching techniques.

II. A Game-Theoretic Pipeline for Recognition

Following [19], we base our matching framework on the recently introduced Game-Theoretic techniques for inlier selection. The complete pipeline we are proposing is made up of a preprocessing step and two non-cooperative games (see Fig. 2). The preprocessing is performed both on the model and on the scene. This step involves an initial selection of relevant points on the respective surfaces. The relevance criteria will be explained in the next section, however, in this context the general meaning of the culling is to avoid surface patches that are not significant from a matching standing point, such as flat areas. All the interest points on the model are kept while those on the scene are uniformly subsampled. This makes sense for many reasons. In many applications the set of models does not change in time, and thus descriptors must be computed just once. In addition, as explained in the following sections, the direction of the matching will be from the scene to the model and having less source than target points allows the game to proceed faster without compromising accuracy. Finally, the model tends to be measured with greater accuracy (either because more time can be spent on it or because it comes from a CAD model). A descriptor is computed for all the retained points, and these are used to build the initial candidates that will be fed to two matching games. The games are played respectively to build a coarse initial set of fiducial correspondences and to make those into dense matches by exploiting neighborhood relationships.

In general, a matching game [19] can be built by defining just four basic entities: a set of model points M, a set of data points D, a set of candidate correspondences $S \subseteq M \times D$ and a pairwise compatibility function between them $\Pi : S \times S \rightarrow \mathbb{R}^+$. The goal of the gameplay is to operate a (natural) selection among the elements in the initial set S. This happens by setting up a non-cooperative game where the set S represents the available strategies and Π the payoffs between them. In this game, a real-valued vector $\mathbf{x} = (x_1, \ldots, x_{|S|})^T$ that lies in the $|S|$-dimensional standard simplex

$$\Delta^{|S|} = \left\{ \mathbf{x} \in \mathbb{R}^{|S|} \ : \ x_i \geq 0, i = 1 \ldots |S|, \ \sum_{i=1}^{|S|} x_i = 1 \right\}$$

represents the amount of population that plays each strategy i at a given time. The game starts by setting the initial population around the barycenter (to be fair with respect to each strategy). Then, the population can be evolved at discrete steps by applying the replicator dynamics equation:

$$\mathbf{x}_i(t+1) = \mathbf{x}_i(t) \frac{(\Pi \mathbf{x}(t))_i}{\mathbf{x}(t)^T \Pi \mathbf{x}(t)} \tag{1}$$

where Π is a matrix that assigns to row i and column j the payoff (compatibility) between strategies (correspondences) i and j. Under very weak assumptions it can be shown that such dynamics must converge (in an infinite time) to a *Nash equilibrium*, *i.e.*, a point in the simplex where the average payoff obtained by the population is a local maximum constant for each strategy. In addition, the values of the elements of $\mathbf{x}$ are proportional to the degree of compatibility of each strategy with the equilibrium [19]. In practice, a much faster convergence to the equilibrium can be obtained by replacing the iteration in equation (1) with the adaptive exponential replicator dynamics introduced in [20]. Since we defined the payoff as the compatibility between candidates, these are all desirable properties from a selection standpoint. In our context, M and D always correspond to the retained model and scene points, while S and Π will be defined differently for the sparse and dense matching game. Specifically, for the sparse game the construction of S will be driven by descriptor similarity, whereas positional information can be used in the segmentation game. Likewise, the payoff Π will be proportional to the different notions of compatibility.

A. Feature Detection and Description

For both efficiency and robustness reasons, the proposed matching technique works on a subset of the model and scene data. First, a culling of all the vertices is performed. This happens by computing for each point a single-component *Integral Hash* [21] at a given support scale σ, and thus retaining only those samples that obtain a negative value (*i.e.*, that belongs to a concave surface patch). In practice, this means that we are avoiding flat and convex areas which we experimented to be less distinctive. By modulating the value of σ a more or less selective sample selection can be made (see Fig. 3). All the model relevant points are kept. By contrast, an optional uniform subsampling can be performed on the relevant points in the scene. Finally, a descriptor vector must be computed for each vertex to be matched. To this extent, any of the descriptors discussed in the introduction could be used; however, after an initial round of tests, SHOT [13] was chosen as it obtains the best performance over the whole pipeline. In the experimental section both the influence of the relevant point selection and of the adopted descriptor are studied.

B. Sparse Matching Game

In this matching game the set of candidates S is built by associating each reference point in the scene with the k nearest points in the model in terms of the descriptor:

$$S = \{(a, b) \in D \times M | b \in dn_k(a)\}, \tag{2}$$

where $dn_k(a)$ is the set of the k model vertices with the nearest descriptor with respect to the descriptor of a. In practice, this means that each sample in the scene is considered to be a possible match with samples in the model that exhibit similar surface characteristics, and we limit the

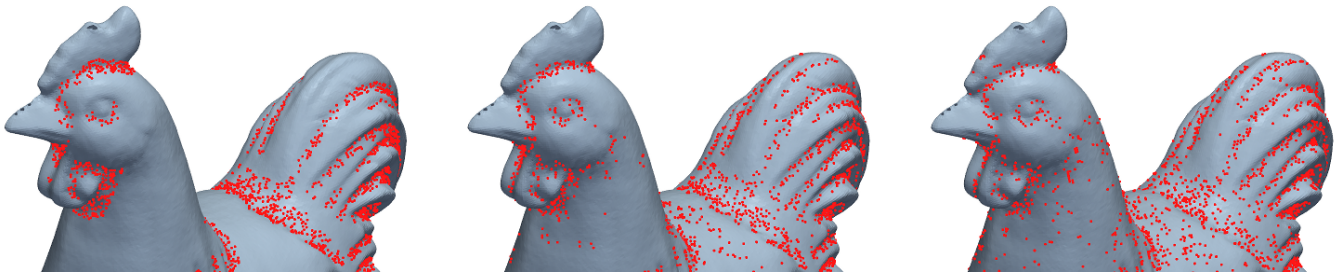


Figure 3. In order to avoid mismatches and reduce the convergence time it is important to use only relevant points. Model vertices selected with a σ respectively equal to 8, 5 and 2 times the median model edge are shown from left to right.

number of "attempts" to k. It should be noted that candidates are built from scene to model. Since we are interested in finding a correspondence between the model and part of the surface in the scene, we are looking for a subset of candidates that enforce the Euclidean rigidity constraint. Such candidates are likely to lay on the same surface both in the scene and in the model and thus to be a viable solution. To this extent, we define this distance measure between pairs of strategies in S as

$$\delta((a_1, b_1), (a_2, b_2)) = \frac{min(|a_1 - a_2|, |b_1 - b_2|)}{max(|a_1 - a_2|, |b_1 - b_2|)} \quad (3)$$

where a_1, a_2, b_1 and b_2 are respectively the two model and scene vertices in the compared strategies. The value of δ will be 1 if the corresponding source and destination points are separated by exactly the same Euclidean distance. By contrast, δ will be small when the two pairs exhibit very different distances. This kind of check will succeed with correct pairs and will give false positives only for a small amount of cases, those preserving the rigid constraint by chance. However, since our game is seeking for a large group of candidates with large mutual payoff, such sneaky outliers will be filtered out with high probability by the other strategies that participate to the Nash equilibrium. Finally, we also want to avoid many-to-many matches, since we do not expect any point in the scene to correspond to more than one point in the model. This can be done easily by forcing to 0 the compatibility between candidates that share the same source or destination vertex [19]. Thus, the final payoff for the sparse matching game that we are defining will be

$$\Pi = \begin{cases} \delta((a_1, b_1), (a_2, b_2)) & \text{if } a_1 \neq a_2 \text{ and } b_1 \neq b_2 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

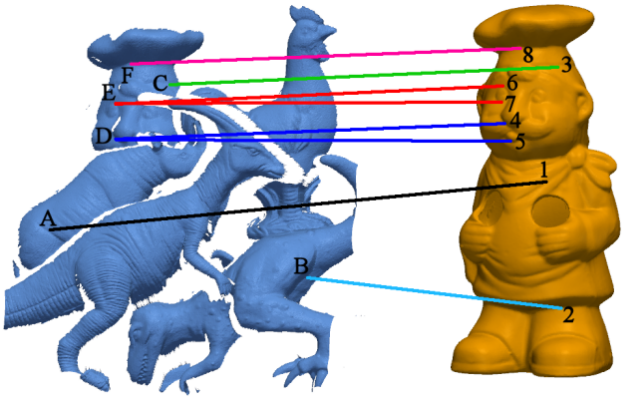
Once the candidate set and the payoff matrix are built, the game is started from the barycenter of the simplex: when a stable state is reached, all the strategies supported by a large percentage of the population (above a threshold based on the most played strategy) are considered non-extinct and retained as correct matches (see Fig. 4). If the surviving matches are more than a fixed minimum (set to 8 in our experiments), then the object is recognized and its pose can be computed.

C. Dense Matching Game

If the matching game succeeds, then a fiducial set of correspondences has been found; it would be interesting to use these matches as a seed for segmenting the surface belonging to the model from the scene. In most cases, growing from the fiducial points to the connected part of the range surface would be enough. However, in cluttered range images the unintentional merging between surfaces of different objects is quite frequent, thus a better selection mechanism could be useful. In order to demonstrate the flexibility of the Game-Theoretic framework we define another type of game to solve these problems (albeit other more direct solutions are also possible). We start by using the initial correspondences to estimate the rigid transformation between the model and the scene using the closed form method proposed in [22]. The computed transformation is then used to register the model within the scene coordinate system. At this point, if the initial matches are correct, the model vertex corresponding to each scene point should be in its neighborhood. For this reason we define the set of candidates S as

$$S' = \{(a, b) \in D \times M | b \in en_k(a)\} \quad (5)$$

where $en_k(a)$ is the set of the k nearest model vertices with respect to the Euclidean distance from a. Note that since we trust the alignment to be good (even if it is not perfect) we do not need point descriptors anymore. We want to enforce the rigidity constraint for this game as well, thus the compatibility δ defined in the previous game could still be used. However we would like to apply two modifications to the payoff function. The first one is the introduction of an exponent α to the measured compatibility. This is needed because within this game all the points are very close to each other and small variations in the position of a point in the model and its scene correspondence can easily lead to low compatibility. The second modification is related to the observation that we are interested in operating a model-driven segmentation of the scene, thus we are not really



	A1	B2	C3	D4	D5	E6	E7	F8
A1	0	0.12	0.77	0.83	0.98	0.77	0.66	0.75
B2	0.12	0	0.05	0.21	0.37	0.07	0.32	0.17
C3	0.77	0.05	0	0.99	0.6	0.99	0.99	0.7
D4	0.83	0.21	0.99	0	0	0.99	0.99	0.96
D5	0.98	0.37	0.6	0	0	0.9	0.88	0.69
E6	0.77	0.07	0.99	0.99	0.9	0	0	0.98
E7	0.66	0.32	0.99	0.99	0.88	0	0	0.99
F8	0.75	0.17	0.7	0.96	0.69	0.98	0.99	0

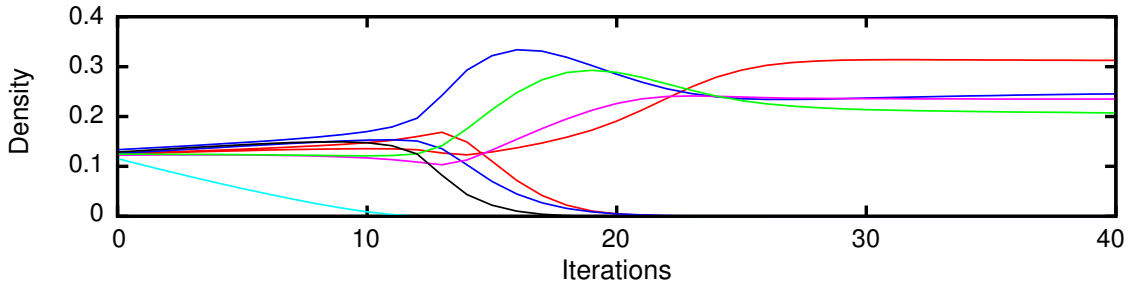


Figure 4. An example of the evolutionary process (with real data). A set of 8 matching candidates is chosen (upper left), a payoff matrix is built to enforce their respective Euclidean constraints (upper right, note that cells associated to many-to-many matches are set to 0) and the replicator dynamics are executed (bottom graph). At the start of the process the population is set around the barycenter (at 0 iterations). This means that initially the vector $\mathbf{x}$ represents a quasi-uniform probability distribution. After a few evolutionary iterations the matching candidate B2 (cyan) is extinct. This is to be expected since it is a clearly wrong correspondence and its payoff with respect to the other strategies is very low (see the payoff matrix). After a few more iterations, strategy A1 vanishes as well. It should be noted that strategies D4/D5 and E6/E7 are mutually exclusive, since they share the same scene vertex. In fact, after an initial plateau, the demise of A1 breaks the tie and finally E6 prevails over E7 and D4 over D5. After just 30 iterations the process stabilizes and only 4 strategies (corresponding to the correct matches) survive.

looking for a one-to-one correspondence between points, but rather we are trying to match each vertex in the scene to at least one reasonable vertex in the model to which it belongs. To this extent, the one-to-one constraint enforced in the previous game can be relaxed to a many-to-one constraint and the payoff function can be defined as

$$\Pi' = \begin{cases} \delta((a_1, b_1), (a_2, b_2))^\alpha & \text{if } a_1 \neq a_2 \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

Again, this segmentation game can be played by starting from the barycenter of the standard simplex and letting the population evolve by means of appropriate dynamics.

III. Experimental Results

In order to evaluate the performance of the proposed pipeline we performed a wide range of tests and comparisons with recent techniques. To offer a fair comparison we used the model/scene dataset adopted in [14], [16], [17]. This dataset is composed of five high resolution models scanned from real objects (chef, dino1, dino2, chicken and rhino), plus about two hundred range scans of these objects under various conditions of occlusion (due to the overlap of objects and limits on the field of view of the sensor) and clutter (due to the presence of many objects in the scene). All the tests were performed on a standard desktop PC equipped with a Core Duo processor clocked at 1.6Ghz. The evolutionary process makes use of the adaptive exponential replicator dynamics [20]. The minimum number of matches to assume the model as recognized in the scene was 8. The value of α for the segmentation game was 0.2. For the sparse matching game, the SHOT descriptor [13] was used.

A. Comparison with the State-of-the-art

In Fig. 5 we compare our results in terms of recognition rate with recent state-of-the-art algorithms (respectively [14], [16], [17]) and with the well-known 3D Spin Image matching technique [10], which is often used as a baseline. Looking at the recognition rate with respect to model occlusion, the proposed pipeline outperforms even the most recent techniques. Regarding the evaluation of the effects of clutter we could compare our algorithm only to [14], since an implementation for the other approaches and the data they used were not available. Still, it is apparent that the Game-Theoretic approach obtains good recognition with uniform performance. Some examples of critical scenes

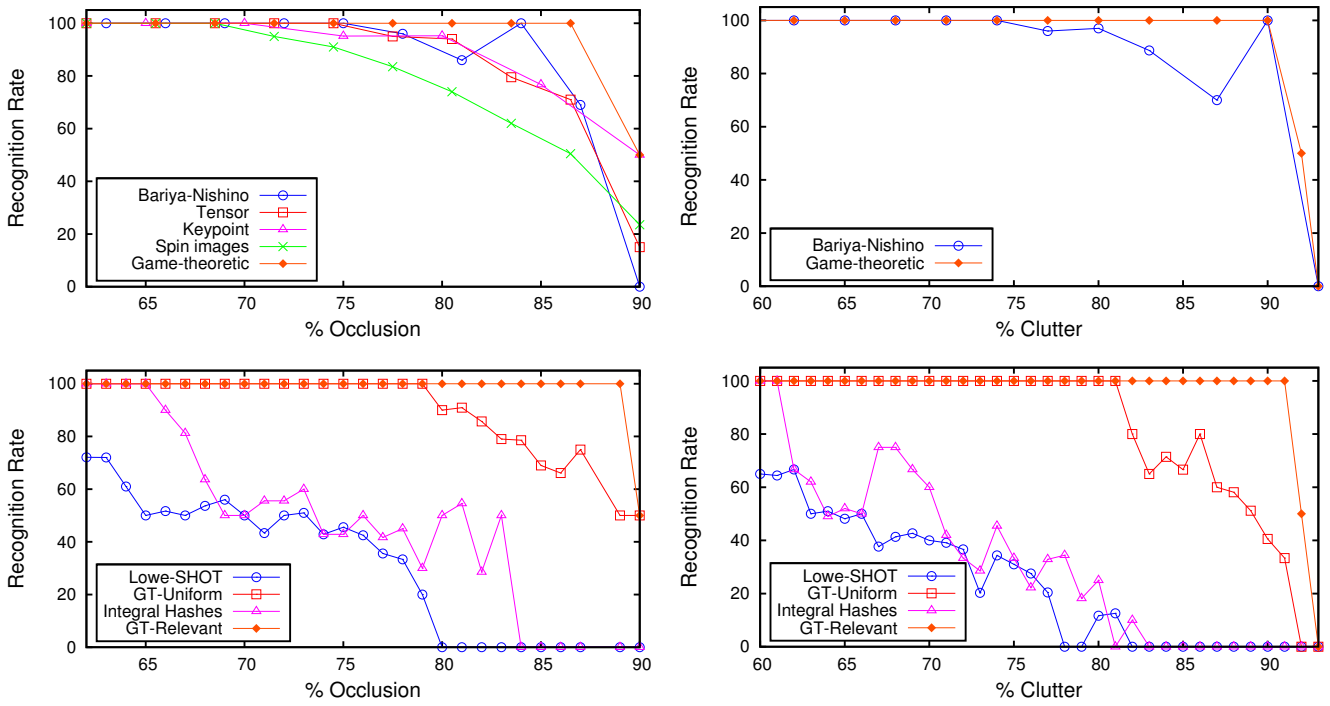


Figure 5. In the top row the recognition rate of our pipeline is compared with state-of-the-art techniques, which are outperformed with respect to both occlusion and clutter. In the bottom row the contribution of each part of the overall approach is tested separately (see text for details).

where the proposed technique fixes matches missed by the other methods are shown in Fig. 7. The behavior with respect to false positives has not been plotted since the proposed pipeline does not get any throughout the whole dataset. In the second row of Fig. 5 several combinations of components of the pipeline are evaluated one at a time in order to shape their respective contribution. Specifically, we show the results obtained using the same descriptor [13] with the classical matcher proposed by Lowe [5] (Lowe-SHOT), the Game-Theoretic matcher without operating the initial relevance-based sampling (GT-Uniform), the descriptors and matching proposed in [21] (Integral-Hashes) and finally the full proposed pipeline (GT-Relevant). It is apparent that the proposed pipeline only works with all the components in place (note that with these latter experiments the sampling of the plots is more dense).

B. Resilience to Noise

All the experiments so far have been done using a dense model and slightly less dense scenes produced with a range scanner. Although there is not an exact correspondence between model and scenes, they are still very similar by construction. It would be interesting to study the performance of the proposed method in presence of positional noise. To do so, we added Gaussian displacement of varying intensity to each vertex in the scene. In Fig. 6 the results obtained with two different SHOT parameterizations are shown. As expected, performance gets lower as the noise level increases; still, reasonable recognition rates are maintained also with a moderate amount of noise (with standard deviation equal to 30% the median edge length).

C. Sparse to Dense Matching

An example of the dense matching game used to segment the parts of the scene belonging to the model is shown in the last row of Fig. 7. Segmented points are highlighted both on the model and on the scene. In this case the naive growing approach would have failed since in the range scene the chef's foot is partially merged with the hind foot boundaries of dino1.

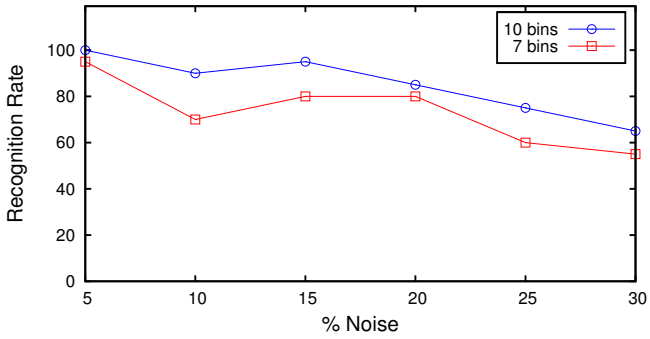


Figure 6. Evaluation of the robustness of the proposed pipeline with respect to increasing positional noise applied to the scene.

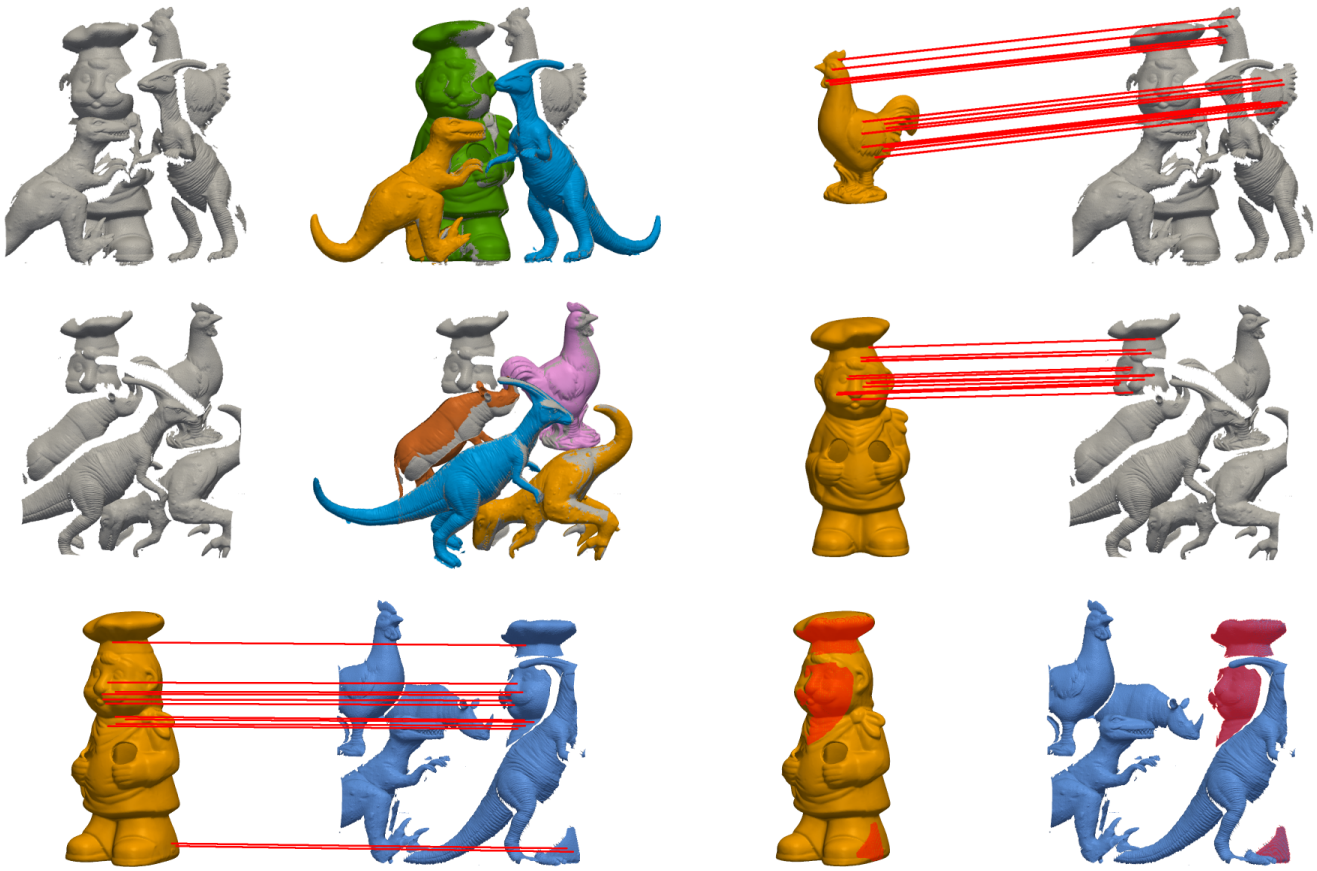


Figure 7. In the first and second rows we show an example of models correctly matched in scenes that break the method by Mian *et al.* (the chicken in the first row has been missed) and the method by Bariya-Nishino (the chef in the second row has been missed). In the third row we show the results obtained by playing a segmentation game (right) starting with the matches produced by a sparse game (left).

D. Performance Considerations

We did not systematically evaluate the performance of the proposed pipeline. In practice, the most demanding step from a computational point of view is the evolutionary process. Since this is an iterative process, it is difficult to give an upper bound for its convergence time. However, the time required for each iteration is roughly proportional to the square of the number of strategies, which in turn means that the overall complexity could reach $O(n^4)$ with respect to the number of mesh points. However, since only the initial subset of strategies is used, the actual complexity is much lower and can be controlled by the parameter k (described in section 2). Empirically we always observed convergence of the process within 50 iterations (tens of seconds).

IV. Conclusions and Future Work

We described and empirically evaluated a novel pipeline for model-based 3D object recognition and segmentation in cluttered range scans. The pipeline starts with the detection of distinctive keypoints in the scene, which in turn is composed of a relevance filter, a subsampling step and the calculation of a descriptor for each sample kept. Such keypoints are then pairwise matched with all the relevant points of the model and a set of candidate pairings is obtained. Finally, two non-cooperative games are played: a rigid-matching game and a dense-growing game. The first one performs the actual recognition step and returns a sparse set of reliable matches. The second game expands these matches to segment all the surface patches in the scene that are compatible with the model. The overall approach combines a simple but effective relevance sampling schema with a recent local surface descriptor and with techniques borrowed from the emerging field of Game-Theoretic inlier selection. An extensive experimental evaluation shows that the proposed method outperforms recently proposed state-of-the art techniques on the same dataset. The contribution of the sampling schema is highlighted by testing the performance of the same pipeline leaving out this step; moreover, different keypoints descriptors are shown to give worse results. Finally, resilience to noise and the ability to obtain dense correspondences are evaluated individually obtaining encouraging results. The running time of the

matching algorithm is in line with the techniques currently found in literature. In the immediate future we are aiming at the extension of the proposed framework to both non-rigid and scale-invariant object recognition. We believe that such an extension could take place by introducing in the payoff function of the selection game a measure taking into account the geodesic path between the pairs of matching candidates, rather than attempting to preserve their Euclidean distance.

ACKNOWLEDGMENTS

We wish to thank Dr. Samuele Salti for contributing code to compute SHOT descriptors, Prof. Ajmal S. Mian and Dr. Prabin Bariya for providing us with the experimental results used to compare our approach with their methods. We acknowledge the financial support of the Future and Emerging Technology (FET) Programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open project SIMBAD grant no. 213250.

REFERENCES

[1] T. S. Newman and A. K. Jain, "A system for 3d cad-based inspection using range images," *Pattern Recognition*, vol. 28, no. 10, pp. 1555 – 1574, 1995.

[2] D. Borrmann, J. Elseberg, K. Lingemann, A. Nüchter, and J. Hertzberg, "Globally consistent 3d mapping with scan matching," *Robot. Auton. Syst.*, vol. 56, pp. 130–142, February 2008.

[3] Y.-K. Ahn, Y.-C. Park, K.-S. Choi, W.-C. Park, H.-M. Seo, and K.-M. Jung, "3d spatial touch system based on time-of-flight camera," *WSEAS Trans. Info. Sci. and App.*, vol. 6, pp. 1433–1442, 2009.

[4] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. of the International Conference on Computer Vision*, 1999, pp. 1150–1157.

[5] ——, "Distinctive image features from scale-invariant keypoints," in *Int. J. Comput. Vis.*, vol. 20, 2003, pp. 91–110.

[6] E. Akagündüz, O. Eskizara, and I. Ulusoy, "Scale-space approach for the comparison of hk and sc curvature descriptions as applied to object recognition," in *Proc. of the 16th IEEE International Conference on Image processing*, ser. ICIP'09, Piscataway, NJ, USA, 2009, pp. 413–416.

[7] A. Zaharescu, E. Boyer, K. Varanasi, and R. P. Horaud, "Surface feature detection and description with applications to mesh matching," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, June 2009.

[8] C. S. Chua and R. Jarvis, "Point signatures: A new representation for 3d object recognition," *Int. J. Comput. Vision*, vol. 25, pp. 63–85, October 1997.

[9] Y. Sun, J. Paik, A. Koschan, and M. A. Abidi, "Point fingerprint: A new 3-d object representation scheme," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 33, pp. 712–717, 2003.

[10] 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, 1999.

[11] H. Chen and B. Bhanu, "3d free-form object recognition in range images using local surface patches," *Pattern Recognition Letters*, vol. 28, pp. 1252–1262, July 2007.

[12] A. Frome, D. Huber, R. Kolluri, T. Bülow, and J. Malik, "Recognizing objects in range data using regional point descriptors," in *ECCV 2004, 8th European Conference on Computer Vision*, 2004, pp. 224–237.

[13] F. Tombari, S. Salti, and L. di Stefano, "Unique signatures of histograms for local surface description," in *ECCV 2010 - 11th European Conference on Computer Vision*, 2010, pp. 356–369.

[14] P. Bariya and K. Nishino, "Scale-hierarchical 3d object recognition in cluttered scenes," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2010*, 2010, pp. 1657–1664.

[15] J. Novatnack and K. Nishino, "Scale-dependent/invariant local 3d shape descriptors for fully automatic registration of multiple sets of range images," in *Proc. of the 10th European Conference on Computer Vision*. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 440–453.

[16] A. S. Mian, M. Bennamoun, and R. Owens, "Three-dimensional model-based object recognition and segmentation in cluttered scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, pp. 1584–1601, October 2006.

[17] ——, "On the repeatability and quality of keypoints for local feature-based 3d object retrieval from cluttered scenes," *Int. J. Comput. Vision*, vol. 89, pp. 348–361, September 2010.

[18] A. Albarelli, S. R. Bulò, A. Torsello, and M. Pelillo, "Matching as a non-cooperative game," in *ICCV 2009: Proc. of the 2009 IEEE International Conference on Computer Vision*. IEEE Computer Society, 2009.

[19] A. Albarelli, E. Rodolà, and A. Torsello, "Robust game-theoretic inlier selection for bundle adjustment," in *International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT2010)*, 2010.

[20] M. Pelillo and A. Torsello, "Payoff-monotonic game dynamics and the maximum clique problem," *Neural Computing*, vol. 18, pp. 1215–1258, May 2006.

[21] A. Albarelli, E. Rodolà, and A. Torsello, "Loosely distinctive features for robust surface alignment," in *ECCV 2010 - 11th European Conference on Computer Vision*, 2010, pp. 519–532.

[22] B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *J. of the Optical Society of America. A*, vol. 4, no. 4, pp. 629–642, Apr 1987.