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

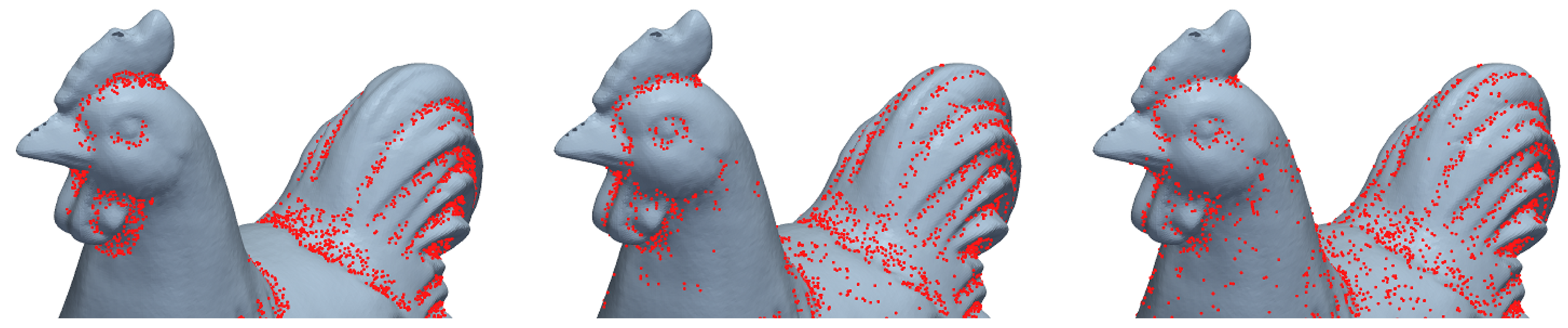
Andrea Albarelli, Emanuele Rodolà, Filippo Bergamasco and Andrea Torsello
Dipartimento di Scienze Ambientali, Informatica e Statistica - Università Ca' Foscari

Introduction

We focus on the recognition of known objects in cluttered and incomplete 3D scans, offering a novel 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. The same technique is used to shift from an initial sparse correspondence to a dense matching, leading to a novel pipeline for 3D object recognition.

Feature Selection

Feature selection is performed on both the model and the scene by computing single-component Integral Hashes at a given support scale σ, and thus retaining only those samples with a negative value (corresponding to concave surface patches). By modulating the value of σ, a more or less selective sample selection can be made.

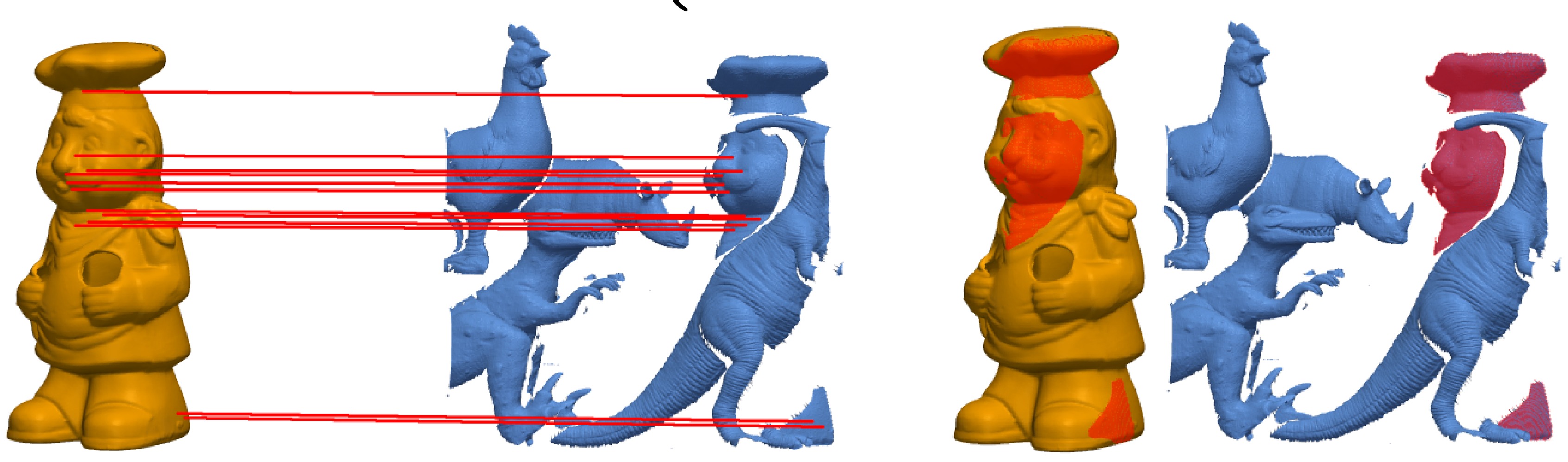


Finally, a descriptor vector is computed for each vertex. The SHOT descriptor was ultimately chosen as it obtains the best performance over the whole pipeline.

Dense Matching

We wish to operate a model-driven segmentation of the scene, and as such we wish to match each vertex in the scene to at least one reasonable vertex in the model (many-to-one). If the initial matches are correct, a rigid motion bringing the model over the scene can be computed, and the model vertex corresponding to each scene point will be in its neighborhood.

The new payoff is: $$\Pi' = \begin{cases} \delta((a_1,b_1),(a_2,b_2))^{\alpha} & \text{if } a_1 \neq a_2 \\ 0 & \text{otherwise} \end{cases}$$



Sparse Matching

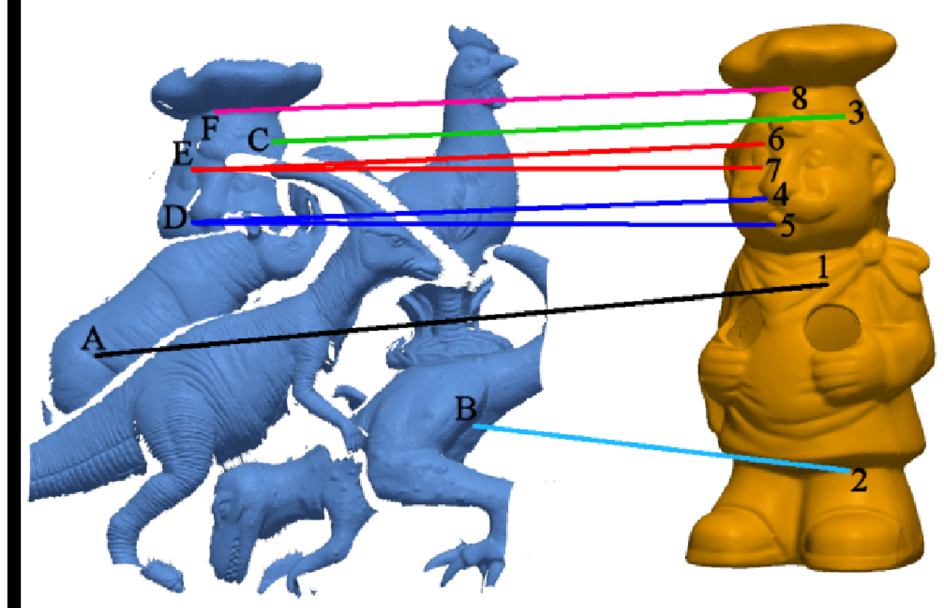
We base our matching framework on the recently introduced Game-Theoretic techniques for inlier selection.
First, a set of candidate matches (or *strategies*) is built by associating each interest point in the scene with the *k* nearest points in the model in terms of the descriptor:

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

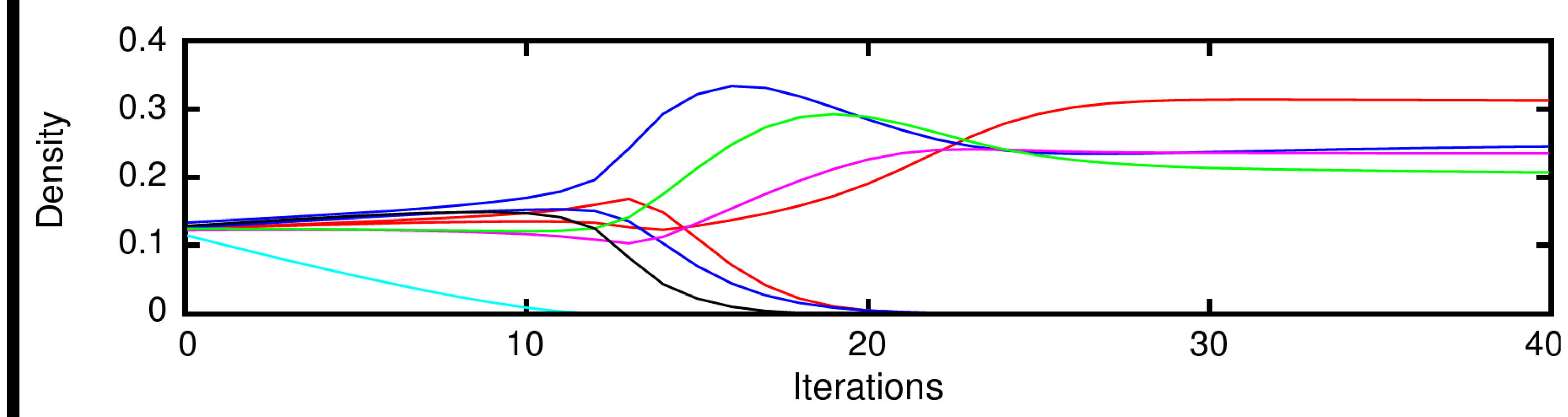
Then we look for a subset of candidates enforcing some rigidity constraint. To this extent, we define a distance measure between pairs of strategies in *S* as:

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

From this we build a symmetric matrix which is given as an input to a natural selection process. When the process comes to convergence, all the non-extinct strategies are retained as correct matches.



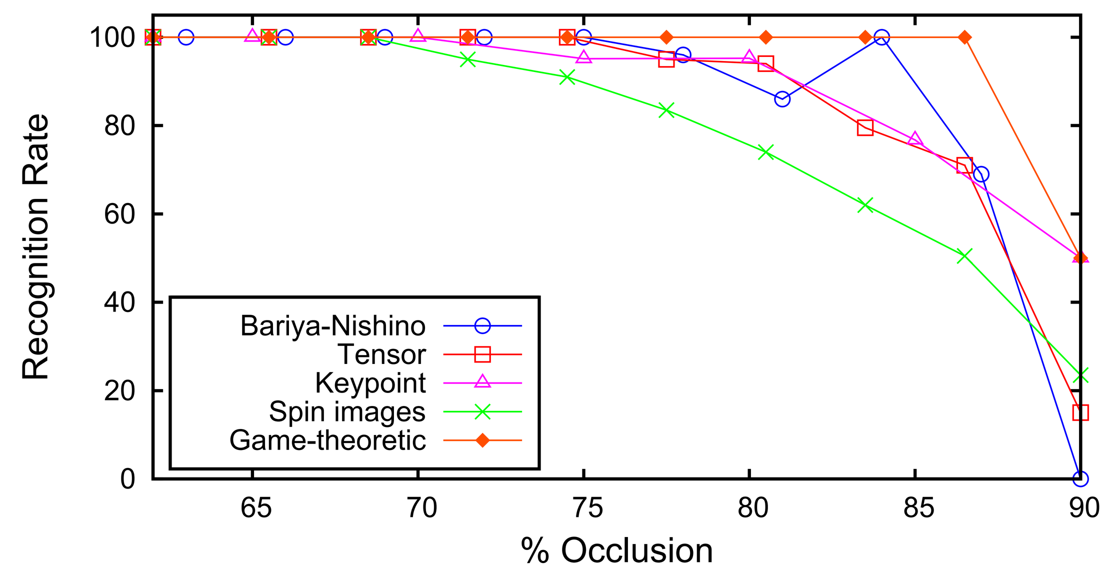
	A1	B2	C3	D4	D5	E6	E7	F8
A1	0	0.12	0.77	0.83	0.98	0.77	0.66	0.75
B2	0.12	0	0.05	0.21	0.37	0.07	0.32	0.17
C3	0.77	0.05	0	0.99	0.6	0.99	0.99	0.7
D4	0.83	0.21	0.99	0	0	0.99	0.99	0.96
D5	0.98	0.37	0.6	0	0	0.9	0.88	0.69
E6	0.77	0.07	0.99	0.99	0.9	0	0	0.98
E7	0.66	0.32	0.99	0.99	0.88	0	0	0.99
F8	0.75	0.17	0.7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

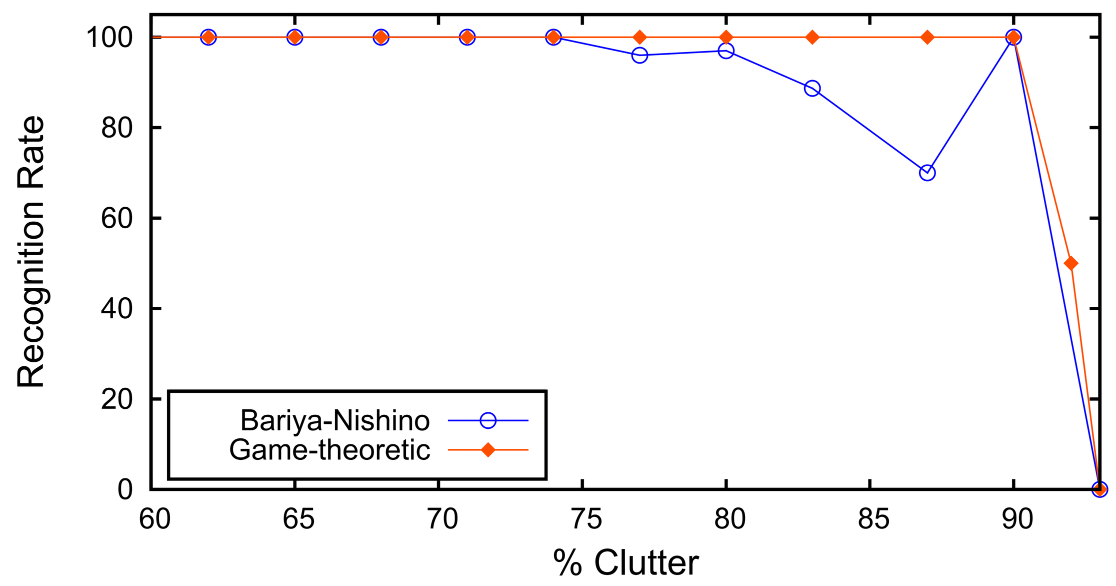


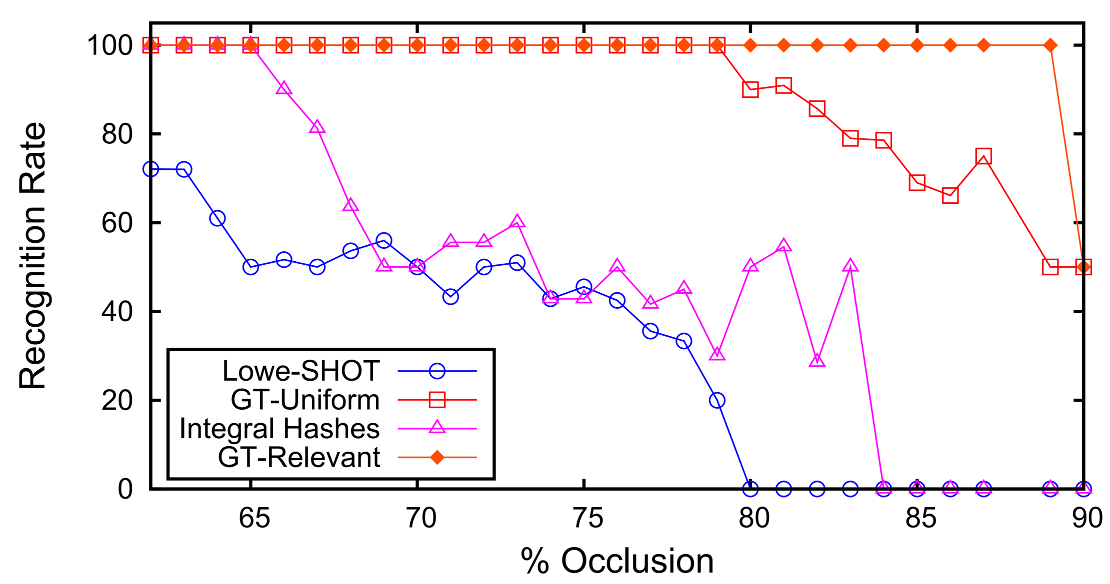
Experimental Results

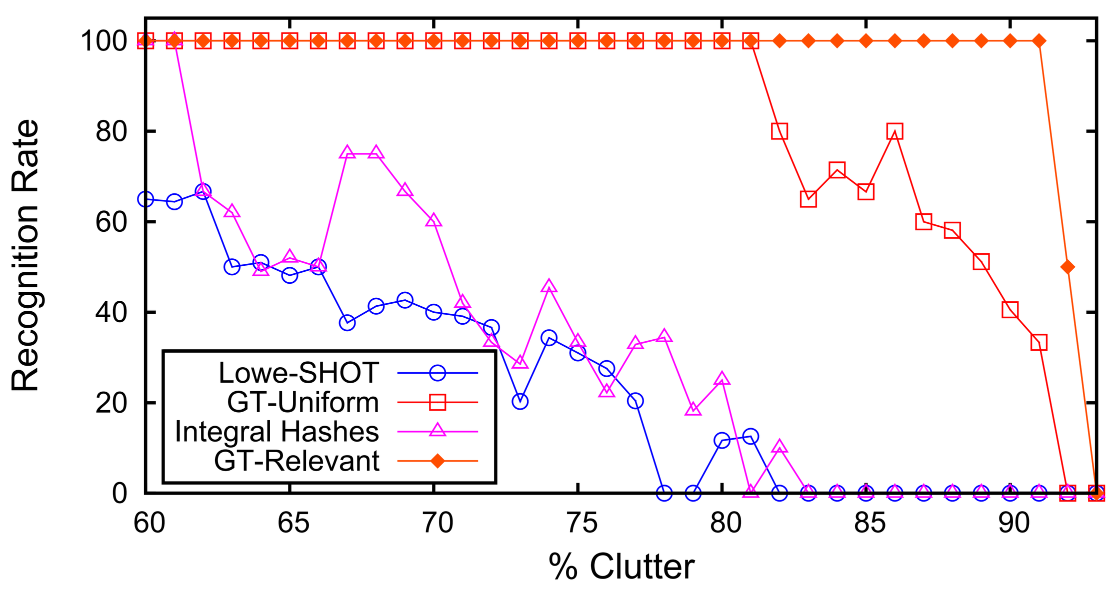
We performed a wide range of tests and comparisons with recent techniques, under different conditions of noise, occlusion and clutter, on the same dataset.

- The minimum number of matches to assume the model as recognized in the scene was 8
- The proposed technique did not report any false matches









An example of "critical" scene where other techniques fail: