An Experimental Benchmark for Point Set Coarse Matching

Ferran Roure[1], Yago Díez [1], Xavier Lladó[1], Josep Forest[1], Tomislav Pribanic[2] and Joaquim Salvi[1]

[1]*ViCOROB Research Group, University of Girona, Girona, Spain*

[2]*Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, Croatia*

{froure, yago, llado, forest, qsalvi}@eia.udg.edu, tomislav.pribanic@fer.hr

Keywords: Coarse Matching, Algorithm Standardization, Datasets, Benchmark, 3D Registration

Abstract: Coarse Matching of point clouds is a fundamental problem in a variety of computer vision applications. While many algorithms have been developed in recent years to address its different aspects, the lack of unified measures and commonly agreed upon data hampers algorithm performances comparison. Additionally, a large number of contributions are tested only with synthetic or processed data. This is a problem as the resulting scenario is somewhat less challenging and does not always conform to practical application conditions. In this paper, we present a new, publicly available database that aims at overcoming the existing problems, provide researchers with a useful tool to compare new contributions to existing ones and represent a step towards standardization. The database contains both processed and unprocessed data with attention to specially challenging datasets. It also includes information on correct solution, presence of noise, overlap percentages and additional information that will allow researchers to focus only on specific parts of the matching pipeline.

1 INTRODUCTION

3D registration[1] is a fundamental problem in a variety of areas such as medical imaging, heritage reconstruction or shape retrieval. Specific applications include the alignment of temporal 3D images for lesion monitoring, the modelling of structures, the reconstruction of an object giving several views or the bin picking problem.

While the ICP algorithm [Besl and McKay, 1992] [Rusinkiewicz and Levoy, 2001] has been adopted as a ”de facto” standard for the problem of ”fine matching” (i.e. determining the best match between two sets once an adequate initial approximation has been found), finding that initial approximation remains an active research field known as ”coarse matching”. The coarse matching problem encompasses different communities and contributions appear steadily in all the steps of the matching pipeline (Fig.1): 1) Detection and Description [Bronstein, 2010] [Salti et al., 2011] [Yu et al., 2013] 2) Searching Strategies [Gelfand et al., 2005] [Aiger et al., 2008] [Albarelli et al., 2010] or 3) Fine matching [Besl and McKay, 1992].

Given the diverse origin of the contribution to this research field as well as the divergence in focus between papers dealing with separate parts of the matching pipeline, most contributions are evaluated with particular datasets that are not accessible to the research community. Within this data, we distinguish two types: synthetic (or processed) and real data. The former usually consists of scanned objects that have been de-noised, smoothed or similarly post-processed. The later consist of scanned objects without any such post-processing. This distinction is crucial when evaluating the performance of algorithms, as data with noise or a low degree of overlap presents a much more challenging problem.

Some publicly available datasets do exist [Bronstein et al., 2008] [Bronstein, 2010] [Bogo et al., 2014] or the Stanford Repository[2] and some of them are widely used by the community. However, not much background information is available for result comparison. Specifically: 1) No ”correct output” sets are provided. 2) Similarly, no assessment on the level of noise or on the final overlap to be achieved between sets is given 3) No intermediate data concerning the different steps of the matching pipeline is included. For example, researchers developing new search strategies (at the later part of the matching pipeline) need to either first implement state-of-the-art descriptors or not use them at all.

[1]Note that we understand the words ”registration”, ”matching” and ”alignment” as synonyms, and we use them interchangeably throughout the paper.

[2]http://graphics.stanford.edu/data/3Dscanrep/

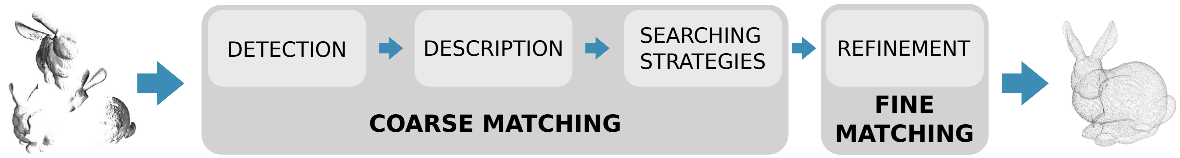

Figure 1: Point Registration Pipeline

In this work we present a new, publicly available, database that aims at overcoming these limitations and provide a valuable tools for researchers working in the coarse matching field. The main characteristics of our database are:

- It contains datasets targeting different aspects of the matching problems. Special attention is given to noise and overlap.
- Proposed solutions are included as well as measures on the quality of final registration (overlap percentage and residue).
- The data it contains makes it possible to test different parts of the pipeline separately. For example, descriptor data is provided as well as output after ICP execution.
- Several practical applications and problems are targeted, so, for example, we include, as well as the usual data were rigid motion needs to be determined, sets where the rotation is provided separately. This makes it possible to test methods that determine the two parts of the motion separately [Larkins et al., 2012].
- Data from real application problems allow to test algorithms in increasingly challenging scenarios.
- It can be accessed online at: http://eia.udg.edu/3dbenchmark

2 Overview on the State Of the Art of 3D registration

New acquisition techniques provide higher-resolution scanned objects. This results in point clouds that represent object surfaces more precisely but use also larger point cloud that demand more efficient algorithms. Many contributions exists in the literature targeting different parts of the matching pipeline (Figure 1). In terms of published papers, Detectors and Descriptors are the most active field. Some examples of the most used methods are Spin Image [Johnson, 1997], SHOT [Tombari et al., 2010], Heat Kernel Signature (HKS) [Sun et al., 2009], Intrinsic Shape Signatures (ISS) [Zhong, 2009], Fast Point Feature Histogram (FPFH) [Rusu et al., 2009] or Integral Invariants [Manay et al., 2004].

Most methods in the literature are tested with in-house data, making it difficult to compare their performances to that of new contributions. Moreover, comparison to other approaches in the state of the art is often not provided. Noticeable exceptions to this last point are [Bronstein, 2010], [Salti et al., 2011] or [Dutagaci et al., 2012], which provide meaningful comparison among several methods. Furthermore, in most papers the data used has been processed to filter noise and outliers. While this makes it possible to obtain better results and enhance the range of applicability of algorithms, such post-processes are often not available in real application situations. For example, the Heat Kernel HKS descriptor [Sun et al., 2009] shows high repeatability and distinctiveness with processed models [Bronstein, 2010], yet with laser-scan data or image-based reconstructions, it is deemed too selective to be considered a robust registration algorithm [Kim and Hilton, 2013]. Other methods that have been extensively tested with real application data are the Intrinsic Shape Signatures (ISS) [Zhong, 2009] and Key-Point Quality (KPQ) [Mian et al., 2010]. While their performances with real data are, once more, worse than with processed data, they still manage to obtain good results even with real data (70% detector repeatability, for example).

Perhaps the most significant contributions in terms of computational gain made recently correspond to searching strategies [Aiger et al., 2008] [Mellado et al., 2014] although the main limitation here is the lack of comparison between different approaches. Finally, although research in fine matching is still going on, the Iterative Closest Point (ICP) algorithm [Besl and McKay, 1992] has become a "de-facto" standard for the problem and obtains acceptable results in a variety of situations. In the following sections, we present a new database aimed at helping future researchers overcome these limitations of the current state of the art.

3 Data Base Description

The database contains several models that cover a variety of scenarios in 3D coarse registration. From low to high complexity, we provide two processed models and two "real application" models. The processed models are modified versions of the well known Buddha and Bunny models from the Stanford Repository[3]. We include five of the original views for every model as well as data concerning correct alignment results between them. Additionally, we present modified versions of the Bunny dataset with varying levels of gaussian noise added. Concerning the real datasets, the first one corresponds to five views of a bust model reconstructed using a structured-light system [Pribanić et al., 2010]. This datasets is more challenging due to noise in acquisition and includes the possibility of determining the rotation and the translation part of the motion separately. Finally, the Joints dataset present a particular and challenging problem of interest in industrial settings. The data presented corresponds to an unstructured heap of manufactured parts along with a model part that has to be located (possibly many times over) in the heap. In the remainder of this section we describe the general layout of the database and all the characteristics shared by all datasets. In the following subsections we provide specific details on each particular datasets.

All models are provided in *.ply format. Moreover, we added the normal vectors computed at every point. We provide this information in order to offer a common starting point for algorithms that use normals as their main geometric primitive. Such algorithms include descriptor algorithms such as [Johnson, 1997] [Tombari et al., 2010] but also algorithms that use Fourier analysis of normal distribution to determine matching [Larkins et al., 2012]. Additionally, all models include:

- Five consecutive views in .*ply format with different overlapping ratios. For each view, we provide the non-aligned view as well as its properly alignment pose. All computations were performed automatically and re-checked manually.
- 4x4 transformation matrices in homogeneous coordinates to align all views.
- Alignment residue computed using the Root Mean Squared Distance (RMSD) criterion.
- Overlap ration for correctly aligned view. This is computed as the percentage of paired points after coarse matching and ICP were successfully run. A point was considered matched if its nearest neighbour in the other set was closer than $2 \times MMD$ where MMD stands for the mean nearest neighbour distance for the set.
- All this information is publicly accessible online at *http://eia.udg.edu/3dbenchmark*.

[3]http://graphics.stanford.edu/data/3Dscanrep/

3.1 Processed data

Two of the most well-known objects in the literature are the Bunny and Buddha models the from Stanford Repository. Both original datasets consist of several views with smoothed surfaces. These objects also appear to have undergone noise and outlier filtering. This type of data presents less challenging problems that other datasets.

The Bunny model is the simplest model with $\approx$ 37000 points per view (Figure 2). All features are clearly defined, without noise, outliers or symmetries. Pairwise views present decreasing overlap increasing the difficulty of registration. Specifically, views *bun0-bun1* have $\approx$ 90% of overlap while views *bun3-bun4* have only $\approx$ 40% (see Table 1 for details).

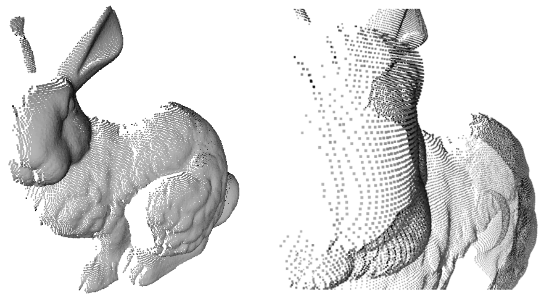

Figure 2: Left: *bun1* view. Right: Detail of *bun1* view. Notice that there are no noise or outliers.

For this model we also include a four-level noisy versions of three different views for more challenging tests (Figure 4). Gaussian noise was added to all the views with varying modulus. This modulus of the noise vector associated to each point was randomly chosen but had a limit that changed for each set. Specifically, in the first view this maximum was set to $1 \times MMD$. Values of $2 \times MMD$, $3 \times MMD$ and $4 \times MMD$ were also considered in order to make up the three remaining modified views.

The Buddha model is a bit more challenging than the Bunny because of its bigger size ($\approx$ 75000 points per view). Additionally, it has much smaller details and a higher degree of symmetry. These represent challenges both for Descriptor functions as well as for Searching Strategies. Furthermore, the base of the figure is a rounded pedestal which induces quite a noticeable source of symmetry and hampers normal space analysis.

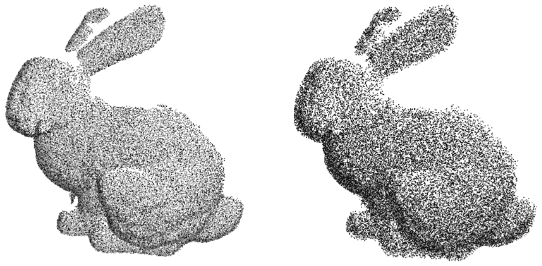

Figure 3: Left: *bun0* view with $1 \times MMD$ of gaussian noise. Right: *bun0* view with $4 \times MMD$ of gausian noise.

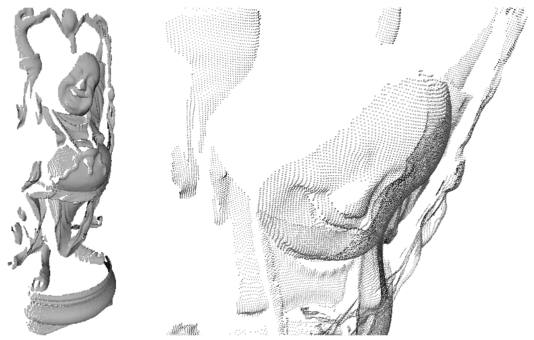

Figure 4: Left: *buddha0* view of Buddha model. Right: Detail of *buddha0* view.

Table 1: Example of data base results from processed data with a $MMD \approx 5 \times 10^{-4}$ for Bunny yhe model and $MMD \approx 3 \times 10^{-4}3$ for the Buddha model.

	Views	Residue	Ovlp A-B	Ovlp B-A
Bunny	0 - 1	3×10^{-4}	91.66%	89.91%
	1 - 2	3×10^{-4}	48.67%	79.07%
	2 - 3	5×10^{-4}	44.45%	27.53%
	3 - 4	4×10^{-4}	38.45%	48.30%
Buddha	0 - 1	2×10^{-4}	79.74%	85.16%
	1 - 2	2×10^{-4}	80.04%	89.28%
	2 - 3	2×10^{-4}	76.99%	90.84%
	3 - 4	2×10^{-4}	75.64%	79.10%

3.2 Real data

Our real scanned data consist of a set of views from two different models that have been acquired using different techniques. The Bust model is a real-sized mannequin of a human body, scanned with a 3D structured-light system [Pribanić et al., 2010, Pribanic et al., 2013]. The Joints model consist on an unsorted lot of metal joints acquired using a range scan (a laser and a single camera). This particular model was conceived for solving the bin picking problem, where a robot is expected to identify a certain part from a stack of unsorted similar parts. Figures 5 and 6 present example views of these datasets.

Both datasets represent a more challenging scenario for any registration method, due to the presence of noise, the low overlapping ratios and the outliers.

The Bust model was acquired using a structured light system that consist on projecting a pattern on a real object and tacking a photo of the scene. Then, the differences between the original pattern and the captured one provide the 3D information. The views of this model contain $\approx$ 450000 points. There was no post-processing step and the noise comes from acquisition system (Figure 5). The overlapping ratio is $\approx$ 60%, depending on the view. All the characteristics of this model (noise, overlap and number of points) represents a challenging problem for registration algorithms.

Figure 5: Left: *bust0* view of Bust model. Right: Detail of *bust0* view.

For this particular case we also provide a single translation version of data, where an approximation of the rotation is already computed using a gyroscope added to the scanning system [Pribanic et al., 2013]. This approximation reduces the final computation costs, because only a single correspondence needs to be found. As the estimation of the rotation is noisy due to the nature of the sensor used, this datasets aim at being useful for researchers who tackle the matching problem by determining rotations and translations separately [Larkins et al., 2012]. The data provided allows to compare rotation determination algorithms to the data obtained by the sensor and also to prove the usefulness of robust translation determination algorithm with noisy rotation data.

The most complex registration problem in the database is the Joints model (Figure 6). This scenario is intended for industrial applications such as quality control. Specifically, it is an instance of the "bin picking" problem, where a robot arm is expected to pick an industrial part from an unstructured heap of possibly defective similar parts. This data was obtained using a range scan composed by a laser and a single camera and presents abundant noise and outliers. We provide this model without any post-processing in order to offer a registration problem as close as possible to real application conditions. The dataset consists of a big heap of equal metal pieces, unsorted inside a box. As well as an "ideal" model of the piece to be found within the heap. Although, many possible matches are possible the noise and the outliers make it a very difficult task. Providing a scoring function that ranks all possible matching candidates is also an interested associate problem. Table 4 presents a summary of the alignment data obtained for the real datasets.

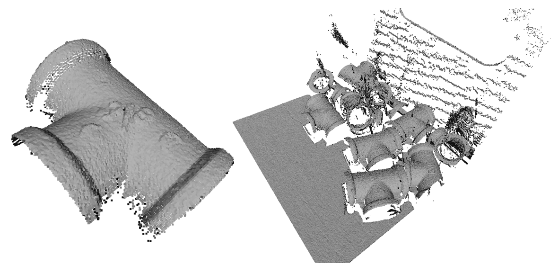

Figure 6: Left: Single *joint1* view. Right: Heap of unsorted joints.

Table 2: Example of data base results from real data with a $MMD \approx 0.59$ for Bust model and $MMD \approx 0.16$ for Joints model.

	Views	Residue	Ovlp A-B	Ovlp B-A
Bust	0 - 1	0.68	86.34%	53.94%
	1 - 2	0.66	72.28%	57.31%
	2 - 3	0.65	61.16%	60.06%
	3 - 4	0.66	61.69%	76.50%
Joints	heap - 1	0.14	4.27%	86.92%
	heap - 2	0.19	3.73%	73.60%

4 Examples of application

The main aim of this database is to be used to compare new matching algorithms, but it can also be used to study aspect of the coarse matching problem. In this section we illustrate how the presented database can be used to study commonly agreed upon "truths" of coarse matching algorithms.

4.1 Application 1: ICP needs a "good enough initial pose" to succeed.

The very distinction between coarse and fine matching algorithms is based on the accepted fact that ICP fails to converge or stalls at a local minimum if the initial pose it is provided with is "not good enough". It is, however, infrequent to see quantifications of what a local minimum looks like or exactly how good the initial pose needs to be.

In this experiment we run ICP with the Bunny and Bust datasets. For each of the two objects, one view was registered against its consecutive view. The original pose stands for the best possible initial alignment. We then perturbed this initial pose by rotating the second view along one of the three axes. We repeated the process independently for axes X,Y and Z and also by rotating along all the axes at the same time. Table 3 contains the summary of this experiment. Column *Deg. fail* shows the angle (in degrees) were ICP failed to converge to the global minimum for the first time. The table also shows how in some executions, ICP fell in local minimum (indicated in the table by an asterisk) while in other executions ICP was not able to converge. Notice that, once again, we were able to observe differences in behavior for processed and real data. Specifically, the Bunny dataset was much more robust to the perturbation of the initial pose than the Bust dataset. We also observe, how the resilience against this type of rotational perturbation increases when the total overlap between views is higher. As showed in Table 1, the overlap between *bun0* and *bun1* is higher ($\approx 91\%$) than *bun1* and *bun2* ($\approx 48\%$).

Table 3: Table of ICP test. An asterisk indicates that the algorithm stalled at a local minimum. The dash indicates that the ICP was not able to find any solution.

	Views	Rot. Axis	Deg. fail	Ovlp B-A
Bunny	0 - 1	X	50°	14.24%*
		Y	70°	12.78%*
		Z	40°	17.45%*
		XYZ	25°	9.98%*
	1 - 2	X	40°	-
		Y	60°	-
		Z	30°	4.21%*
		XYZ	20°	4.12%*
Buddha	0 - 1	X	20°	-
		Y	10°	-
		Z	10°	-
		XYZ	2°	-
	1 - 2	X	15°	-
		Y	15°	-
		Z	15°	-
		XYZ	3°	-

4.2 Application 2: Descriptor performance drops in the presence of noise

It is often claimed that noise in data negatively affects the behaviour of shape descriptors. In this experiment we aimed at determining how much noise in data it takes to get well-established descriptor to fail. Specifically, we chose the well-established SHOT method [Tombari et al., 2010]. We then run a search based on coupling points according to their descriptor value:

- A three point basis in the first set was chosen.
- Each point in the basis was tentatively matched to k neighbours in order of decreasing descriptor similarity.
- Once three correspondences were determined, distances between the points in the two basis were checked for consistency.
- If the two basis presented similar distances, then a rigid motion between the two sets was computed.
- ICP was used to complete the matching process. The percentage of matched points (also referred to as overlap percentage) and the residue between the two sets was computed.

In order to test the effect of noise, we use the sets with increasing quantity of noise described in section 3.1. Table 4 presents the results obtained.

Table 4: Results of registration process with SHOT descriptor without ICP refinement. Timeout was set at 15 hours. The overlap presented is the best obtained (at the end of execution or at timeout). An asterisk indicates that the overlap obtained was not the best possible and, thus, the algorithm stalled at a local minimum.

Noise	Residue	Ovlp A-B	k
-	7×10^{-4}	97.32%	500
$1\times MMD$	5×10^{-4}	94.58%	500
$2\times MMD$	1×10^{-3}	36.72%*	500
$3\times MMD$	1×10^{-3}	21.59%*	500
$4\times MMD$	1×10^{-3}	22.10%*	500

The results show how in the absence of noise the descriptors-based search is able to find correspondences very quickly while achieving the total degree of overlap. Descriptors manage to discriminate points very well and we only need to consider a low number of possible correspondences k in order to obtain the best possible matching. As soon as noise is added to the data the behaviour of the search suffers. For the less noisy set, the algorithm still manages to find the correct matching but needs to consider many more correspondences. For the remaining sets, containing more noise, the algorithm was allowed to run for 15 hours before being stopped. During all that time even when considering a very high number of correspondences, only local minimum were reached and the algorithm was unable to output the correct alignment for any of the three sets.

5 Conclusions and Future Work

In this paper we have introduced a new database aimed at providing researchers in the coarse matching research field with a usable tool that overcomes some of the current limitations in the field while providing insight in a variety of aspect of the problem. Some of the aspects that the database focuses on are: Providing correct registration results for the publicly accessible data, with special attention to overlap percentages between sets and amount of noise present in data. Including intermediate data such as surface normals, descriptor values or separate values for rotations and translations (coming from realistic hardware sources). Finally, the fact that part of the data comes from realistic applications such as surface reconstruction (Bust model) or industrial applications (Joints model) aims at providing a benchmark for researchers to show the potential of new contributions to the field in specially challenging scenarios.

Regarding future work, we expect to increase the number of models in the database as well as include outputs from existing and future state of the art algorithms.

REFERENCES

Aiger, D., Mitra, N. J., and Cohen-Or, D. (2008). 4-points congruent sets for robust pairwise surface registration. In *ACM Transactions on Graphics*, volume 27, page 85.

Albarelli, A., Rodola, E., and Torsello, A. (2010). A game-theoretic approach to fine surface registration without initial motion estimation. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 430–437. IEEE.

Besl, P. J. and McKay, N. D. (1992). A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256.

Bogo, F., Romero, J., Loper, M., and Black, M. J. (2014). FAUST: Dataset and evaluation for 3D mesh registration. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Piscataway, NJ, USA. IEEE.

Bronstein, A. e. (2010). Shrec 2010: robust feature detection and description benchmark. *Eurographics Workshop on 3D Object Retrieval*, 2(5):6.

Bronstein, A. M., Bronstein, M. M., and Kimmel, R. (2008). *Numerical geometry of non-rigid shapes*. Springer.

Dutagaci, H., Cheung, C. P., and Godil, A. (2012). Evaluation of 3d interest point detection techniques via human-generated ground truth. *The Visual Computer*, 28(9):901–917.

Gelfand, N., Mitra, N. J., Guibas, L. J., and Pottmann, H. (2005). Robust global registration. In *Eurographics Symposium on Geometry Processing*, pages 197–206.

Johnson, A. E. (1997). *Spin-images: A representation for 3-D surface matching*. PhD thesis, Citeseer.

Kim, H. and Hilton, A. (2013). Evaluation of 3d feature descriptors for multi-modal data registration.

Larkins, R. L., Cree, M. J., and Dorrington, A. A. (2012). Verification of multi-view point-cloud registration for spherical harmonic cross-correlation. In *Proceedings of the 27th Conference on Image and Vision Computing New Zealand*, pages 358–363. ACM.

Manay, S., Hong, B.-W., Yezzi, A., and Soatto, S. (2004). Integral invariant signatures. *European Conf. on Computer Vision*, pages 87–99.

Mellado, N., Aiger, D., and Mitra, N. J. (2014). Super 4pcs fast global pointcloud registration via smart indexing. In *Computer Graphics Forum*, volume 33, pages 205–215. Wiley Online Library.

Mian, A., Bennamoun, M., and Owens, R. (2010). On the repeatability and quality of keypoints for local feature-based 3d object retrieval from cluttered scenes. *International Journal of Computer Vision*, 89(2):348–361.

Pribanic, T., Diez, Y., Fernandez, S., and Salvi, J. (2013). An efficient method for surface registration. In *VISAPP (1)*, pages 500–503.

Pribanić, T., Mrvoš, S., and Salvi, J. (2010). Efficient multiple phase shift patterns for dense 3d acquisition in structured light scanning. *Image and Vision Computing*, 28(8):1255–1266.

Rusinkiewicz, S. and Levoy, M. (2001). Efficient variants of the icp algorithm. In *IEEE International Conference on 3D Digital Imaging and Modeling*, pages 145–152.

Rusu, R. B., Blodow, N., and Beetz, M. (2009). Fast point feature histograms (fpfh) for 3d registration. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, pages 3212–3217.

Salti, S., Tombari, F., and Stefano, L. D. (2011). A performance evaluation of 3d keypoint detectors. In *IEEE International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, pages 236–243.

Sun, J., Ovsjanikov, M., and Guibas, L. (2009). A concise and provably informative multi-scale signature based on heat diffusion. In *Computer Graphics Forum*, volume 28, pages 1383–1392.

Tombari, F., Salti, S., and Di Stefano, L. (2010). Unique signatures of histograms for local surface description. *European Conf. on Computer Vision*, pages 356–369.

Yu, T.-H., Woodford, O. J., and Cipolla, R. (2013). A performance evaluation of volumetric 3d interest point detectors. *International Journal of Computer Vision*, pages 1–18.

Zhong, Y. (2009). Intrinsic shape signatures: A shape descriptor for 3d object recognition. In *IEEE International Conference on Computer Vision Workshops*, pages 689–696.

ACKNOWLEDGEMENTS

This work has been supported by the FP7-ICT-2011-7 project PANDORA-Persistent Autonomy through Learning, Adaptation, Observation and Replanning (Ref 288273) funded by the European Commission and the project RAIMON-Autonomous Underwater Robot for Marine Fish Farms Inspection and Monitoring (Ref CTM2011-29691-C02-02) funded by the Ministry of Economy and Competitiveness of the Spanish Government. Ferran Roure is supported by a FPI scholarship associated to the RAIMON project.