3D Registration of Multi-modal Data Using Surface Fitting

Amine Mahiddine, Djamal Merad, Pierre Drap, Jean-marc Boï
LSIS UMR CNRS 7296
Centre National de la Recherche Scientifique
Marseille, France
surname@univ-amu.fr

ABSTRACT

The registration of two 3D point clouds is an essential step in many applications. The objective of our work is to estimate the best geometric transformation to merge two point clouds obtained from different sensors. In this paper, we present a new approach for feature extraction which is distinguished by the nature of the extracted signature of each point. The descriptor we propose is invariant to rotation and overcomes the problem of multi-resolution. To validate our approach, we have tested on synthetic data and we have applied to heterogeneous real data.

Keywords

Registration 3D/3D, Feature extraction, Surface fitting, Underwater, SONAR, Photogrammetry.

1 INTRODUCTION

Our study is within the framework of French National project which aims at developing innovative tools dedicated to underwater survey.

Dealing with the growing interest in the study of the sea, the understanding of 3D models of underwater scenes is becoming more essential. In this context, we see the emergence of photogrammetry systems more efficient. Indeed, photogrammetry has long been used in underwater environment [Bas69a] and its great advantage in comparison with other techniques is its simplicity of implementation, the fact that it helps researchers to make measurements without physical contact with the scene and finally the diversity of potential outcomes (3D measurements on object, 3D reconstruction, orthophoto, restitution vector). However, photogrammetry is based on the use of optical sensors, the constraints of the underwater environment such as turbidity and the presence of suspended particles, require to work at different scales. The limits of an optical sensor, lead us to the use of new instruments that are not exposed to the same constraints such as the sonar. High resolution sonar systems are an effective way to measure objects on the seabed. Nevertheless the magnitude is however not the same as the optical sensors, the best acoustic active sensors have a range of about 10 meters maximum for accuracy of one centimeter. Therefore, the resolution of the point cloud obtained is rather low in comparison with that obtained by the optical sensor, because the measurement is made at a distance from the object measured with an angular pitch of 1 degr. In our context, the registration of the point cloud from photogrammetry will increase the resolution of the measurements from the sonar and helps to get the textures and color of objects for visualization.

The fusion of optical and acoustic data is a very promising technique for underwater objects survey, which has received increasing attention in recent years. The first attempts at fusion of optical and acoustic data are presented in the work of Andrea Fusiello [Fus00a] where acoustic data were used to approximate the external orientation of the camera in the particular case of the measurement of jacket for offshore. Research work have also integrated acoustic and optical data in an augmented reality environment, such as in [Fus04a][Piz09a][Sing00a]. The authors describe the use of techniques that allow the superposition of photo mosaics on 3D bathymetric digital maps of terrain. Negahdaripour *et al.* [Neg07] propose a new technique for calibrating a stereo opti-acoustic system. This technique is used to estimate the best relative orientation between the two sensors in order to model the epipolar geometry of the system for the 3D reconstruction of the scene. Research are consistent with the integration of different sensors. Finally, the research works of Hurtos and Cufi [Hur10a] are consistent with the integration of different sensors, as is already the case for land survey where laser and camera are often integrated into a single tool.

More broadly, it can be found out that few studies have developed innovative techniques for fusion of 3D data obtained by utilizing optical and acoustic sensors in an underwater environment.

In the next section, we will present a state of the art methods of 3D/3D registration and we will discuss the choice of the descriptor. Section 3 is devoted to the description of our method, including the presentation of our 3D descriptor. Thereafter, we propose to evaluate our algorithm on pairs of synthetic data. These results will be discussed in section 4, and highlight the application of our method on heterogeneous data. Finally, section 5 contains the concluding remarks.

2 RELATED WORK

The goal of the 3D registration is to find the best rigid transformation between two point clouds. To find the right transformation, we need a set of corresponding points, and the fact that they are unknown a priori, makes the challenge difficult and interesting at the same time.

Many studies concerning the registration of 3D point clouds have been made over the last thirty years. However, the registration of two point sets of different resolutions remains an open problem with the issue of the invariance of the signatures is a major point. There are various ways to address the problem, we can classify these techniques into two categories: iterative methods and locales descriptors based techniques.

The iterative methods are often variants of the ICP (Iterative Closest Point) method proposed by Besl and McKay [Bes92a] remain the most used in the majority of tools for automatic registration. One drawback to this method is the fact that it converges to the first local minimum that is often due to initialization with false matches. Several solutions have been implemented to solve this kind of problem, such the one proposed by Chen and Medioni [Med91a], which replaces the measurement of distance between points by measuring the distance between a point and a tangential plane which makes the algorithm less sensitive to local minima. Rusinkiewicz and Levoy [Lev01a] compare several variants of the standard algorithm in terms of convergence time. They also propose an optimized method where they obtained good results using a method to classify points in the direction of their normal, then sampling each class and rejecting the outliers. Variants of the ICP algorithm [Lev01a] [Tur94a] [Mas96a] [Wei97a] are still trying to improve the initialization of matched points step to calculate an approximation of rigid transformation which can minimize the quadratic error between the matched points.

The second category contains all methods that aim to extract meaningful information (descriptor) of one point thanks to its neighborhood. Jarvis and Chua [Chu97a] present a signature for each point by calculating the distances between the latter and all neighboring points. Rusu *et al.* [Rad08a] propose the PFH (Point Feature Histograms) method that characterizes the local geometry of a 3D point and stores the information in 16-bin histograms by using the method that Wahl and Hirzinger [Wah03a] proposed for a pair-relation histogram. And in [Rus09a], the authors have optimized their method for real time use. Johnson and Hebert [Joh99a] introduced the concept of spin images where each spin image is a local descriptor of a surface in a point P defined by its position and its normal n. This method computes a 2D histogram which is the projection of all points on the cylindrical coordinates knowing that this method requires a point cloud with uniform resolution. Within the same category, 3D shape context [Fro04a] accumulates 3D histograms of points within a sphere centered on the basis point P and its north pole oriented with the surface normal. In the same context, Tombari *et al.* [Tom10a] present SHOT (Signature of Histograms of OrienTations), which concatenates a set of local histograms incorporating geometric information about the position of neighboring points within a sphere divided into several sectors. For each sector of the sphere, a one-dimensional histogram is calculated by accumulating the angles between the normal to the point of interest with normal neighboring points. Mian *et al.* [Mia10a] propose an interesting method for finding keypoints on 3D models and they also used 3D surface fitting techniques to detect the best matching. However, this method requires a learning phase on a base of 3D models. On the other hand, as they apply a principal component analysis (PCA) to represent all objects in the same referential, this induces an ambiguity of 180 degrees on each principal axis. Which makes the phase of matching very difficult. Since the PCA is also sensitive to the dispersion of points, this would have been more appropriate to apply one of the methods cited by Petrelli and Di Stefano [Pet11a] to calculate a local reference frame.

In Table 1, we present a comparative of the most commonly used descriptors in terms of use for matching between whole objects (Global), partial objects or data with different resolutions (heterogeneous data).

Methods	Global	Partial	Heterogeneous data
Fast PFH	Yes	Yes	No
SHOT	Yes	Yes	No
Spin image	Yes	Yes	No
Shape index	Yes	No	No
Mian *et al.*	Yes	Yes	No

Table 1: Comparison of feature extraction methods

3 FEATURE EXTRACTION AND MATCHING

The purpose of this study is to find a descriptor for each point, which will be invariant to a Euclidean transformation (rotation and translation) for application to real data. Recall that in this study we used two different data sources. The first one is an optical sensor that produces photographs from which we get a 3D model with high resolution. The second data source is a high-frequency SONAR that provides a 3D point cloud with a lower resolution but with an extended range.

Since the resolutions of the two point clouds are different and one point from the sonar will never match with a point from the optical sensor, we are oriented towards a method of resampling surfaces. Our method is inspired by the work of Mian *et al.* [Mia10a], who used an algorithm for fitting surface presented by D'Erico [Eri08a].

3.1 Overall description of the method

The approach that we developed is based on the use of a 3D descriptor invariant to rotation and on the 3D surface fitting method. This is a novelty as far as we know, this formalism has not been used in the context of multimodal registration.

The principle of this method is illustrated in Figure 1. We can establish three major phases necessary for the registration. The first phase is the extraction of keypoints. We used the method LSP (Local Surface Patches) [Che07a] that calculates the saliency of a point according to its Shape Index. The second phase consists in establish a good quality measure between two sets of points. In our application, this task is difficult because this measure should assess a degree of similarity between two surfaces of different orientations and resolutions. The procedure for extraction of the descriptor will be detailed in the section 3.2. The last phase is devoted to the computation of the rigid transformation between the set of matched points. This transformation is computed by using singular value decomposition (SVD) to decompose the covariance matrix H as follows:

$$H = \sum_{i=1}^{N} (p_i - centroid_p) \times (p'_i - centroid_{p'}) \quad (1)$$

$$[U,S,V] = SVD(H) \quad (2)$$

$$R = V \times U^T \quad (3)$$

$$t = -R \times centroid_p + centroid_{p'} \quad (4)$$

where N is the number of matched points.

3.2 3D Feature Descriptor

We were inspired by existing methods in the literature, that try to extract a descriptor for each point relative to its neighborhood. The neighborhood of a point is determined by a sphere centered at that point (see Figure 2). This sort of descriptor is a local description of the geometry.

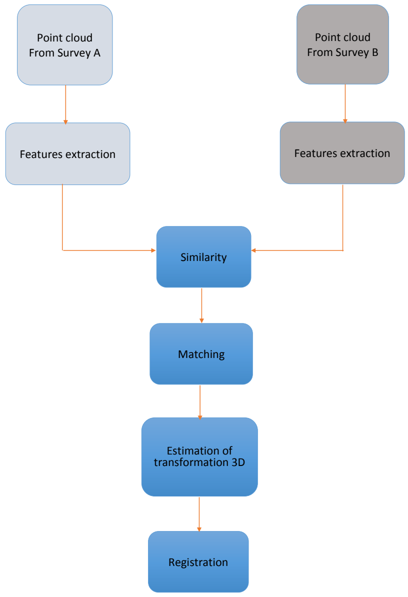


Figure 1: Global illustration of our approach.

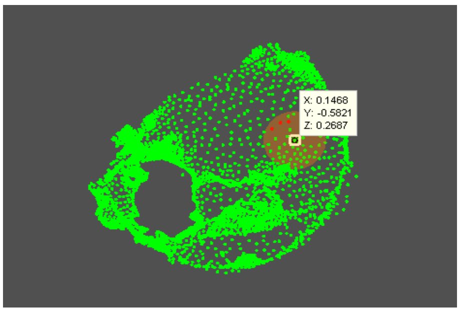


Figure 2: Local surface extraction using a sphere.

The Figure 3 helps to explain our implementation. We start first by extracting some keypoints from each point cloud. As mentioned above, we used the LSP method for the extraction of keypoints. This method computes the saliency of a point according to its Shape Index, which is an indication of the shape of a surface at a point. This measure is defined by (5) where κ_1 and κ_2 are the maximum and minimum principal curvatures respectively.

$$S_i = \frac{1}{2} - \frac{1}{\pi} \tan^{-1} \frac{\kappa_1(p) + \kappa_2(p)}{\kappa_1(p) - \kappa_2(p)} \quad (5)$$

The point is considered as keypoint, if its shape index S_i satisfies the following condition:

$$S_i = max\ of\ shape\ indices\ and\ S_i \geq (1+\alpha)\times\mu \quad (6)$$
$$S_i = min\ of\ shape\ indices\ and\ S_i \leq (1-\beta)\times\mu \quad (7)$$
$$\mu = \frac{1}{N}\sum_{j=1}^{N}(S_j) \qquad 0 \leq \beta \leq 1 \quad (8)$$

where α and β are two scalar parameters that serve to control the keypoints selection and N is the number of neighbors within a sphere.

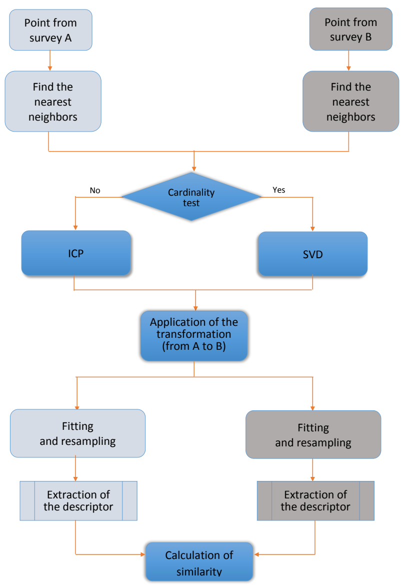


Figure 3: Feature extraction process.

The descriptor that we introduce in this section is based on surface fitting method that was implemented by D'Erico [Eri08a]. This method is not invariant to rotation, where for the same set of points oriented differently, we can have different surfaces. For this reason, before extraction of the features from keypoints, we first try to align the sets of neighbors of two keypoints to solve the problem of rotation of the surface fitting method. We used two methods to align sets of neighbors. In the case where the two sets have the same cardinality, the rigid transformation between these sets is computed by using singular value decomposition (SVD) using equation (3) and (4). In the second case where the cardinalities are different, we use ICP to compute the best transformation. Since we use this iterative method on a local surface, this step will not consume a lot of time.

After the computation of the best transformation between two sets of neighboring points, we fit a surface to the nearest neighbors of each keypoint. Thereafter, the fitted surface is sampled into a grid of $n \times n$ (where $n = 20$) (see Figure 4). The values of the Z coordinate

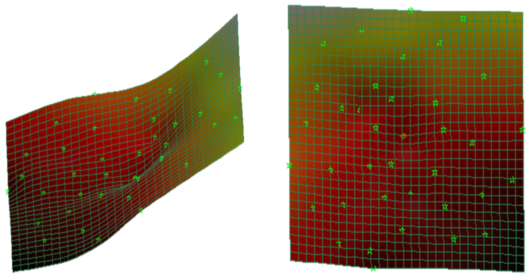


Figure 4: Fitted surface on nearest neighbors.

of the normal vector at each point along the grid represent information on the topology of the surface that is invariant to rotation of the point cloud. These values are stored in a vector, which represents a local descriptor for each point. The matching between two points is defined by the Euclidean distance between their descriptors.

4 EXPERIMENTS

Although our main goal is to use this approach with real data, we first tested it on synthetic data to verify the results in the lack of any noise and to validate the feasibility of this approach.

4.1 Validation of the method on synthetic models

We created a virtual scene with three theoretical amphorae and we added the 3D model of the bunny found in the Stanford[1] database. The purpose of this test is to identify the model of the bunny and align it with the scene thereafter.

For this first test on synthetic data, the model and the scene have the same resolution. We launched the test on all the points of the query model (bunny) and we found 82% of the points correctly matched. This result is illustrated in Figure 5. On the left image is shown the model and the right image the scene. The matched points are red and for clarity, we have shown just 3 point matches. We also notice in this same figure that the points that are not matched are close to the bounds of the model, which means that the descriptors of these points have been influenced by the points of the amphora which is located close. A rigid transformation between the model and the scene is calculated using the equations (3) and (4), but to infer if the model is well aligned with the scene, we calculate the mean square error between the matched points. The error obtained in this experiment is 0.05, which leads us to say that the object is fully recognized and correctly aligned with the scene (see Figure 6).

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

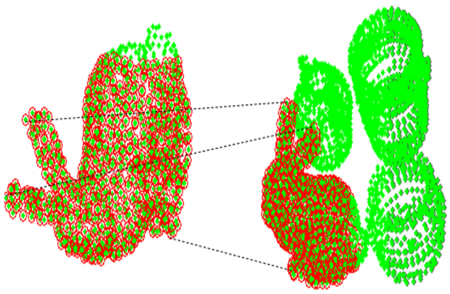


Figure 5: Features matching applied on synthetic data.

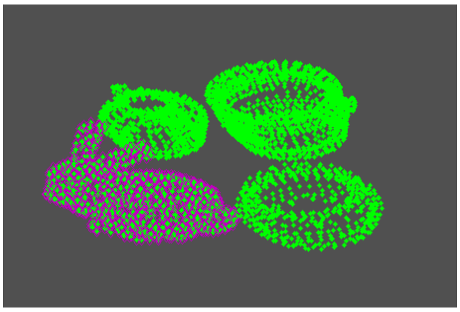


Figure 6: Data alignment and 3D object recognition.

In order to test the robustness of our descriptor with respect to the change of resolution, we took the same scene of Figure 5 which initially contained 3616 points and the bunny model which contains 1333 points. We downsampled the bunny model with different sampling step and we performed several tests. The test with a model of 846 points gives the best results (see Figure 7), below this value, the quality of registration is degraded. We did the same experiment using SHOT and FPFH descriptors that we have implemented using Point Cloud Library [Rus11a]. The results obtained are summarized in Table 2. These results show that our method is much better than SHOT and FPFH descriptor fails to find matched points in the case of heterogeneous data, which is confirmed in the documentation of the PCL library.

Methods	Points from model	Points from scene	Matched points	RMS
FPFH	307/846	1447/3616	0	-
SHOT	289/846	761/3616	45	15.5
Our approach	846/846	3616/3616	299	0.045

Table 2: Comparison of the results of our approach with that of SHOT and FPFH methods.

As we mentioned above, the radius of the sphere is used to select the set of neighbors of a point. We examined the impact of this radius on the quality of registration. We varied the radius of the sphere and we calculated the mean square error of the registration using matched points. The results are shown in Figure 8 and we found that the best result is obtained with a radius equivalent to 2% of the smallest dimension of the bounding box.

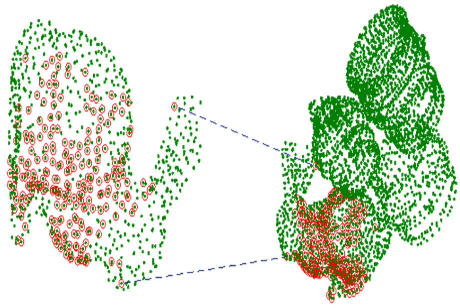


Figure 7: Matching between two point clouds with different resolutions.

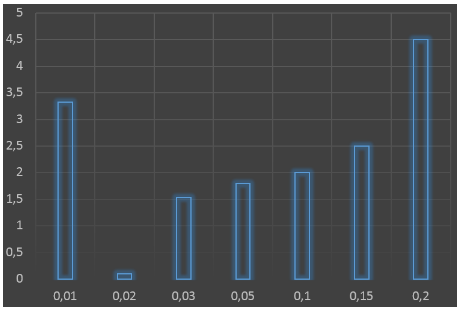


Figure 8: Radius impact on quadratic error (RMS).

4.2 Validation of the method on real data

For this first experiment on real data, we used surveys made in collaboration with the team of Prof. Guido Vannini, Department of Medieval Archaeology at the University of Florence, Italy. These surveys have accompanied the excavation of filling a vault at the last level of the steeple of the church of San Domenico in Prato, Italy. (see Figure 9).

Excavation for these surveys are comparable to underwater archaeology excavations on shipwreck ceramics, but with an access time to the site not limited by the constraints of diving. Therefore, the acoustic measurements were replaced here by the laser scanner surveys.

We used for this experiment, a model reconstructed from 53 photographs representing a small area of one of the five strata excavated (see Figure 10(a)) and the same area measured with a laser scanner. The reconstruction of the photogrammetric 3D model is obtained firstly by SFM approach (Structure From Motion); extraction of features, matching, pose estimation [Sna10a] and bundle adjustment [Lou09a]. Then densification of the 3D point cloud obtained with a patch approach [Fur10a] that provides us 3D model with high resolution of different areas of the site that have been photographed.

The vessel surrounded by a red curve in Figure 10(b) is used as a query object. A rotation of 30 degrees is applied to the model in order to test the invariance to rotation of the descriptor. The Figure 10(c) shows the result of the matching. The red dots represent the matched points, we can find that the object has been recognized in the scene. This result also supports our remark noted

(a)

(b)

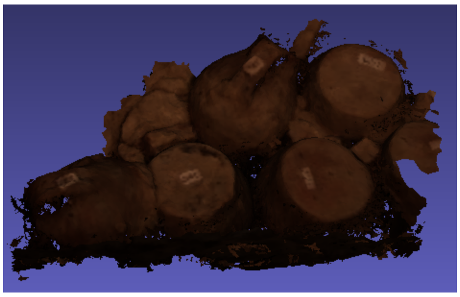
(c)

Figure 9: (a) Steeple of San Domenico, (b) Vault at the last level, (c) Stratum of the excavation filling.

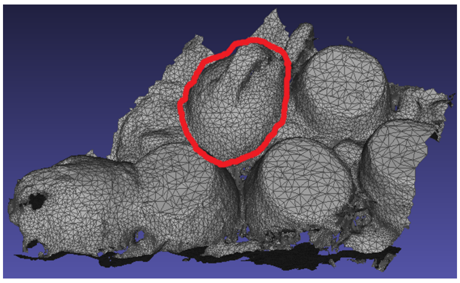
above. The unpaired points are located near the boundaries of the model. This means that the descriptor shape changes by changing its neighborhood, which is obvious, while the rest of the points whose neighborhood has not changed are well matched.

This last experiment presented below is made out on an underwater cave named *Imperial de Terre*, which lies off Marseille, France. Surveys were obtained during a joint mission between the COMEX, the LSIS laboratory and Septentrion Environment association. Three days were required, one day of preparation, a dive for taking photographic views and a day for the sonar acquisition.

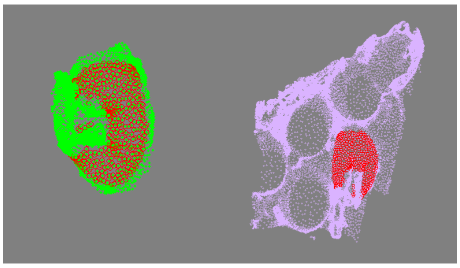
We used three synchronized cameras and 2260 photographs were used to produce the 3D model with very high resolution, shown in Figure 11(a). For the second data source, we used the high-frequency SONAR



(a)



(b)



(c)

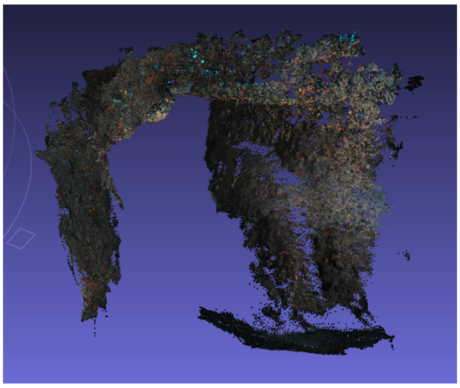
Figure 10: Recognition and registration of an amphora from real data. (a) 3D point cloud measured by photogrammetry, (b) 3D Mesh mesured by laser scanner, the query object is highlighted by red curve, (c) Features extraction and matching points

(BlueView BV5000) which provides a 3D point cloud with a lower resolution, without color information but with an extended range (see Figure 11(b)).

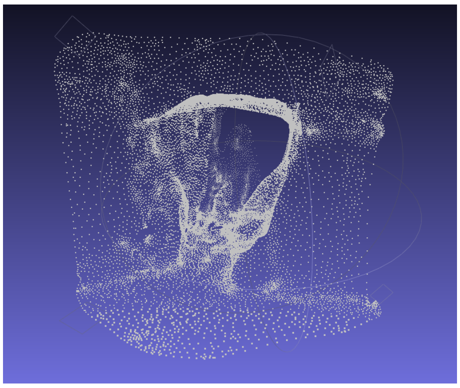
In Figure 12, we can see the result of the registration of photogrammetric data with Sonar model. Since we had a large volume of data with the presence of the noise, we thought about a semi-automatic approach. The user should select two areas of interest, one from the photogrammetric model and the second from model Sonar, to accelerate the research phase of homologous points. We specify that this approach is better appreciated by our partners archaeologists than automatic methods.

5 CONCLUSION AND FUTURE WORK

In our work, we are required to use an acoustic sensor to overcome the limitations of an optical sensor in an underwater environment. Both sensors that were used,



(a)



(b)

Figure 11: The *Impérial de terre* undersea cave (30m depth). (a) Photogrammetry data. (b) Acoustic data.

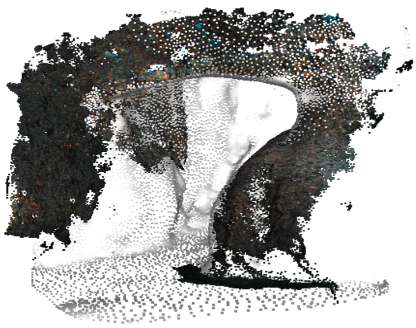


Figure 12: Registration of acoustic and optical data.

produce point clouds with different resolutions. The purpose of this study was to find a method of multi-modal registration to merge the point clouds obtained from each sensor. In this paper, we proposed an approach based on a descriptor invariant to rotation. The descriptor for each point is computed according to these neighboring points. To solve the problem of the resolution, we fit a surface on neighboring points. Thereafter, the fitted surface is resampled to extract a descriptor, that will be independent of the resolution of the point cloud.

Fusion using object recognition technique of the point clouds obtained from the SONAR and photogrammetry, as we have shown with the experience on vases, has two consequences: (1) the use for pose estimation (transformation between two clouds), (2) identification of the type of the object. This technique can be used in the case of Underwater Archaeology on shipwrecks with a large number of objects known a priori such as amphorae. This will not only help archaeologists to merge data, but also to specify the type of objects in the scene and investigate the variations with respect to theoretical models.

The purpose of this study was to find a method of multi-modal registration to merge the point clouds from each sensor. In this paper, we proposed an approach based on a descriptor invariant to rotation approach. A handle at each point is calculated based on these neighboring points.

As for future work, we plan to test more state-of-the-art descriptors, and we are still working on how best to choose the neighborhood radius automatically for use with any data without *a priori* information on the geometry of the point clouds.

6 REFERENCES

[Bas69a] G. F. Bass, Archaeology under Water, Penguin Book ed. Harmondsworth: Pelican Book, 1969.

[Fus00a] A. Fusiello and V. Murino, Calibration of an Optical-Acoustic Sensor, Machine GRAPHICS & VISION, vol. 9, pp. 207-214, 2000.

[Fus04a] A. Fusiello and V. Murino, "Augmented Scene Modeling and Visualization by Optical and Acoustic Sensor Integration," IEEE transaction on Visualization and computer graphics, vol. 10, pp. 625-636, 13 september 2004.

[Piz09a] O. Pizarro, R. M. Eustice, and H. Singh, "Large Area 3-D Reconstructions From Underwater Optical Surveys," IEEE Journal of Oceanic Engineering, vol. 34, pp. 150-169, Apr 2009.

[Sing00a] H. Singh, C. Roman, L. Whitcomb, and D. Yoerger, "Advances in Fusion of High Resolution Underwater Optical and Acoustic Data," in Proceedings of the 2000 International Symposium on Underwater Technology, Tokyo Japan, 2000.

[Neg07] S. Negahdaripour, H. Sekkati, and H. Pirsiavash, "Opti-Acoustic Stereo Imaging, System Calibration and 3-D Reconstruction," in Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on, 2007, pp. 1-8.

[Hur10a] N. Hurtós, X. Cufí, and J. Salvi, Integration of optical and acoustic sensors for D underwater scene reconstruction., Instrumentation Viewpoint, 2010.

[Bes92a] P. J. Besl and N. D. McKay, A Method for Registration of 3-D Shapes, IEEE Trans. Pattern Anal. Mach. Intell., vol. 14, pp. 239-256, 1992.

[Med91a] Y. C. a. G.Medioni., Object Modeling by Registration of Multiple Range Images, International Conference on Robotics and Automation, pp. 2724-2729, 1991.

[Lev01a] S. R. a. M. Levoy, Efficient Variants of the ICP Algorithm, Third International Conference on 3D Digital Imaging and Modeling (3DIM), 2001.

[Tur94a] G. a. L. Turk, M. , Zippered Polygon Meshes from Range Images, Proc. SIGGRAPH, 1994., 1994.

[Mas96a] T. Masuda, Sakaue, K., and Y okoya, N., Registration and Integration of Multiple Range Images for 3-D Model Construction, Proc. CVPR, 1996., 1996.

[Wei97a] S. Weik, Registration of 3-D Partial Surface Models Using Luminance and Depth Information, Proc. 3DIM, 1997.

[Rad08a] N. B. Radu B. Rusu, Zoltan C. Marton, and Michael Beetz., Aligning point cloud views using persistent feature histograms, pp. 3384-3391, September 2008 2008.

[Wah03a] U. H. E. Wahl, and G. Hirzinger., Surflet-pair-relation histograms: A statistical 3D-shape representation for rapid classification, In Proceedings of the fourth International Conference on 3-D Digital Imaging and Modeling, pp. 474-481, 2003.

[Rus09a] R. B. Rusu, N. Blodow, and M. Beetz, Fast Point Feature Histograms (FPFH) for 3D Registration, in Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), ed. Kobe, Japan, 2009.

[Joh99a] 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, pp. 433-449, 1999.

[Fro04a] A. Frome, Huber, D., Kolluri, R., Bülow, T., Malik, J., Recognizing objects in range data using regional point descriptors., In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3023, pp. 224-237, 2004.

[Fur10a] Y. Furukawa and J. Ponce, Accurate, Dense, and Robust Multiview Stereopsis, Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 32, pp. 1362-1376, 2010.

[Chu97a] C. S. Chua and R. Jarvis, Point Signatures: A New Representation for 3D Object Recognition, Int. J. Comput. Vision, vol. 25, pp. 63-85, 1997.

[Tom10a] F. Tombari, S. Salti, and L. D. Stefano, Unique signatures of histograms for local surface description, presented at the Proceedings of the 11th European conference on computer vision conference on Computer vision: Part III, Heraklion, Crete, Greece, 2010.

[Mia10a] 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. Comput. Vision, vol. 89, pp. 348-361, 2010.

[Pet11a] A. Petrelli and L. D. Stefano, On the repeatability of the local reference frame for partial shape matching, presented at the Proceedings of the 2011 International Conference on Computer Vision, 2011.

[Rus11a] R. B. Rusu and S. Cousins, 3D is here: Point Cloud Library (PCL), in IEEE International Conference on Robotics and Automation (ICRA), ed. Shanghai, China, 2011.

[Eri08a] J. D'Erico, Surface Fitting using Gridfit, MATLAB Central File Exchange, 2008.

[Sna10a] N. Snavely, I. Simon, M. Goesele, R. Szeliski, and S. M. Seitz, Scene Reconstruction and Visualization From Community Photo Collections, Proceedings of the IEEE, vol. 98, pp. 1370-1390, 2010.

[Lou09a] M. I. A. Lourakis and A. A. Argyros, SBA: A software package for generic sparse bundle adjustment, ACM Trans. Math. Softw., vol. 36, pp. 1-30, 2009.

[Che07a] Chen, H., & Bhanu, B. (2007). 3D free-form object recognition in range images using local surface patches. Pattern Recognition Letters, 28(10), 1252-1262.