Signature Extraction from 3D Point Clouds using Frame Theory for Environmental Modeling

Fabio Martino, Cosimo Patruno, Roberto Marani and Ettore Stella
Via Amendola n° 122 D/O,70126,Bari, Italy
martino@ba.issia.cnr.it

Abstract— In this paper a method based on the well-known frame theory is presented for the identification and classification of objects inside a scene. Three-dimensional (3D) point clouds have been firstly acquired using a laser triangulation system exploiting a high resolution camera, in order to derive accurate datasets for the method validation. The method performs a quadratic fit on the acquired samples and then extracts local curvatures from the analytical reconstructed surfaces. Such information is referred to a vocabulary of curvatures, created making use of the frame basis. Meaningful signatures can be finally analyzed to derive the recurrences of objects in the investigated scene. Specifically, by fixing a threshold value ζ, similarities can be estimated and thus objects can be recognized. Results prove the capability of the method to distinguish surface properties among several objects, validating this algorithm against the contributions of the measurement noise.

Keywords– Object Recognition; Mean and Gaussian Curvatures; Frame Theory; Robot Pose Estimation.

I. INTRODUCTION

Object recognition is one of the most important issue in the field of three-dimensional (3D) environmental modeling. Although considerable progresses have been made in Literature, it is still widely accepted that the recognition of the real world is a difficult task, especially when the scene representation is affected by occlusions [1]. The scene understanding becomes challenging especially whenever complex-shaped objects with ambiguous geometries are investigated [2].

Model-based algorithms can be classified in those looking at global or local features [3].

Global methods can exploit a new reference system, referred to the object under investigation. These objects are then modeled in terms of a parametric function defined in the new system of coordinates. Among them, the 3D generalized cylinder (GC) can be used to approximate the shape of objects [4]. However, the parameters involved in the GC can fit objects only in intensity images, thus disabling their applicability for the analysis of complex 3D point clouds, where samples are expressed in terms of spatial coordinates. Also super-quadratics are often used to derive a global representation of complex-shaped objects [5]. However, only few shapes can be represented with small residuals by means of this simple formulation and, consequently, intricate scenes incur huge computational costs, since objects have to be further segmented. Another global method is the one based on the Gauss Map [6], which confers good results that are independent of translations and perspective deformations. In this case, the limit of the technique is given by the presence of errors when the scene is partially occluded.

On the other hand, local methods try to represent objects by segmenting them. An example, of local descriptor can be found in the Normal Aligned Radial Feature (NARF) algorithm [7], where different kinds of edges are investigated and classified to derive the properties of the scene objects. In this case, once edge points are found, radial tangent patches are analyzed to classify whether the considered edge divides foreground objects or background/occluded regions. Despite of the good results provided [8], the analysis of edges can suffer from 3D measurement noise.

The problem of recognizing objects in a complex environment is one of the most considerable interest in the field of robotics, since it can be directly inverted for the motion estimation of autonomous vehicles (visual odometry [9]). In particular, whenever a couple of images can be acquired by a camera mounted on a mobile robot in two different consecutive time instants, it is possible to match the distinguishable features, i.e. signatures. As also valid for the well-known Iterative Closest Point (ICP) and RANdom SAmple Consensus (RANSAC) algorithms, relations can be inverted to estimate the robot movement in the environment [10]. In this context, the reference techniques for object recognition are the Scale Invariant Feature Transform (SIFT) [11] and the Speeded Up Robust Feature (SURF) [12], which operate only in two-dimensional images. As a consequence, this kind of algorithms need sophisticated camera calibrations to translate pixel displacements in terms of robot movements in world coordinates.

In this paper a method based on the frame theory for the definition of local surface descriptors obtained by 3D datasets is developed. Preliminary analyses have been performed on 3D point clouds derived by means of a triangulation-based laser scanner assisted by a high definition camera. Each returned measurement is labeled by a set of coordinates in a given reference system. For each sample of the dataset, a set of points belonging to a spherical volume with known radius is extracted, and then processed to define a least square (LS) fit on a quadratic function. The curvature information is then represented on a vocabulary, defined following the frame theory. The non-orthonormal but complete basis of the equivalent space is able to express the contribution of precomputed Gaussian and mean curvatures on the actual surface under analysis, made of the actual points. The final feature descriptor is thus derived in terms of a surface signature, which is directly linked to the weight of each element of the initial characteristic basis (frame). The paper is

organized as follows: Section II discusses on the theory beside the proposed method and focuses attention of the steps for the method implementation, whereas Section III introduces results on real data derived by the analysis of actual objects. Final conclusion and remarks on further activities will be presented in Section IV.

II. Mathematical Framework

A. Frame Theory

As stated previously, the presented method represents the object properties, derived by the analysis of 3D data, in terms of a projection on an equivalent space, following the frame theory [13].

A frame $\mathbf{F}$: $\mathbb{R}^n \rightarrow \mathbb{R}^l$ can be expressed as a $l \times n$ matrix which constitutes a basis, or in other words, the dictionary of the equivalent space. The matrix $\mathbf{F}$ can be defined as:

$$\mathbf{F} = [\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \boldsymbol{\varphi}_3, \ldots, \boldsymbol{\varphi}_n] \qquad \text{given } \boldsymbol{\varphi}_i \in \mathbb{R}^l \tag{1}$$

where the vectors $\boldsymbol{\varphi}_i$ that set up the dictionary are called atoms. Under these hypotheses, given an input vector $\boldsymbol{\alpha}$, which is related to the input points, it is possible to compute the corresponding signature vector $\mathbf{c}$ by means of the following formulation:

$$\mathbf{F} \cdot \mathbf{c} = \boldsymbol{\alpha} \rightarrow \mathbf{c} = \mathbf{F}^* \cdot \boldsymbol{\alpha} \qquad \text{given } \boldsymbol{\alpha} \in \mathbb{R}^l \tag{2}$$

where $\mathbf{F}^*$ is the pseudo-inverse of the Frame $\mathbf{F}$.

Accordingly to the frame theory, the matrix $\mathbf{F}$ can be effectively considered as a frame whether it satisfies a fundamental condition: the eigenvalues of the matrix $\mathbf{F} \cdot \mathbf{F}^T$ have to be higher than zero, i.e. comparable to the order of magnitude of the entries of the input vector $\boldsymbol{\alpha}$.

After this brief introduction to the Frame theory, it is fundamental to define the way the input vector $\boldsymbol{\alpha}$ has to be created. As already described, the entries of the vector $\boldsymbol{\alpha}$ are defined as an ordered set of couples of Gaussian and mean curvatures, computed punctually over a grid resampling the input surface. Given a generic point p of the input differentiable surface $\mathbf{S}$, two principal curvatures, i.e. the maximum and the minimum curvatures (k_1 and k_2, respectively), can be determined. As known, the product of the principal curvatures is named as Gaussian curvature (K) of $\mathbf{S}$ in p, while the mean of the same is called mean curvature (H) of $\mathbf{S}$ in p. Qualitatively, these terms are linked to the maximum and minimum radii of a curve passing through p and belonging to the surface $\mathbf{S}$. Therefore, elementary surfaces, such as planes, cylinders and spheres, have known curvature values, which are reported in Tab. I.

From a computational point of view, given the differentiable surface $z = f(x,y)$, where x and y are the independent variables belonging to $\mathbb{R}$, the Gaussian and the mean curvatures are defined as:

$$K = \frac{(f_{xx} f_{yy} - f_{xy}^{\,2})}{(1 + f_x^{\,2} + f_y^{\,2})^2} \tag{3}$$

$$H = \frac{(1 + f_x^{\,2}) f_{yy} - 2 f_x f_y f_{xy} + (1 + f_y^{\,2}) f_{xx}}{2(1 + f_x^{\,2} + f_y^{\,2})^2} \tag{4}$$

where:

$$f_x = \frac{df(x,y)}{dx}; f_y = \frac{df(x,y)}{dy}; f_{xx} = \frac{df_x(x,y)}{dx}$$

$$f_{xy} = \frac{df_x(x,y)}{dy}; f_{yy} = \frac{df_y(x,y)}{dy}$$

For more details the reader can refer to Ref. [14].

TABLE I: Gaussian and Mean Curvatures of Common Surfaces

Surfaces	K	H
Plane	0	0
Cylinder	0	$1/2r$
Sphere	$1/r^2$	$1/r$

In this case, the local Gaussian and mean curvatures have been chosen as characteristic descriptors of the surface properties, since these are perspective invariant. In particular, the Gaussian curvature takes advantage of the *theorema egregium*, which states that this feature is only ascribable to the properties of the object surface, regardless the observer point of view. In actual context, also mean curvatures own the same property, since wrapped surfaces can be observed in a one-shot acquisition from a single point of view. As a consequence, perspective ambiguities due to the possible existence of negative values of the mean curvature are practically avoided. By this way, the method can overcome one of the most relevant problem in object recognition and modeling, regarding the perspective deformation of scenes, when the same environment is acquired in two consecutive time instants from different points of view [15].

B. Frame generation

In the presented framework, a generic indoor environment is modeled as a combination of quadratic surfaces. In particular, paraboloid surfaces (Fig. 1) are the best candidate to fast and easily represent many kinds of actual primitive surfaces, such as spheres, cylinders and planes. The analytical formulation exploits within these lines is the following:

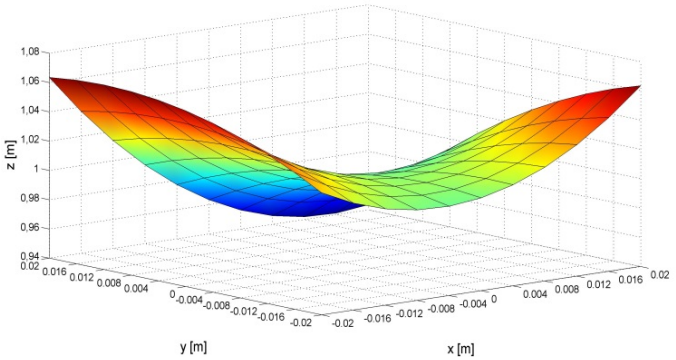

Figure 1. Surface of a paraboloid in the 3D space

$$z = ax^2 + by^2 + cx + dy + exy + f \quad (5)$$

where, the values assumed by the coefficients in Eq. (5), determines whether the generic paraboloid is hyperbolic or elliptic.

The generation of the dictionary, i.e. of the frame, is the first step in the dataset processing. As shown previously, curvatures can be easily derived by exploiting Eqns. (3) and (4). As a consequence a square domain of x and y, having side length of 40 mm, has been created and sampled by steps of 4 mm. In this domain, several paraboloids have been artificially created by varying the values assumed by the parameters in Eq. (5). For this purpose, parameters have been limited within the boundaries in Tab. II.

TABLE II: ADMISSIBLE RANGES OF PARABOLOID PARAMETERS

Parameter	Minimum value	Maximum value
a	-100	100
b	-100	100
c	-1	1
d	-1	1
e	-100	100
f	-0.02	0.02

Once all paraboloids have been generated, these are further selected by looking at their extensions. Since actual measurements will be limited by the sensor depth of field, only those paraboloids inscribed in a sphere of diameter equal to $\sqrt{3}$ times the patch size are kept in the frame generation. Once again, in this way, the real surfaces and the ones that artificially constitute the dictionary will undergo the same geometrical constrains.

Then, the Gaussian and mean curvatures of the generated paraboloids are computed for each point of the planar grid by using Eqns. (3) and (4). Here the partial derivatives are easily derivable:

$$f_x = 2ax + c + ey \quad (6)$$

$$f_y = 2by + d + ex \quad (7)$$

$$f_{xx} = 2a \quad (8)$$

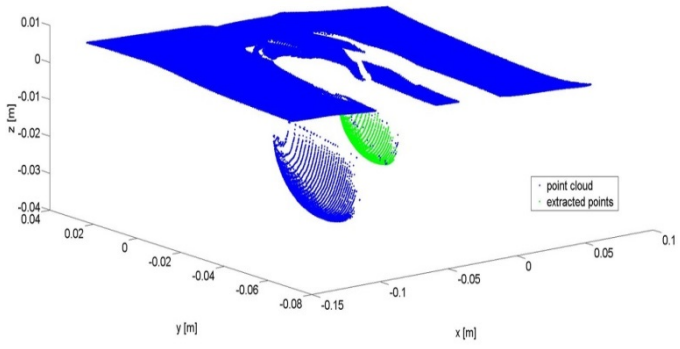

Figure 2. Real data of two spheres of different radius.

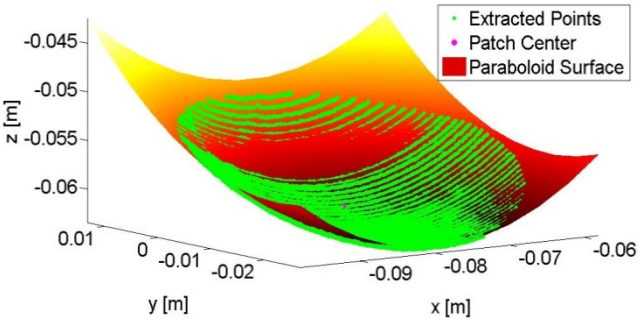

Figure 3. Approximating quadratic surface of the extracted real data (see the green dots in Fig. 2).

$$f_{xx} = 2b \quad (9)$$

$$f_{xy} = e \quad (10)$$

Therefore, **K** and **H** are two different matrices of the same size of the mesh grid (11×11 entries), representing the punctual values of the surface curvatures of the analytical functions z in Eq. (5). The two matrices are then reshaped columnwise in arrays of 121 elements each and then are concatenated to produce the specific atom $\boldsymbol{\varphi}_i$ of the frame. Therefore, each column of the frame comprises the Gaussian and mean curvatures of the analytical surfaces satisfying the previous fixed condition.

The final check for the produced matrix consists of the verification of the condition of non-vanishing eigenvalues of the product $\mathbf{F}\cdot\mathbf{F}^{\mathbf{T}}$. In this case, the smallest returned value was about 2.5×10^{-4}, i.e. comparable with the size of the geometrical square domain of existence of the generated paraboloids.

C. *Processing of real data*

As shown in the previous paragraphs, the validity of the proposed method has been tested by processing real data. In particular, several primitive objects, such as spheres, cylinders and planes, have been acquired under altered conditions, residing in different relative displacements and orientations. For each object under investigation, a 3D point cloud is extracted. An example of real data is reported in Fig. 2, where two spheres having different radii are displayed. The green dots represent the points used for the signature extraction. These are enclosed in a sphere of diameter equal to the initial patch side D, and centered on a sample belonging to the object under analysis. It is important to notice that, from now on, the reference system is oriented oppositely with reference to the vertical axis, and therefore heights are developed along the negative part of the z-axis.

The six parameters of the quadratic surface that best-fits the extracted points in the least square sense are determined by solving a system of nonlinear equations. Corresponding results are shown in Fig. 3, where the paraboloid approximating the green points in Fig. 2 is highlighted. It is worth noticing that the sampling grid is centered on the specific reference point around which the green dots in Fig. 2 are extracted. In this case, it is clear how real data suffers from the presence of measurements errors that produce holes in the acquired

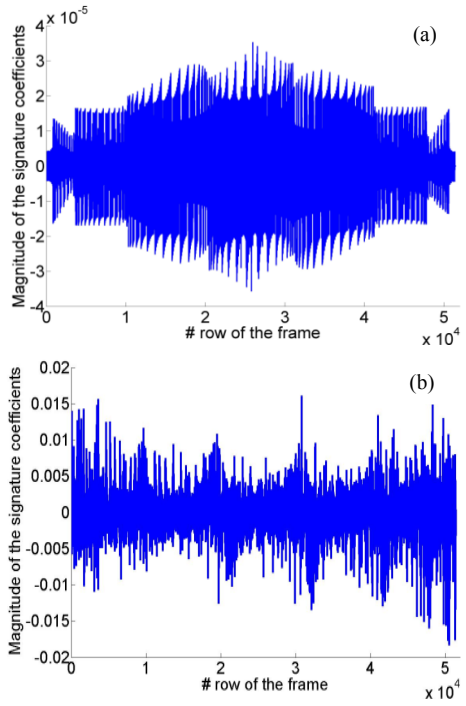

Figure 4. Examples of signature vectors extracted from a planar (a) and spherical (b) surface.

samples. Nevertheless, these issues do not affect the data processing, since the extraction of information is performed on the analytically reconstructed surface resulting from the non-linear least square fit.

The object under investigation is represented in terms of curvatures, which can be compared with the atoms of the vocabulary, i.e. the columns of the frame, by computing the partial derivatives in Eqns. (6) – (10), together with Eqns. (3) and (4).

The local signature can be obtained as a linear combination of the frame columns. In other terms, the vector **α**, having information about the surface curvatures, is used in Eq. (2) to compute the signature vector **c** that contains the weights of the linear combination of the frame columns. Therefore, different measurements of the same object generates the same signatures. An example of signature vectors derived by the analysis of a planar and spherical surface is reported in Fig. 4.

III. Results

As already discussed, object recognition is a complex task aimed to derive the recurring instances of a specific object in an actual environment, captured in several frames from different points of view. The following results are computed by comparing local signatures belonging to several real objects, scanned by means of a triangulation-based laser scanner.

The comparison is performed using a metrics defined as the dot product of the two signature vectors. Exploiting the definition of the dot product in Eq. (11), the distance between vectors is carried out using the normalized form in Eq. (12).

$$dot(v_1, v_2) = \|v_1\| * \|v2\| \cos\theta \tag{11}$$

$$\zeta = \cos\theta = \frac{|dot(v_1, v_2)|}{\|v_1\| * \|v2\|} \tag{12}$$

The proposed normalize distance ζ can also play the role of a confidence level bounded between 0 and 1. In this case, the following classification can be derived:

- $\zeta \geq 0.9$: optimal similarity;
- $0.5 \leq \zeta < 0.9$: good similarity;
- $0 \leq \zeta < 0.5$: dissimilarity.

Although many experiments have been performed, for the sake of simplicity, only twenty acquisitions are reported in this manuscript. Tab. III lists the set of acquisitions performed for the model validation and the corresponding identification number used in the following lines. Specifically, acquisition 1, 2 and 3 are derived from the inspection of three horizontal planes. The first is made of plastic, and the remaining of aluminum. The acquisitions of index $i = 4,\ldots,7$ are obtained by tilting the plastic plane, following the direction described by the normal vectors $\mathbf{n}_i$. Acquisitions 8 and 9 are related to different cylinders with comparable radii (r_8 and r_9), whereas the acquisitions of index $j = 10,\ldots,19$ are obtained by scanning the surfaces of seven spheres having different radius r_j. Acquisitions 16, 18 and 19 are redundant, since these are obtained by applying rigid displacements to the spheres of acquisition 11, 14, and 15, respectively. Finally, acquisition 20 is performed on an irregular surface.

TABLE III: List of Analyzed Objects

# acq	Kind of surface	Object property
1	Horizontal plane 1	Plastic medium
2	Horizontal plane 2	Aluminum
3	Horizontal plane 3	Aluminum
4	Tilted plane 1	$\mathbf{n}_4$ = [0.781, -0.499, -0.375]
5	Tilted plane 2	$\mathbf{n}_5$ = [-0.507, -0.028, -0.862]
6	Tilted plane 3	$\mathbf{n}_6$ = [-0.381, -0.334, -0.862]
7	Tilted plane 4	$\mathbf{n}_7$ = [-0.435, 0.268, -0.860]
8	Cylinder 1	r_8 = 24.27 *mm*
9	Cylinder 2	r_9 = 38.27 *mm*
10	Sphere 1	r_{10} = 18.85 *mm*
11	Sphere 2	r_{11} = 34.42 *mm*
12	Sphere 3	r_{12} = 29.70 *mm*
13	Sphere 4	r_{13} = 27.25 *mm*
14	Sphere 5	r_{14} = 14.15 *mm*
15	Sphere 6	r_{15} = 20.5 *mm*
16	Sphere 2	$r_{16} = r_{11}$
17	Sphere 7	r_{17} = 35.12 *mm*
18	Sphere 6	$r_{18} = r_{15}$
19	Sphere 5	$r_{19} = r_{14}$
20	Irregular surface	-

Fig. 5 reports the point cloud corresponding to the acquisition 5 (tilted plane 2). Green dots again displays the extracted patch used to derive the surface signatures. The inspection of Fig. 5 gives evidence to the contribution of noise due to the presence of cloud outliers, which can lead to processing errors.

The signature vector is compared with those derived by the other acquisitions, producing the results in Tab. IV, which

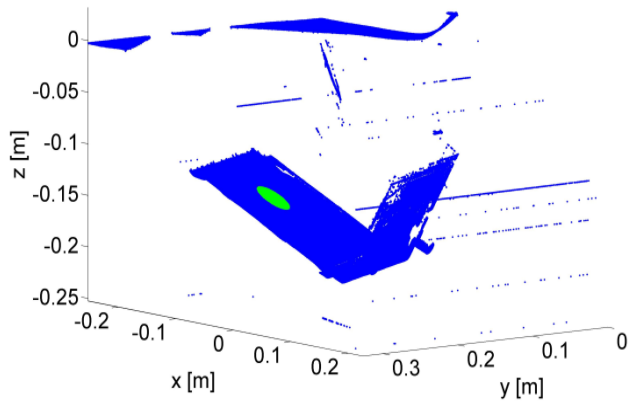

Figure 5. Point cloud of acquisition 5 (blue) and corresponding extracted patch used for the computation of the signature vector (green).

reports the corresponding ζ values. Confidence values reported in bold are the ones that verify the similarity condition expressed previously, i.e. with the best matching of signatures. Quantitatively, ζ values are very close to the unity, regardless the tilt experienced by the planar regions. Equivalently, this outcome states that planar objects always produce the same signatures, as expected independently of the point of view of the acquisition system.

TABLE IV: COMPARISON OF SIGNATURE VECTORS WITH THE ONE RELATED TO ACQUISITION 5 (TILTED PLANE 2)

# acq	ζ value	# acq	ζ value
1	**0.9996**	11	0.1207
2	**0.9996**	12	0.1687
3	**0.9996**	13	0.0107
4	**0.9667**	14	0.0174
5	-	15	0.0955
6	**0.9997**	16	0.1954
7	**0.9945**	17	0.1855
8	0.2615	18	0.0692
9	0.1036	19	0.0066
10	0.0795	20	0.2438

The dataset reported in Fig.6 is related to acquisition 8, where a cylinder is scanned by the laser profilometer. Results of the comparison of signatures are highlighted in Tab. V,

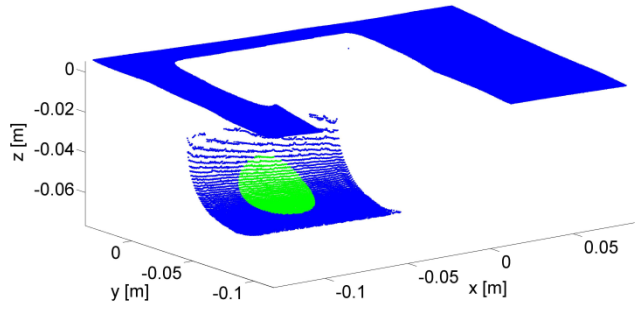

Figure 6. Point cloud of acquisition 8 (blue) and extracted points considered for the signature definition (green).

where the term ζ is shown. Also in this case bold values are those that indicate the best matching of signatures. Note that the only one value that passes the confidence check is the one related to the comparison with the second cylinder. In this case, the comparison produces a lower ζ value, since the considered cylinders have actually slightly different radii. On the contrary, the other values are much more different than the previous one, since different kinds of surfaces are compared.

Finally, Fig. 7 shows the point cloud related to acquisition 14, where a 28-*mm*-diameter sphere is investigated. Corresponding results of signature comparison are reported in Tab. VI.

TABLE V: COMPARISON OF SIGNATURE VECTORS WITH THE ONE RELATED TO ACQUISITION 8 (CYLINDER 1)

# acq	ζ value	# acq	ζ value
1	0.2710	11	0.0146
2	0.2727	12	0.0182
3	0.2727	13	0.0484
4	0.3062	14	0.0024
5	0.26154	15	0.0114
6	0.2719	16	0.0230
7	0.3052	17	0.0012
8	-	18	0.0156
9	**0.5951**	19	0.0025
10	0.0005	20	0.0731

TABLE VI: COMPARISION OF SIGNATURES WITH REFERENCE TO ACQUISITION 14 (SPHERE 5)

# acq	ζ value	# acq	ζ value
1	0.0170	11	0.0308
2	0.0170	12	0.0012
3	0.0173	13	0.0478
4	0.0164	14	-
5	0.0174	15	0.3034
6	0.0172	16	0.0003
7	0.0175	17	0.0011
8	0.0021	18	0.1959
9	0.0013	19	**0.6935**
10	0.2306	20	0.0391

Also the results in Tab. VI prove the capability of the proposed method to recognize the sphere 5 in the other measurements: the highest value, marked in bold, is obtained for acquisition 19, which is performed on the same object under different conditions. On the contrary, the remaining spheres produce lower values of ζ as the radius difference increases.

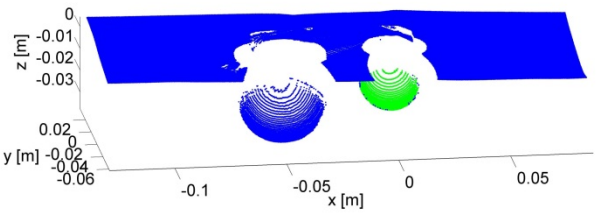

Figure 7. Point cloud derived by the inspection of spheres 5 (green) and 6.

IV. Conclusion and future works

In this paper a method object detection has been described and tested on real data. The method is implicitly unaffected by perspective deformations since it exploits Gaussian and mean curvatures to produce robust signatures, in accordance with the frame theory. An analytical basis of linearly dependent elements, the frame, is created by inspecting the curvature of parametric surfaces. This matrix is then used to derive the local signature of the actual surface, which is decomposed in terms of a weighted summation of the frame entries. Results confirm that the method can distinguish elementary objects among several primitives, i.e. planes, cylinders and spheres, regardless the entity of the measurement noise and without any ambiguity. This preliminary work opens the way for further activities which will lead to the use of the propose method for the motion estimation of autonomous vehicles through the pose estimation of recognized objects. Also the use of corner detectors for the unsupervised determination of interesting surface regions can improve the proposed method since it enables the automatic detection of the patch center, which constitutes the starting point of the proposed algorithm.

Acknowledgment

This work was funded within the CNR-ISSIA project "PI-LOC – Technological system for the automation of logistics processes in critical contexts" (P.O. Puglia fesr 2007-2013 line 1.2 – action 1.2.4).

References

[1] F. Dell'acqua,R. Fisher,."Reconstruction of planar surfaces behind occlusion in range images", IEEE Trans. Pattern Anal. Mach. Intell., Vol. 24, No. 4, pp. 569-575, April 2002.

[2] R. Benlamri, "Curved Shapes Construction for Object Recognition", in Proc. IEEE Geometric Modeling and Processing (GMP), Wako, Saitama, Japan, 10-12 July 2002, pp. 197-204.

[3] H. Delingette, M. Hebert and K. Ikeuchi, "A Spherical Representation for the Recognition of Curved Objects", in Proc. IEEE Computer Vision, 11-14 May 1993, pp. 103-112.

[4] A. D. Gross and T. Boult, "An algorithm to recover Generalized Cylinders from a single intensity view", in Proc. IEEE Trans. Robot. Autom., 13-18 May 1990, pp. 790-795.

[5] A.P. Pentland, "Perceptual Organization and the Rapresentation of Natural Form", Artificial Intelligence, Vol. 28, No. 3, pp. 293-331, May 1986.

[6] K. Ikeuchi, K.S. Hong, "Determining Linear Shape Change: Toward Automatic Generation of Object Recogniction Program" in Proc. IEEE Computer Vision and Pattern Recogniction (CVPR), 4-8 June 1989, pp. 450- 457.

[7] B. Stader, R.B. Rusu, K konolige and W.Bugard, "Point Feature extraction on 3D Range Scans taking into Account Object Boundaries", in Proc. IEEE Trans. Robot. Autom., 9-13 May 2011, pp. 2601-2608.

[8] M. Oshima, Y. Shirai, "Object recogniction using three-dimensional information", IEEE Trans. Pattern Anal. Mach. Intell., Vol. PAMI-5, No. 4, pp. 353–361, July 1983.

[9] F.Fraundorfer, D.Scaramuzza, "Visual Odometry part II: Matching, Robustness, Optimization and applications", IEEE Trans. Robot. Autom. Mag., Vol. 19, No. 2,June 2012.

[10] D.Scaramuzza, F. Fraundorfer, "Visual Odometry part I: The first 30 years and fundamental", IEEE Trans. Robot. Autom. Mag., Vol. 18, No. 4, December 2011.

[11] P. Scovanner, S. Ali, M. Shah, "A 3-Dimensional SIFT Descriptorand its application to Action recognition", in Proc. 15th International Conference on Multimedia, pp. 357-360, 2007.

[12] E. Oyallon, J. Rabin, "An analysis and implementation of the SURF method, and its comparison to SIFT", Image Processing On Line, 2013.

[13] E. Stella, "Novel Perspective on Automatic classification of Defects", Institute on Intelligent Systems for Automation, National Research Council, Bari, Italy, Rep. RI-ISSIA/CNR 04-2008, July 2008.

[14] M. P. Do Carmo, "Differential Geometry of Curves and Surfaces", Prentice Hall, 1976, pp. 162-164.

[15] A. Pogorelov, "Geometry", Mir Publishers Moscow, 1987, pp. 175-184.