3DOR based Global Pose Estimation for Service Robotics

Tiberiu T. Cocias, Sorin M. Grigorescu and Florin Moldoveanu

Transilvania University of Brasov
Dept. of Automatics

Abstract: Because of the large availability of 3D models, search engines for 3D shapes, also known as 3D Object Retrieval (3DOR), has begun to be widely used in research areas as object reconstruction or complex scene understanding. This paper addresses the problem of shape recognition with the main purpose of pose (position and orientation) estimation. The major advantage of the approach is that no GPS signal is needed for pose estimation. The perceived surface is searched over a database with possible candidates for similarities. The model with the highest confidence is then used to approximate the global position of the perceived surface using a series of geo-tags attached to each database model. Next, the robot pose is determined relatively to the perceived surface.

Keywords: 3DOR, shape recognition, partial matching, scene understanding

1. Introduction

There large amount of 3D models available over the internet or specific database has led to the development of reliable search engines for 3D shapes [1][2][3]. Thanks to the affordable and reliable 3D sensing devices like laser or structured light sensor (MS Kinect®), 3DOR engines can be successfully used in any robot vision related areas. Tasks like 3D object recognition, complex model segmentation or scene reconstruction can be easily tackled by a 3DOR approach [4][5]. The main issue related to the 3D object retrieval topic is reduced to the problem of determining the similarity of two given shapes or surfaces. Based on the representation of the shape, in the literature are proposed several approaches that deals also with 2D contours, 3D surfaces, 3D volumes or statistics [6][7]. To produce correct correspondences, the object retrieval process must follow a validation step. In [9] the validation occurs based on a ration between two distance samples whereas in [8] is presented a fast and simple validation algorithm which is based on the slope of the line connecting the corresponding points. This paper is organized as fallows. In chapter 2 the framework of a generic 3DOR algorithm is presented, followed by the global pose estimation principle in chapter 3. In the end, experimental results and conclusions will be presented.

2. 3D object retrieval framework

Conceptually, a 3D Object Retrieval framework (see Fig 1) aims at identifying a template model (query) among a large number of shapes. The origin of the technique lies in the abundance of available 3D representations on the internet. It is used mainly for object identification and surface reconstruction.

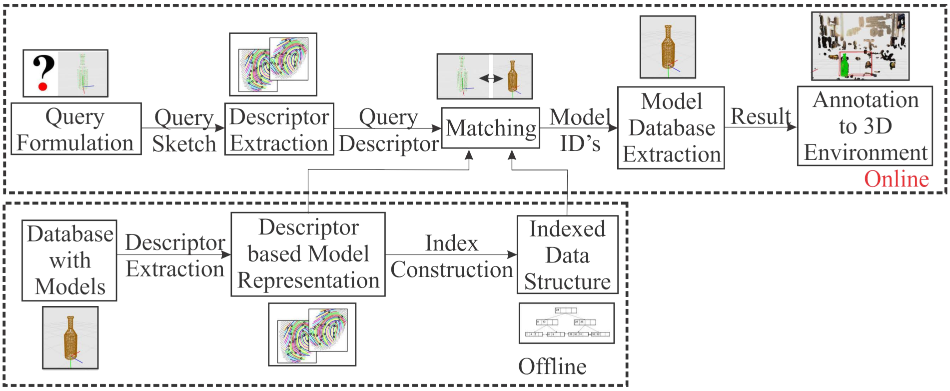

Figure 1: Generic 3D object retrieval framework.

For a large objects database, the algorithm cannot offer real-time model extraction, but it compensates with precision and reliability. Each model involved in this process has to be described in a common frame, also known as a descriptor. As depicted in Fig. 1, the overall structure is coarse divided in two sections: online and offline. The first sub-structure is used to process only variable information, such as templates descriptors extraction, or correspondence identification and validation, whereas the second sub-structure deals with static data such database models or database descriptors extraction. The similarity between the 3D shape models can occur only in the direction of partial matching. Finally, the database shape with the highest confidence (similarity) is used to represent the query model.

2.1. Shape unique representation

Since it is impossible to directly determine a similarity only by simply comparing two *Point Distribution Models* (PDM), a method which embeds surrounding information in a common frame representation has to be used. A descriptor collects surrounding feature information (Euclidean distance, color etc.) and embeds it into local representations which can be further used to identify similarities with other regions or models. The data is stored within a histogram which encodes neighbouring geometric information. Several descriptors used are: Signature of Histograms of OrienTations (SHOT), Spin Images, Points Features Histograms etc.

2.2. Measuring similarity

The matching process tries to determine similar regions between two given models. Since one of the models (query) can contain large occluded areas, the matching process occurs at a local level. As a direct result, only local descriptors can be used. The similarity identification is performed using brute force matching algorithm for independently comparing each point from the query with each point from the database models. Several similarity measure methods are used: L1-Norm, L2-Norm, Battacharyya distance, correlation coefficient, histogram intersection or chi-squared distributions. Fig 2 illustrates a series of correspondences between to similar PDMs.

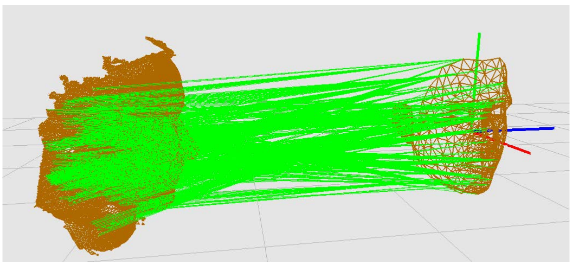

Figure 2: Determined correspondences between two similar models.

The L2-Norm, as depicted in Eq 1, is one of the most utilized comparison methods because it is discriminative and because of the math simplicity it can be computed in a short time. Similar, the Battacharyya approach offers improved precision but at a much high computational cost.

$$dist_{L_{2_{norm}}}(H_1,H_2) = \sqrt{\sum_i (H_1(i) - H_2(i))^2}, \quad (1)$$

where, H_1 and H_2 are the histogram of which similarity has to be measured.

2.3. Correspondence validation

By this stage, incorrect correspondences are eliminated. Because of similar histograms representations of some features, multiple correspondences can be determined. To avoid this, only points describing unique geometric regions are used to validate the similarity between two models. In this sense, flat or regular objects are difficult to match. A correspondence is considered to be valid only if the ratio between the closest $dist_{NN}$ and the second closest $dist_{secondNN}$ similarity measure is above 0.7 (see Eq. 2). This value ensures the selection of only distinctive point correspondences.

$$ratio = dist_{NN}/dist_{secondNN}, \tag{2}$$

where, $ratio = [0{,}1]$. A high value for the ratio indicates a distinctive correspondence because among all searched samples the selected correspondence is considered to be unique and distinctive. On the other side, a low ratio value indicates that several feature points have similar representation and choosing a random one can be ambiguous.

3. Global Pose Approximation

The main purpose of using a 3DOR search engine for pose identification is to enable global coordinate retrieval in any environment. Conventional systems based on the GPS signal simply cannot work in indoor environment because of the poor signal coverage. In this sense, a 3DOR approach can be used to identify an representative landmark (e.g. statues, distinctive pylons or any other unique construction) inside the environment, and further compute the relative position of the milestone from the closest known global coordinate. For the case of a building, the global robot coordinates can be determined using the geo-tagged building corners and the relative pose of the perceived landmark.

4. Experimental results

The experimental set-up used during test was as fallows. For scene depth perception, an MS Kinect ® RGBD sensor has been used. During tests, constant illumination was ensured. For principle evaluation purposes, the datasets used as model database were extracted from a large laser scan of the *Matchu Pichu* monument [10]. The similarity between the scene perspective and the database models [10] was performed using the L2-Norm and for correspondence validation the threshold ratio technique has been used. Through these experimental results, the authors wanted to prove and validate the concept of perspective identification with the main purpose of location retrieval. In Fig ‘3 is presented the recognition of a distinctive part of a temple chamber. With red are marked correspondent points between the compared models whereas with blue are marked the points resembling strictly the model database.

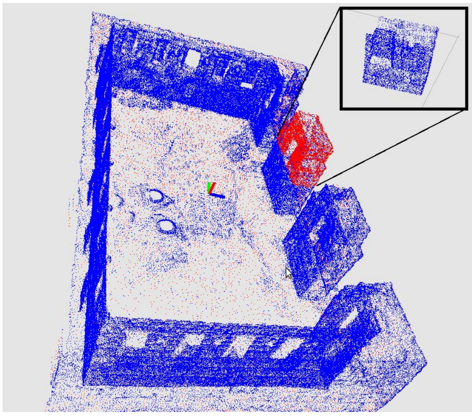

Figure 3: Identifiing a specific region from a building.

As can be seen in Fif 3, a large amount of sparse outlier correspondences, marked with red dots, were determined. The highest cluster of matched points is concerning exclusively the region depicting the robot perspective. In this sense, the larger cluster of corresponding points is considered to be the best candidate sample out of which the relative robot position can be estimated. For regions with regular geometry it is almost impossible to create correspondent points clusters because of the ambiguity of the distribution. The perceived scene model is represented trough 12.721 points whereas the building model is defined by 245.685 points. Because of the large amount of points, the computation time was 10 minutes and 35 seconds.

Conclusions

In this paper, a global pose estimation using 3DOR search engine for indoor service robots has been proposed. The goal of the approach is to enable accurate global pose estimation using only visual information extracted directly on the scene. Particular, the system can be used to augment the perceived scene enabling thus the understanding of the scene. As future work the authors consider the time computation enhancement of the proposed procedure through parallel computational devices (e.g. Graphic Processors), as well as the application of the method to other computer vision areas, such as 3D medical imaging.

Acknowledgements

This paper is supported by the Sectoral Operational Program Human Resources Development (SOP HRD), financed from the European Social Fund and by the Romanian Government under the projects POSDRU/107/1.5/S/76945 and POSDRU/89/1.5/S/59323.

References

[1] Min, P., Halderman, J.A., Kazhdan, M. and Funkhouser, T.A.: *Early experiences with a 3D model search engine*, Web3D Symposium, (2003), pp.7–18.

[2] Chen, D.Y., Tian, X.P., Shen, Y.T. and Ouhyoung, M.: *On visual similarity based 3D model retrieval*, Computer Graphics Forum, Vol. 22, No. 3, (2003).

[3] Suziki, M.T.: *A search engine for polygonal models to support development of 3D e-learning applications*, Tenth International World Wide Web Conference Poster Proceedings, (2001), pp. 182–183.

[4] Sebastian, T.B., Klein, P.N. and Kimia, B. B.: *Recognition of shapes by editing shock graphs*, In ICCV, (2001), pp. 755–762.

[5] Hartley, R.I.: *An object-oriented approach to scene reconstruction*, IEEE International Conference on Systems, Man, and Cybernetics, Vol. 4, New York, (1996), pp. 2475-2480.

[6] Zhang, H. and Fiume, E.: *"Shape matching of 3-D contours using normalized fourier descriptors*, Proceedings of the shape modeling international (SMI'02), (2002), pp.261-268.

[7] Tangelder, J.W. and Veltkamp, R.C.: *A survey of content based 3D shape retrieval methods*, Multimedia Tools Appl., Vol. 39, No. 3, (September 2008), pp. 441-471.

[8] Grigorescu, S.M.: *On robust 3D scene perception and camera egomotion estimation*, Proceedings of the 33rd. Colloquium of Automation, Leer, Germany,(2011).

[9] Čech, J., Matas, J. and Perdoch M.: *Efficient Sequential Correspondence Selection by Cosegmentation*, Vol. 32, No. 9, (2010), pp. 1568-1581.

[10] *** "*Computer Modelling of Heritage*", 2012. [online]. Available: http://www.cast.uark.edu/home/research/archaeology-and-historic-preservation/archaeological-geomatics/archaeological-laser-scanning/laser-scanning-at-machu-picchu.html.