

Grasp Exploration for 3D Object Shape Representation using Probabilistic Map

Diego R. Faria¹, Ricardo Martins¹ and Jorge Dias¹,

¹ Institute of Systems and Robotics,
Department of Electrical Engineering and Computers, University of Coimbra – Polo II,
3030-290 Coimbra, Portugal
{diego, rmartins, jorge}@isr.uc.pt

Abstract. In this work, the representation of 3D object shape acquired from grasp exploration is presented. Electromagnetic motion tracking sensors are attached to the fingertips for object contour following, acquiring thus the 3D points to represent the object shape using a probabilistic volumetric map. The object-centric representation is adopted by finding the centroid through 3D moment invariants. The occupancy of each individual voxel in the map is assumed to be independent from the other voxels occupancy, and the posterior information is obtained by using the Bayes' rule, thus achieving the probability distribution on the occupations percentage for each voxel. The conversion of the probabilistic map in Cartesian space to the spherical coordinate system for visualization with more resolution in an egocentric view is proposed.

Keywords: Grasp exploration; Probabilistic map; Object shape representation.

1 Introduction

Applications of service robots will require advanced capabilities of grasping objects and also skills that allow a robot to recognize an object through the grasp exploration. Human uses the hand for recognizing some object properties such as size, texture, etc. The skills of grasp exploration to acquire object properties are important also in the robotic field to complement other sensors such as vision and laser to obtain more information of an object. The in-hand exploration strategies performed by humans have motivated the development of analogous approaches to promote the exploration of surfaces of objects by robotic hands. These exploratory procedures are used to create internal representations of objects in order to proceed to its identification and integrating its 3D spatial and physical characteristics distribution to control the interaction with it.

In this work, the representation of the shape of a 3D object extracted from the data acquired during the human hand exploration of the object (by contour following) is proposed. The representation of object shape is built using a probabilistic volumetric map. The frame of reference associated to the representation of the object shape is located in the estimated object centroid.

2 Contribution to Technological Innovation

Nowadays, there is a tendency to move the development of robotic manipulators from simple grippers towards more natural human inspired articulated robotic hands, with integrated multimodal sensing technologies. This new generation of autonomous dexterous robotic hands is playing an important contribution to the evolution of the robotic research paradigms. The field of robotics is moving to the development of new systems, adapted to be introduced in new environments far from the traditional robotic applications in the industrial production lines. There is an emerging trend in the development of methods and strategies to endow these robotic platforms with the ability to autonomously explore these environments, proceed to the characterization of objects, use them as tools or interact with them. This work intends to contribute to the development of autonomous dexterous robotic hands by developing methods to extract a representation of the manipulated object in order to estimate the object shape and size. These methods are developed using contact points acquired from demonstrations performed by humans during the in-hand exploration. The developed methodology will be then transferred to robotic dexterous hands in order to estimate the shape and size of a manipulated object, from contact points acquired by a robotic hand through contour following. Achieving the 3D object shape estimation, it is possible to determine, through its geometrical properties, the best region of the object to perform a specific grasp. Using this knowledge (object representation) allows to endow a robot to grasp different types of objects including unknown objects.

3 Related Work

Researches about human perception concerning haptic exploration disclose that contour following is an ordinary way of “exploratory procedures” that people use for determining the geometry of an object [1]. Many approaches have been proposed regarding robot haptic exploration for object recognition and object shape representation. In [2], the authors present haptic object recognition using a dexterous robot hand with a manipulator arm. Through the hand contact by enclosing the objects at predefined positions was possible evaluate joint angles and force readings. The object shape recovery was performed using sparse contact points from hand. The point clouds were fitted to superquadric models defined by a set of parameters describing shape and pose. In [3], the authors developed a framework for haptic exploration by contour following that can be used with robot or human hand. The object shape contour is acquired by a human hand using a data glove where the human operator visually guides the contact sensor along the contours of the object by the index fingertips. An extended superquadric function for primitive modeling which can represent a variety of cubical and spherical geometries is used. They address modeling of basic superquadric shapes. Stereo vision is also used for wrist tracking by particle filter. A marker bracelet attached to the wrist of the subject for its tracking was used. The proposed work in [4] presents superquadric functions for shape recovery from haptic exploration with multi-fingered robot hands using fingertip tactile sensors. They have applied a hybrid minimization method using a genetic

algorithm by considering the contact normal information to recover superquadric primitives from synthetic exploration data. In a previous work [5], we developed a method to extract basic shapes (primitives) from data acquired from grasp exploration. Gaussian Mixture Models for points clustering and outlier removal were used. After a learning phase, the object shape was classified through Bayesian techniques. Least square minimization of a distance metric to find the shape orientation and scale for its representation was adopted.

4 Research Contribution and Innovation

This section describes the methodology adopted for object representation using probabilistic map. Mapping techniques as occupancy grid [6] [7] have been used in robotics field to describe an environment of a mobile robot. Two-Dimensional grid has been very used for static indoor mapping [8]. The idea is to verify the probability of each cell to be full or empty after the sensors observation. In [8], the authors describe a standard grid mapping algorithm as a version of Bayes filter. This filter is used to calculate the posterior over occupancy of each cell. Probabilistic volumetric maps are also useful as presented in [9] for data fusion (visual and auditory perception). The main idea of using the probabilistic map is due to the fact that it is a simpler way of static object reconstruction and representation, dealing with uncertainty of the sensor noise due to real world (i.e., the sensor probability model depends on the characteristics of the sensor and the object being sensed). In the next subsections the detailed content about the employed methodology for object representation through grasp exploration by contour following is presented.

4.1.1 Probabilistic Volumetric Map for Object Representation

This work presents a model for grasp exploration that allows data fusion from different sensors. The proposed work aims at representing the 3D shape of an object with object-centric representation, extending also its representation in a spherical coordinate system with more resolution. The volumetric map is updated along the exploration in discrete intervals. The proposed methodology can also be used for data fusion (multimodal perception), e.g. using vision and touch.

The occupancy of each individual voxel is assumed to be independent from the other voxels' occupancy and thus O_c is a set of independent random variables, where $c \in M$ represents the index of a cell on the Map; O_c is the probability value describing the occupancy of the cell c ; Z_{grasp} represents the grasp exploration measurement that influences one or more cells c ; this variable represents a set of sensors (attached to the fingertips) returning 3D positions of a movement over the object surface; $P(O_c)$ is the probability distribution of preliminary knowledge of coverage value describing the occupancy of a cell c , initially the prior is a uniform distribution; $P(Z_{grasp} | O_c)$ is the probability distribution corresponding to the set of measurement Z_{grasp} that influences a cell c taken from the grasp exploration data. This distribution is taken from the occupancy model. In case of data fusion, when working with vision, then

$P(Z_{vision} / O_c)$ represents the sensor model. The joint distribution decomposition of the relevant variables shows the dependency assumptions according to Bayes' rule as follows:

$$P(O_c Z_{vision} Z_{grasp}) = P(O_c)P(Z_{vision} | O_c)P(Z_{grasp} | O_c). \quad (1)$$

The *posteriori* is the probability distribution on the occupation's percentage $P(O_c | Z_{vision} Z_{grasp})$ for each voxel:

$$P(O_c | Z_{vision} Z_{grasp}) = \alpha P(Z_{vision} | O_c)P(Z_{grasp} | O_c)P(O_c), \quad (2)$$

where $\alpha = \frac{1}{P(O_c | Z_{vision})P(O_c | Z_{grasp})P(O_c)}$ is a normalization factor.

Since in this work we focus on grasp sensor model, then we simplify the estimation model of occupancy, so that the *posteriori* in this case is the probability distribution on the occupation's percentage $P(O_c | Z_{grasp})$ for each voxel. The occupation's probability is given by:

$$P(O_c | Z_{grasp}) = \frac{P(Z_{grasp} | O_c)P(O_c)}{P(Z_{grasp} | O_c)P(O_c) + 1 - (P(Z_{grasp} | O_c)(1 - P(O_c)))} \quad (3)$$

where $P(O_c | Z_{grasp})$ is the posteriori value, i.e., the cell occupancy given the sensor measurement; $P(Z_{grasp} | O_c)$ is the measurement that comes from the sensor model (likelihood); $P(O_c)$ is the *priori* information (at the beginning it is a uniform distribution representing the state full or empty) and subsequently the last *posteriori* becomes the *prior* for the next computation.

4.1.2 Grasp Exploration Occupancy Model

For the data acquisition (grasp exploration), we are using the Polhemus Liberty system [10]. One magnetic sensor is attached to each fingertip (e.g., thumb, index and middle fingers). Each sensor return the 3D coordinates based on Polhemus Liberty frame of reference. The frame rate of each sensor was configured to be up to 15Hz. During the data acquisition, a workspace (35x35x35cm) for object mapping in the experimental area is defined. This space is subdivided into cells with 0.5x0.5x0.5cm. For object exploration (i.e., contour following), along the displacement of each sensor in the workspace area, the information about the sensors are given (3D positions), being possible to identify in which cell that measurement is inserted. Due to the size of each cell relatively to the standard deviation of the magnetic tracking sensors measurements (up to 0.2 cm for linear movements of 10cm), inside each cell a 3D isotropic Gaussian probability distribution is defined, i.e., the sensor model $P(Z_{grasp} | O_c)$ is assumed to be Gaussian, centred at the cell central point with standard deviation 0.2cm and mean value equal to the cell central point coordinates of the cell. The probability of a measurement belonging to that cell is given by (4). The values are normalized in order to consider that the probability value assigned to a point located at the centre of the cell is equal to 1.

$$P(Z_{grasp} | O_C) = \frac{1}{(2\pi)^{3/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right), \quad (4)$$

where $|\Sigma|$ represents the determinant of Σ (sensor noise variation). It can also represent a scalar value. Due to the normalization used, the equation (4) takes the form:

$$P(Z_{grasp} | O_C) = \exp\left(-\left(\frac{(x-u_x)^2 + (y-u_y)^2 + (z-u_z)^2}{2\sigma^2}\right)\right). \quad (5)$$

4.1.3 Object Frame of Reference & Representation in Spherical System

In order to obtain a better representation of the object, we follow the object-centric representation. To find the object frame of reference (allowing also spherical representation), we compute the object centroid. The centroids of the points' distributions that represents the object is found along the grasp exploration in the workspace. While the data is continuously acquired during the exploration, the object centroid is changing (Fig. 1(a)). The frame of reference of the tracker sensors is kept, so that we just need to translate the points in the tracker frame of reference to the object frame of reference (its centroid as origin). Using the object centroid, we can define the radius of a sphere, allowing thus the object-centric representation as a spherical representation as exemplified in Fig.1 (b). The centroid of a 3D object shape is found by computing the discrete moments (3D moment invariants). The moments are a measure of the spatial distribution of the mass of a shape. The centroid is obtained through some steps, computing the zero moment (sum of the voxels); first moments for x , y and z , and then the centroid for each axis are computed in an iterative way, at some discrete intervals alongside the exploration, then the centroid is updated.

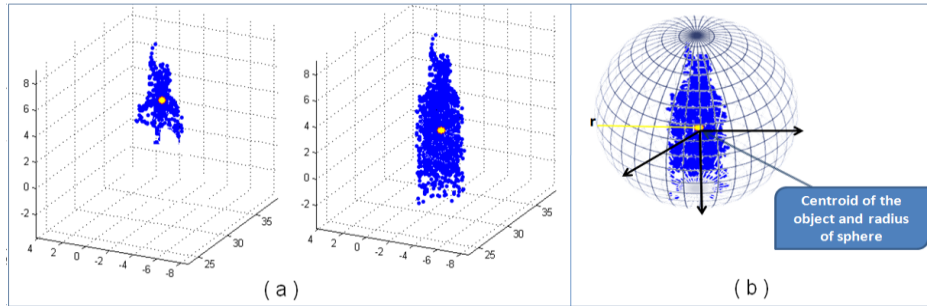


Fig. 1. (a) The center of mass of the object changes along the grasp exploration; (b) Illustration of the object representation in spherical coordinate system.

4.1.4 Experimental Results

Contour following was chosen as exploratory procedure. For this experiment, the subjects were instructed to use only the thumb, index and middle fingers to explore the object surface. We notice that using only these fingers is enough to cover the object shape through the movements around it. The setup for the experiments is composed of a wooden table, without any metallic parts, since the electromagnetic tracker is sensitive to nearby ferromagnetic materials. The rigid 3D object is fixed on the tabletop in a defined workspace for all experiments having the object in the same area. The magnetic tracker emitter unit that determines the frame of reference for the motion tracking system is placed on the tabletop at ~30cm near to the object. Computing a simple translation is enough to transform the 3D points on the sensors frame of reference to the workspace frame of reference. Fig. 2 (a) shows our experimental setup. Each voxel of the volumetric map was defined to represent an area $0.5 \times 0.5 \times 0.5$ cm due to the precision of the sensor position resolution at 30cm range (i.e., less than 1mm) from the sensor emitter. The chosen exploratory procedure was performed during 90 seconds. The volumetric map was updated with the new sensors measurements every 15 seconds. The subject is allowed to perform movements in any direction. The chosen object for this experiment was a bottle of wine. Fig. 2 shows the results obtained with the grasp (in-hand) exploration.

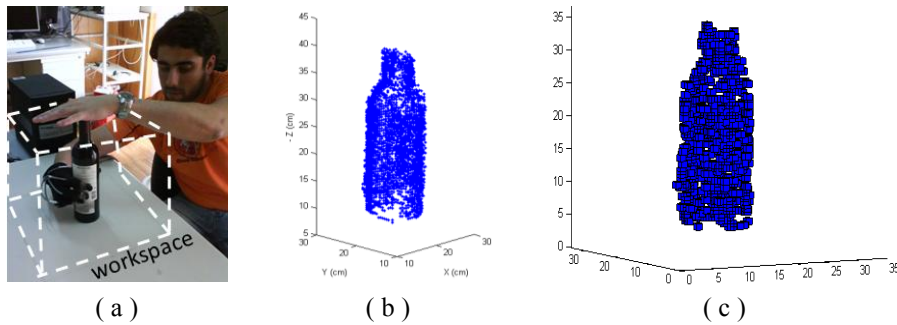


Fig. 2.(a) Experimental area for in-hand exploration; (b) raw data; (c) representation of the object shape (just voxels with probability higher than 0.7).

Fig.3 shows different views of the object centered in the estimated centroid.

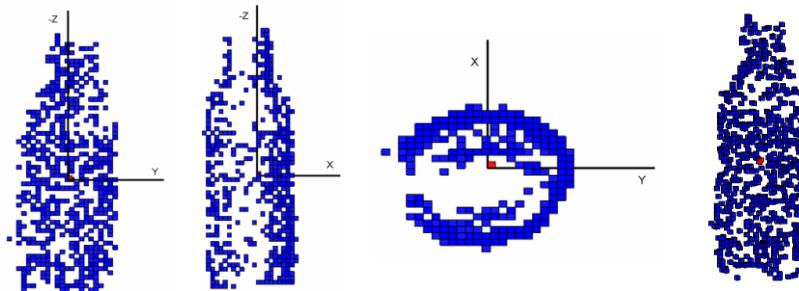


Fig. 3. Representation of the occupied cells with probability higher than 0.7 (object centered in the estimated centroid)

5 Discussion of the Results

Although we can represent the object shape by in-hand exploration, in this work this approach for shape representation has a limitation: the rigid object need to be static. If the object abruptly moves during the exploration, the shape representation can suffer imprecision. The idea of this methodology is to ignore points with low probabilities, so that movements out of the object surface are not represented. The application also ignores all points outside of the object workspace. Our approach uses a unique global map that is updated (i.e., using Bayes rule) when one or more fingers pass through the same cell, increasing the probability of occupancy. After analyzing the exploration performance of some subjects, we observed that the object shape influences the exploratory procedure, e.g., the bottle of wine was explored the most at the middle part, making the top and bottom parts obtaining lower probabilities. The retrieved object attributes such as shape and size were satisfactory. The raw data acquired from the fingers movements inside the workspace was a point cloud with 6169 points to represent the bottle presented in this work. For object shape representation, we have considered only the cells with probability higher than 0.7. After computing the probabilistic volumetric map, the object shape was comprised of 723 cells occupied (probability ≥ 0.7). The use of tactile sensors for contact points could facilitate and improve the results, e.g., by using tactile sensor we can know when exactly the fingers are in contact to the object to initialize the data acquisition, avoiding to start the application with the hand on the object or to avoid fill some voxels during the reach-to-grasp movement. In this work we have introduced the idea of working with multimodal perception. Our future direction will also consider using visual cues to acquire more information of the object, and then through sensors fusion we can obtain better results with the probabilistic representation, allowing features extraction for object characterization.

6 Conclusion and Future Work

This work introduces the strategy for object shape representation using a probabilistic approach to deal with sensor uncertainty. The method allows not only the full 3D shape representation, but also partial volume of the object when it is less explored, i.e., when just few contact points formed by a hand configuration over the object surface are performed. The object-centric representation was adopted by using a 3D occupancy grid to obtain the object frame of reference. Results show that valid models can be obtained with the proposed representation. In this approach, data from different sensors is allowed to improve the object model by means of multimodality. As future work, an extended model for sensor fusion using stereo vision and grasp exploration for better representation of the object will be tested. For that, a calibration between sensors has to be acquired to work with different data in the same frame of reference. In addition, experiments with non-static objects (moving objects) will be performed. For that, transformations to deal with translations and rotations of the object have to be computed, allowing registering the object point cloud into a single frame of reference.

Acknowledgments

This work is partially supported by the European project HANDLE (23-16-40). Diego Faria is supported by the Portuguese Foundation for Science and Technology (FCT).

References

1. Klatzky, R.L., Lederman, S.: Chapter 4 Intelligent exploration by the human hand. In: Dextrous robot hands. Springer-Verlag New York, Inc., New York, NY, USA (1990) 66-81.
2. Allen, P.K., Roberts, K.S.: Haptic object recognition using a multi-fingered dextrous hand. In: IEEE International Conference on Robotics and Automation. (1989) 342-347.
3. Bierbaum, A.; Welke, K., Burger, D., Asfour, T., Dillmann, R.: Haptic Exploration for 3D Shape Reconstruction using Five-Finger Hands. In: IEEE/RAS International Conference on Humanoid Robots. (2007).
4. Bierbaum, A. Gubarev, I., Dillmann, R.: Robust shape recovery for sparse contact location and normal data from haptic exploration. In: IEEE/RSJ International Conference on Intelligent Robots and Systems. (2008) 3200-3205.
5. Faria, D. R., Prado, J. A. S, Drews Junior, P., Dias, J.: Object shape retrieval through grasp exploration. In: ECMR-European Conference on Mobile Robots. (2009).
6. Moravec, H.: Sensor fusion in certainty grids for mobile robots. *AI Mag.* 9 (1988) 61-74.
7. Elfes, A.: Using occupancy grids for mobile robot perception and navigation. *Computer* 22 (1989) 46-57.
8. S. Thrun: Robotic mapping: a survey. In: Exploring artificial intelligence in the new millennium. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2003) 1-35.
9. Ferreira, João F., Pinho, C., Dias, J.: Active exploration using bayesian models for multimodal perception. In: ICIAR '08: Proceedings of the 5th International conference on Image Analysis and Recognition, Berlin, Heidelberg, Springer-Verlag (2008) 369-378.
10. Polhemus Liberty Electromagnetic Motion Tracking System: <http://www.polhemus.com>