Robotics and Autonomous Systems 61 (2013) 960–972

Contents lists available at SciVerse ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

Generic fitted shapes (GFS): Volumetric object segmentation in service robotics

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

Department of Automation, Transilvania University of Brasov, Mihai Viteazu 5, 500174, Brasov, Romania

HIGHLIGHTS

- Generic Fitted Primitives (GFP) for 3D volumetric segmentation using RGB-D sensors.
- 3D object modeling in mobile manipulation.
- Extension of 2D active contours to direct 3D object fitting.
- Energy minimization in the 3D Cartesian space.

ARTICLE INFO

Article history:
Received 7 February 2013
Received in revised form
25 April 2013
Accepted 29 April 2013
Available online 28 May 2013

Keywords:
Active contours
3D segmentation
3D reconstruction
Robot vision systems
RGB-D sensors

ABSTRACT

In this paper, a simultaneous 3D volumetric segmentation and reconstruction method, based on the so-called *Generic Fitted Shapes* (GFS) is proposed. The aim of this work is to cope with the lack of volumetric information encountered in visually controlled mobile manipulation systems equipped with stereo or RGB-D cameras. Instead of using primitive volumes, such as cuboids or cylinders, for approximating objects in point clouds, their volumetric structure has been estimated based on fitted generic shapes. The proposed GFSs can capture the shapes of a broad range of object classes without the need of large a-priori shape databases. The fitting algorithm, which aims at determining the particular geometry of each object of interest, is based on a modified version of the *active contours* approach extended to the 3D Cartesian space. The proposed volumetric segmentation system produces comprehensive closed object surfaces which can be further used in mobile manipulation scenarios. Within the experimental setup, the proposed technique has been evaluated against two state-of-the-art methods, namely superquadrics and 3D Object Retrieval (3DOR) engines.

1. Introduction

In the last decades, the number of service robotics systems centered on human environments has drastically increased, together with the introduction of novel sensors and actuators that push the boundaries of visual perception and mobile manipulation further [1]. Such applications span from common all-day-living assistance platforms [2] to care-giving robots deployed in hospitals and homes [3]. The main goal of a service robot operating in such environments is to autonomously perform an action in order to assist a human person in achieving his/her goal [4]. Among such tasks, autonomous object grasping in mobile manipulation [5] is one of the most researched areas within the robotics community, with imaging and computer vision being a common source of data for performing and improving grasping capabilities [6]. The success or failure of these procedures are directly dependent of the precision through which the imaged objects are reconstructed into the virtual environment of the robot [7].

Reconstructing and segmenting real-world scenes in service robotics scenarios is not a trivial task, especially when the robot perceives the environment from only one perspective. Considering the geometrical complexity of the scene and the uncertainty introduced by the single perspective, the achievement of mobile manipulation tasks is difficult to obtain. Given the challenge of handling complex objects, a robot must deal with both the reconstruction precision, as well as with the computation of the safest and most reliable grasp configuration [8]. The usage of a series of predefined shape models can provide structural information which further enables the estimation of the object of interest's volume. However, such an approach cannot model the particularities of each object to be grasped, thus leading to low grasping configurations.

In order to cope with such challenges, in this paper, we introduce the so-called *Generic Fitted Shapes* (GFS) concept with the goal to model the full volumetric structure of objects present in mobile manipulation scenarios. A GFS model addresses 3D

* Corresponding author. Tel.: +40 268 418-836; fax: +40 268 418 836.
E-mail addresses: tiberiu.cocias@unitbv.ro, tiberiu_cocias@yahoo.com (T.T. Cocias), moldof@unitbv.ro (F. Moldoveanu), s.grigorescu@unitbv.ro (S.M. Grigorescu).

0921-8890/$ – see front matter
http://dx.doi.org/10.1016/j.robot.2013.04.020

segmentation by the usage of a small amount of 3D models which are deformed in order to capture the particularities of each object. The deformation principle is derived from the concept of *active contours* [9], well-established in the 2D image segmentation area. Namely, a GFS structure is designed to capture the particular shape of an object by minimizing a set of *internal* and *external* 3D contour energies [9].

The challenge which we tackle lies in the area of 3D segmentation and reconstruction, which implies the partitioning of a certain n-dimensional space (2D digital image, 3D point clouds etc.) into a subset of meaningful regions representing objects of interest and the background of the scene. While for 2D images the most used characteristics are color or texture, for 3D point clouds, the most relevant attributes describing the scene are the surface's geometry, point normals, surface mesh or point density.

The problem of segmenting an object S from a scene point cloud can be regarded as a simultaneous segmentation and reconstruction of its 3D shape, given a sparse *Point Distribution Model* (PDM). An example of PDM segmentation, where the object is approximated through a GFS, is illustrated in Fig. 1.

1.1. Related work

With the advent of new structured light sensors, such as the MS Kinect®, the 3D segmentation area received a large attention from the computer vision and robotics community [10,11]. A series of implicit segmentation methods, entitled *superquadrics* aim at generating a 3D model $S(b)$, defined by a series of parameters b, for approximating the volume of the given surface [12]. As opposed to the work in [12], where a single superquadric is used to approximate the shape of an object, in [13], multiple superquadrics are merged with the purpose of enhancing the surface estimation precision. The obtained superquadric structure is a shape, describing the segmented object, which can be further used for grasping or mobile manipulation. Since the object is approximated by 3D geometric primitives (e.g. spheres, cubes, toroids, etc.), the final representation is rough. In [14], the authors calculated a refined object model by dividing the initial primitive into more meaningful subregions. In each subregion, a separate superquadric has been fitted, increasing thus the overall estimation accuracy of the model. The approach can be successfully used for simple regular objects like boxes, simple mugs, or cylindrical bottles. For more complex shapes, the algorithm fails to properly segment the object, delivering in the end over-segmented areas.

In [15], a series of kitchen objects (e.g. plates and mugs) are reconstructed using a similar implicit model formulation as for the superquadrics case. Again, simple geometric primitives (e.g. cylinders, cuboids or spheres) are used to estimate the structure of the objects, thus leading to rough representations.

3D deformable shapes have been extensively treated in the work of Terzopoulos [16], where the authors introduce a 3D deformable balloon primitive model with its shape driven by a series of internal forces (e.g. elastic properties). The introduced dynamical model does not need any a-priori knowledge about the object which must be estimated and, in the same time, no *position and orientation* (pose) normalization is required. The major drawback of this approach is that in the regions where no point information is available (occluded regions), the approximated 3D contour is described by a rounded balloon surface.

In comparison to the above presented methods, the so-called *explicit* approaches make use of a series of predefined models (e.g. particular primitives or models) to better approximate a certain surface [17]. The main idea here is to fit into the point clouds complex 3D shapes which best describe its structure. In this case, an appropriate shape model is searched within a database containing multiple predefined models. The similarity between the scene object and the objects from the database can be established using a 3*D Object Retrieval* (3DOR) search engine [18]. The database model, with the lowest convergence error (highest similarity), is further annotated to the scene. The major difference between a 3DOR system and the proposed GFS method lies in the object fitting approach. If, for the case of the 3DOR technique, the selected shape from the database remains static, that is, no adaptation with respect to the nature of the scene is performed on it, in the case of the GFS, the selected shape is automatically deformed in such a way that it captures the particularities of the imaged object.

Other approaches, which exclude the need of active sensors, make use of voxel-based 3D reconstruction techniques from multiple calibrated cameras to obtain an approximated 3D representation of a focused object. In [19], the authors ensure the consistency of the projected reconstruction with the original images by making explicit use of the finite size footprint of a voxel when projecting it into the image plane. The major disadvantage, considering the image based 3D reconstruction techniques, is that multiple views are needed to obtain a full model of the object of interest [20], which, in the case of a service robot, is difficult to obtain.

1.2. Structure and main contributions

The main contributions of the paper may be summarized as follows:

- the introduction of the GFS technique based on a 3D active contour formulation; the deformation of the shape model is performed with respect to energies calculated directly in the 3D Cartesian space;
- the usage of a GFS as an initial contour within the active contours framework, thus improving the computation time and avoiding erroneous approximations;
- usage of the GFS approach for building full 3D volumetric models of objects of interest in the context of mobile manipulation.

The rest of the paper is organized as follows. Section 2, describes the 3D perception apparatus used in imaging a robotic scene. In Section 3, the GFS model is detailed, followed by the segmentation approach given in Section 4. In the end, before conclusions and outlook, experimental results are presented in Section 5.

2. 3D scene perception apparatus

The main motivation for fully segmenting the 3D volumetric shapes of objects is to obtain an exhaustive 3D representation which can be reliably used for any further mobile manipulation task. The block diagram of the proposed scene analysis system is presented in Fig. 2. Starting from the PDM of an object and by the similarity transform applied to a correspondent generic shape, a rough 3D segmentation of the object's structure is obtained.

The process in Fig. 2 is initialized through tabletop segmentation [21], which calculates the initial reference models, or clusters, $C = [c_0, c_1, \dots, c_k, \dots, c_n]$ on which the GFS will be fitted. For the remainder of the paper, the explanations will be given for a single object cluster c_k. Although, in our work, we have used a standard tabletop segmentation algorithm, the proposed GFS approach can be applied on any given point distribution model.

The tabletop segmentation separates object clusters from a point cloud P_{in} by segmenting out the supporting plane on which the objects reside. Points are grouped together into a cluster if the Euclidean distance between them is smaller than 0.02 m. The output of this operation is represented by the vector of object clusters C. On each of $c_k \in C$ a GFS model will be fitted, as described in Section 4.

As opposed to the approach in [21], which concentrated on fitting fixed CAD models onto objects present on a table, in our