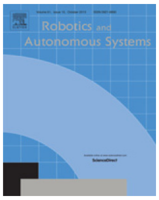
Robotics and Autonomous Systems 83 (2016) 312–325

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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



Incremental scenario representations for autonomous driving using geometric polygonal primitives

Miguel Oliveira [a,b,*], Vitor Santos [b], Angel D. Sappa [c,d], Paulo Dias [b], A. Paulo Moreira [a,e]

[a] INESC TEC - INESC Technology and Science, R. Dr. Roberto Frias s/n, 4200-465 Porto, Portugal

[b] IEETA - Institute of Electronics and Informatics Engineering of Aveiro, Universidade de Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal

[c] Facultad de Ingeniería en Electricidad y Computación, Escuela Superior Politécnica del Litoral, ESPOL, Campus Gustavo Galindo, Km 30.5 vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador

[d] Computer Vision Center, Campus UAB, 08193 Bellaterra, Barcelona, Spain

[e] FEUP - Faculty of Engineering, University of Porto, R. Dr. Roberto Frias s/n, 4200-465 Porto, Portugal

ARTICLE INFO

Article history:
Available online 7 June 2016

Keywords:
Incremental scene reconstruction
Point clouds
Autonomous vehicles
Polygonal primitives

ABSTRACT

When an autonomous vehicle is traveling through some scenario it receives a continuous stream of sensor data. This sensor data arrives in an asynchronous fashion and often contains overlapping or redundant information. Thus, it is not trivial how a representation of the environment observed by the vehicle can be created and updated over time. This paper presents a novel methodology to compute an incremental 3D representation of a scenario from 3D range measurements. We propose to use macro scale polygonal primitives to model the scenario. This means that the representation of the scene is given as a list of large scale polygons that describe the geometric structure of the environment. Furthermore, we propose mechanisms designed to update the geometric polygonal primitives over time whenever fresh sensor data is collected. Results show that the approach is capable of producing accurate descriptions of the scene, and that it is computationally very efficient when compared to other reconstruction techniques.

1. Introduction

Recent research in the fields of pattern recognition suggest that the usage of 3D sensors improves the effectiveness of perception, "since it supports good situation awareness for motion level teleoperation as well as higher level intelligent autonomous functions" [1]. Nowadays, autonomous robotic systems have at their disposal a new generation of 3D sensors, which provide 3D data of unprecedented quality [2]. In robotic systems, 3D data is used to compute some form of internal representation of the environment. In this paper, we refer to this as 3D scene representation or simply 3D representation. The improvement of 3D data available to robotic systems should pave the road for more comprehensive 3D representations. In turn, advanced 3D representations of the scenes are expected to play a major role in future robotic applications since they support a wide variety of tasks, including navigation, localization, and perception [3].

In summary, the improvement in the quality of 3D data clearly opens the possibility of building more complex scene representations. In turn, more advanced scene representations will surely have a positive impact on the overall performance of robotic systems. Despite this, complex scene representations have not yet been substantiated into robotic applications. The problem is how to process the large amounts of 3D data. In this context, classical computer graphics algorithms are not optimized to operate in real time, which is a non-negotiable requirement of the majority of robotic applications. Unless novel and efficient methodologies that produce compact, yet elaborate scene representations, are introduced by the research community, robotic systems are limited to mapping the scenes in classical 2D or 2.5D representations or are restricted to off-line applications.

Very frequently, the scenarios where autonomous systems operate are urban locations or buildings. Such scenes are often characterized for having a large number of well defined geometric structures. In outdoor scenarios, these geometric structures could be road surfaces or buildings, while in indoor scenarios they may be furniture, walls, stairs, etc. We refer to the scale of these structures as a macro scale, meaning that 3D sensor may collect thousands of measurements of those structures in a single scan. A scene representation is defined by the surface primitive that

---

* Corresponding author at: INESC TEC - INESC Technology and Science, R. Dr. Roberto Frias s/n, 4200-465 Porto, Portugal.

E-mail addresses: m.riem.oliveira@gmail.com (M. Oliveira), vitor@ua.pt (V. Santos), asappa@cvc.uab.es (A.D. Sappa), paulo.dias@ua.pt (P. Dias), amoreira@fe.up.pt (A.P. Moreira).

http://dx.doi.org/10.1016/j.robot.2016.05.011
0921-8890/

Table 1
LIDAR sensor systems mounted on some autonomous vehicle systems of recent years.

Institution	Vehicle	Ref.	3D sensor type		Total[a]
			2D laser	3D laser	
Stanford U.	Stanley[b]	[11]	5 × Sick LMS 291	–	67.5
CMU	Sandstorm[b]	[12]	3 × Sick LMS 291	–	50.5
	Highlander[b]		Riegl Q140i	–	
CMU	Boss[b]	[13]	6 × Sick LMS 291	Vel. HDL-64	2305.0
			2 × Continental ISF 172	–	
			2 × IBEO Alasca XT	–	
Stanford U.	Junior[c]	[14]	4 × SICK LMS 291	Vel. HDL-64	2278.0
			2 × IBEO Alasca XT	–	
Virginia Tech	Odin[c]	[15]	4 × Sick LMS 291	–	90.0
			2 × IBEO Alasca XT	–	
			IBEO Alasca AO	–	
MIT	Talos[c]	[6]	12 × Sick LMS 291	Vel. HDL-64	2361.2
U. Munich	MuCar-3[d]	[16]	–	Vel. HDL-64	2200.0
Google	Driverless Car	[17]	–	Vel. HDL-64	2200.0

[a] Estimation of total 3D data throughput of all LIDAR sensors mounted on the vehicle, given as points $\times 10^3$/s.
[b] These vehicles participated in the *Defense Advanced Research Projects Agency* (DARPA) Grand Challenge 2006.
[c] These vehicles participated in the DARPA Urban Challenge 2007.
[d] This vehicle participated in the *Civilian European Land Robot Trial* (ELROB) Trial 2009.

is employed. For example, triangulation approaches make use of triangle primitives, while other approaches such as Poisson surface reconstruction resort to implicit surfaces. Triangulation approaches generate surface primitives that are considered to have a micro scale, since a geometric structure of the scene could contain hundreds or thousands of triangles. Micro scale primitives are inadequate to model large scale environments because they are not compact enough.

In this paper, we present a novel methodology to compute a 3D scene representation. The algorithm uses macro scale polygonal primitives to model the scene. This means that the representation of the scene is given as a list of large scale polygons that describe the geometric structure of the environment. The proposed representation addresses the problems that were raised in the previous lines: the representation is compact and can be computed much faster than most others, while at the same time providing a sufficiently accurate geometric representation of the scene from the point of view of the tasks required by an autonomous system.

The second problem addressed in this paper is the reconstruction of large scale scenarios from a continuous throughput of massive amounts of 3D data. We will use the distinction between the terms scene and scenario. Let scenario refer to a particular location that should be reconstructed, e.g., a city, a road or a building. By scene, we refer to the portion of the scenario that is viewed by the vehicle at a particular time. Thus, the scenario is an integration of scenes over time. In the case of large scale scenarios, the compactness of a given scene representation is even more important. In this paper, we focus also on how the representation may evolve by integrating 3D data from multiple measurements over time.

This is an extended version of [4]. The new material covers mostly the incremental part of the geometric reconstruction. There is also the possibility of adding texture to the geometric scene description. For further details on this see [5].

For testing and evaluation purposes, we use a data-set from the *Massachusetts Institute of Technology* (MIT) Team, taken from their participation in the DARPA Urban Challenge [6]. From this data-set we have extracted a 40 s sequence which will be used to assess the proposed algorithms. For the remainder of the paper, this sequence is referred to as MIT sequence. Using this data-set, we aim at reconstructing large portions of the urban environment in which the competition took place.

The remainder of this paper is organized as follows: Section 2 reviews the state of the art, Section 3 presents the proposed approach. Results are given in Section 4 and conclusions in Section 5.

2. Related work

At first glance, it would seem plain to translate the improvement on the quality of the 3D data into an enhancement of the 3D representations. However, the fact is that the majority of the robotic systems, namely autonomous vehicles, continue to rely on classic 2D or 2.5D scene representations [7], such as occupancy grids [8] or elevation maps [9], or use discretized grid-like approaches as in octrees [10]. The reason for that is that autonomous vehicles commonly require a large array of sensors installed onboard and, as a consequence, collect large amounts of range measurements every second. Table 1 shows an estimate of the amount of 3D data (measured by LIDAR systems alone) generated by several autonomous vehicles. Simplified 2D or 2.5D representations are used so that they can be computed in real time using large amounts of data. More advanced 3D representations have not been introduced in robotics because they fail to abide to the requirements of real time processing using the 3D data produced by new generation LIDAR sensors. One example of this is the methodologies used in the computer graphics research field: traditional algorithms such as building of triangular meshes are unable to operate in real time with the throughput of data provided by new generation 3D sensors. Some authors have tried to optimize triangulation algorithms (e.g., [2,7]), and they report near real time performances. Note that these results were obtained using point clouds from a Microsoft Kinect 3D camera.[1] On the other hand, the results provided towards the end of this paper are obtained using point clouds from a Velodyne HDL-64E Lidar,[2] and therefore results are not directly comparable.

Scene reconstruction is defined as the computation of a geometric 3D model from multiple measurements. These measurements could be obtained from stereo systems, range sensors, etc. Scene reconstruction may also include the texturing of the generated 3D model. Scene reconstruction methodologies are grouped into two different approaches: surface based representations or volumetric occupancy representations. In the first, the underlying surfaces of the scene that generated the range measurements are estimated, while in the second, the range measurements are grouped into cells of a grid, and are then labeled free or occupied. Traditional surface based representations include several 3D triangulations methodologies, such as 3D Delaunay triangulation [18], or *Ball Pivoting*

---

[1] http://en.wikipedia.org/wiki/Kinect.

[2] http://velodynelidar.com/lidar/lidar.aspx.