QuadricSLAM: Dual Quadrics as SLAM Landmarks

Lachlan Nicholson, Michael Milford, and Niko Sünderhauf

***Abstract*— Research in Simultaneous Localization And Mapping (SLAM) is increasingly moving towards richer world representations involving objects and high level features that enable a semantic model of the world for robots, potentially leading to a more meaningful set of robot-world interactions. Many of these advances are grounded in state-of-the-art computer vision techniques primarily developed in the context of image-based benchmark datasets, leaving several challenges to be addressed in adapting them for use in robotics. In this paper, we derive a SLAM formulation that uses dual quadrics as 3D landmark representations, exploiting their ability to compactly represent the size, position and orientation of an object, and show how 2D bounding boxes (such as those typically obtained from visual object detection systems) can directly constrain the quadric parameters via a novel geometric error formulation. We develop a sensor model for deep-learned object detectors that addresses the challenge of partial object detections often encountered in robotics applications, and demonstrate how to jointly estimate the camera pose and constrained dual quadric parameters in factor graph based SLAM with a general perspective camera.**

I. Introduction

In recent years, impressive vision-based object detection performance improvements have resulted from the "rebirth" of Convolutional Neural Networks (ConvNets). Despite these impressive developments, the Simultaneous Localization And Mapping community (SLAM) has not yet fully adopted the newly arisen opportunities to create semantically meaningful maps. SLAM maps typically represent *geometric* information, but do not carry immediate object-level *semantic* information. Semantically-enriched SLAM systems are appealing because they increase the richness with which a robot can understand the world around it, and consequently the range and sophistication of interactions that robot may have with the world, a critical requirement for their eventual widespread deployment at workplaces and in homes.

Semantically meaningful maps should be object-oriented, with objects as the central entities of the map. *Quadrics*, i.e. 3D surfaces such as ellipsoids, are ideal landmark representations for object-oriented semantic maps. Quadrics have a very compact representation, can be manipulated efficiently within projective geometry, and capture information about the size, position, and orientation of an object.

The link between object detections and dual quadrics was recently investigated by [1], [2] and [3]. However, previous work utilized quadrics as a parametrization for landmark *mapping* only [2], was limited to an orthographic camera [1], or used an *algebraic* error that proved to be invalid when landmarks are only partially visible [3]. In this work we formulate a novel geometric error that is well-defined even when the observed object is only partially visible in the image. Furthermore, we investigates the utility of quadric based landmarks in a factor graph SLAM formulation that jointly estimates camera poses and quadric parameters from noisy odometry and object detection bounding boxes using a general perspective camera.

The authors gratefully thank John Skinner for his contributions to the evaluation environment. This research was conducted by the Australian Research Council Centre of Excellence for Robotic Vision (project number CE140100016). Michael Milford is supported by an Australian Research Council Future Fellowship (FT140101229). The authors are with the ARC Centre of Excellence for Robotic Vision, Queensland University of Technology (QUT), Brisbane, Australia.
Contact: lachlan.nicholson@hdr.qut.edu.au

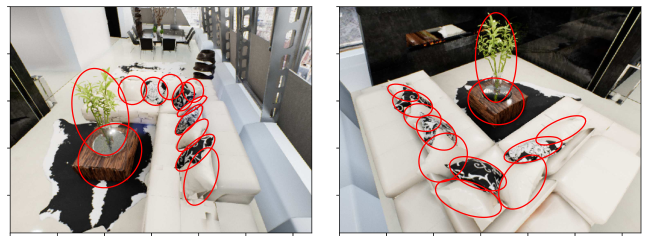

Fig. 1: QuadricSLAM uses *objects* as landmarks and represents them as constrained dual quadrics in 3D space. This figure illustrates how well the estimated quadrics fit the true objects when projected into the camera images from different viewpoints (red ellipses).

II. Dual Quadrics – Fundamental Concepts

A. Dual Quadrics

Quadrics are surfaces in 3D space that are defined by a 4×4 symmetric matrix $\mathbf{Q}$, so that all points $\mathbf{x}$ on the quadric fulfill $\mathbf{x}^\mathsf{T}\mathbf{Q}\mathbf{x} = 0$. Examples for quadrics are bodies such as spheres, ellipsoids, hyperboloids, cones, or cylinders.

A quadric can also be defined by a set of tangential *planes* such that the planes form an envelope around the quadric. This *dual* quadric $\mathbf{Q}^*$ is defined as $\boldsymbol{\pi}^\mathsf{T}\mathbf{Q}^*\boldsymbol{\pi} = 0$. Every quadric $\mathbf{Q}$ has a corresponding dual form $\mathbf{Q}^* = \mathrm{adjugate}(\mathbf{Q})$, or $\mathbf{Q}^* = \mathbf{Q}^{-1}$ if $\mathbf{Q}$ is invertible.

When a quadric is projected onto an image plane, it creates a dual *conic*, following the simple rule $\mathbf{C}^* = \mathbf{P}\mathbf{Q}^*\mathbf{P}^\mathsf{T}$. Here, $\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}]$ is the camera projection matrix that contains intrinsic and extrinsic camera parameters. Conics are the 2D counterparts of quadrics and form shapes such as circles, ellipses, parabolas, or hyperbolas. Just like quadrics, they can be defined in a primal form via points ($\mathbf{x}^\mathsf{T}\mathbf{C}\mathbf{x} = 0$), or in dual form using tangent lines: $\mathbf{l}^\mathsf{T}\mathbf{C}^*\mathbf{l} = 0$.

TABLE I: Comparison of the average RMSE errors for the trajectory and landmark position (cm), as well as the landmark shape and quality defined by the centered Jaccard distance and the standard Jaccard distance respectively.

	ATE_{trans}	LM_{trans}	LM_{shape}	$LM_{quality}$
Odometry	58.95	-	-	-
SVD solutiuon	-	57.86	0.61	0.85
QuadricSLAM	**20.49**	**17.14**	**0.44**	**0.59**

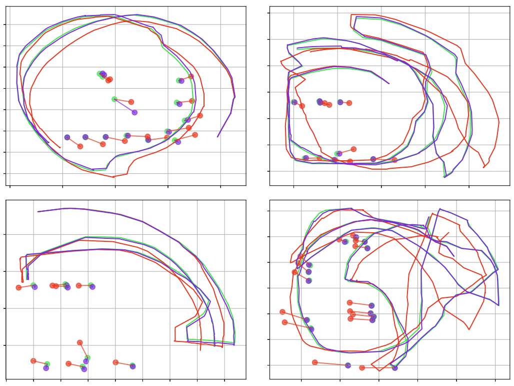

Fig. 2: Example trajectories and landmark centroids from a top-down perspective showing odometry (red) SLAM solution (blue) and ground truth (green).

B. Constrained Dual Quadric Parametrization

In its general form, a quadric or dual quadric can represent both closed surfaces such as spheres and ellipsoids and non-closed surfaces such as paraboloids or hyperboloids. As only the former are meaningful representations of object landmarks, we use a constrained dual quadric representation that ensures the represented surface is an ellipsoid or sphere. In the following, we compactly represent a constrained dual quadric with a 9-vector $\mathbf{q} = (\theta_1, \theta_2, \theta_3, t_1, t_2, t_3, s_1, s_2, s_3)^\mathsf{T}$ and reconstruct the full dual quadric $\mathbf{Q}^*$ as in [2].

III. Dual Quadrics as Landmarks in SLAM

In order to implement quadrics as SLAM landmarks, we define the geometric error as the difference between the observed detection $\mathbf{d}_{ij} = (x_{\text{min}}, y_{\text{min}}, x_{\text{max}}, y_{\text{max}})$ obtained by state-of-the-art object detection approaches such as [4], [5] and the predicted detection $\beta(\mathbf{x}_i, \mathbf{q}_j)$ of landmark $\mathbf{q}_j$ at camera pose $\mathbf{x}_i$. We calculate the predicted detection by projecting the estimated quadric into the image and calculating the on-image bounds of the resulting conic.

We can calculate the correct on-image conic bounding box by the following algorithm:

1) Find the four extrema points of the conic $\mathbf{C}$, i.e. the points $\{\mathbf{p}_1, ..., \mathbf{p}_4\}$ on the conic that maximise or minimise the x or y component respectively.
2) Find the up to 8 points $\{\mathbf{p}_5, ..., \mathbf{p}_{12}\}$ where the conic intersects the image boundaries.
3) Remove all non-real points and all points outside the image boundaries from the set $\mathcal{P} = \{\mathbf{p}_1, ..., \mathbf{p}_{12}\}$.
4) Find and return the maximum and minimum x and y coordinate components among the remaining points.

IV. Experiments and Evaluation

We implemented the SLAM problem as a factor graph where the robot poses and dual quadrics, X^* and Q^*, populated the latent variables of the graph, connected with odometry factors U and 2D bounding box factors D. We evaluate the resulting camera trajectory and landmark parameters in a high-fidelity simulation environment of 250 trajectories, comparing the odometry estimate, initial quadric solution, and SLAM solution to the ground truth camera trajectory and each objects 3D axis-aligned bounding boxes.

A. Results and Discussion

We summarize the results of our experiments in Table I and provide qualitative examples illustrating the improvement in the accuracy of the estimated quadric surfaces and camera trajectory in Figures 1 and 2 respectively.

The results show that quadric landmarks significantly improve the quality of the robot trajectory and the estimated map, providing accurate high level information about the shape and position of objects within the environment. Explicitly, QuadricSLAM gains a 65.2% improvement on trajectory error and a 70.4%, 26.7% and 30.6% improvement on landmark position, shape and quality.

V. Conclusions and Future Work

QuadricSLAM is a step towards integrating state-of-the-art object detection and SLAM systems in order to expand the range of applications for which we can deploy our robotic systems. Object based landmarks are essential to the development of meaningful object-oriented robotic maps.

Our paper has demonstrated how to use dual quadrics as landmark representations in SLAM with perspective cameras. We provide a method of parametrizing dual quadrics as closed surfaces and show how they can be directly constrained from bounding boxes as they originate from typical object detection systems. We develop a factor graph-based SLAM formulation that jointly estimates camera trajectory and object parameters in the presence of odometry noise, object detection noise, occlusion and partial object visibility. This has been achieved by defining a geometric error over the quadric and object detection that is robust to partial object observations. We provide an extensive evaluation of trajectory *and* landmark quality, demonstrating the utility of object-based landmarks for SLAM.

References

[1] M. Crocco, C. Rubino, and A. Del Bue, "Structure from motion with objects," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4141–4149.

[2] C. Rubino, M. Crocco, and A. Del Bue, "3d object localisation from multi-view image detections," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

[3] N. Sünderhauf and M. Milford, "Dual quadrics from object detection boundingboxes as landmark representations in slam," *arXiv preprint arXiv:1708.00965*, 2017.

[4] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems (NIPS)*, 2015, pp. 91–99.

[5] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.