Robotics and Autonomous Systems 62 (2014) 1657–1667

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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

Vistas and parallel tracking and mapping with Wall–Floor Features: Enabling autonomous flight in man-made environments

D.-N. Ta *, K. Ok, F. Dellaert

Center for Robotics and Intelligent Machines, Georgia Institute of Technology, Atlanta, GA, USA

HIGHLIGHTS

- We enable autonomous flight for a lightweight quad-rotor in indoor environments.
- Distant features, called vistas, are used for steering towards open spaces.
- Parallel tracking and mapping framework with odometry overcomes challenges in monoSLAM.
- Special Wall–Floor Features cope with feature-poor environments.
- Wall inference using Wall–Floor Features helps avoid lateral collisions.

ARTICLE INFO

Article history:
Available online 28 March 2014

Keywords:
Quadrotor
MAV
Autonomous navigation
Indoor
SLAM
Monocular
Parallel tracking and mapping
Vistas
Wall–Floor Features

ABSTRACT

We propose a solution towards the problem of autonomous flight in man-made indoor environments with a micro aerial vehicle (MAV), using a frontal camera, a downward-facing sonar, and odometry inputs. While steering an MAV towards distant features that we call *vistas*, we build a map of the environment in a parallel tracking and mapping fashion to infer the wall structure and avoid lateral collisions in real-time. Our framework overcomes the limitations of traditional monocular SLAM approaches that are prone to failure when operating in feature-poor environments and when the camera purely rotates. First, we overcome the common dependency on feature-rich environments by detecting *Wall–Floor Features (WFFs)*, a novel type of low-dimensional landmarks that are specifically designed for man-made environments to capture the geometric structure of the scene. We show that WFFs not only reveal the structure of the scene, but can also be tracked reliably. Second, we cope with difficult robot motions and environments by fusing the visual data with odometry measurements in a principled manner. This allows the robot to continue tracking when it purely rotates and when it temporarily navigates across a completely featureless environment. We demonstrate our results on a small commercially available quad-rotor platform flying in a typical feature-poor indoor environment.

Published by Elsevier B.V.

1. Introduction

We address the problem of vision-based autonomous navigation in man-made environments for Micro Aerial Vehicles (MAVs). We utilize a lightweight frontal camera, a downward-facing sonar for height measurements, and odometry inputs, while refraining from using heavy and power-hungry sensors that impose limitations on the MAVs. Research and applications for such lightweight MAVs have been growing significantly in recent years for both military and civilian services.

Man-made environments pose two key difficulties for vision-based methods. First, state-of-the-art monocular vision systems [1,2] rely on many corner-type features in the image to localize the camera and build a map. However, a typical indoor environment does not possess many distinct corner-type features, while the few that are present are mostly far-away, providing only little information to localize the camera reliably. Furthermore, these systems break down when the camera dominantly rotates towards an unknown region, due to the lack of motion parallax needed to triangulate new landmarks in the scene. Such rotation is common in robot navigation where the robot's frontal camera undergoes a yaw motion induced by a change in the robot's heading.

In this paper, we present a vision-based navigation system that steers the robot towards distant features, which we call *vistas*, while inferring the structure of the scene to avoid lateral collisions.

* Corresponding author. Tel.: +1 6174158387.
E-mail addresses: duynguyen@gatech.edu (D.-N. Ta), kyelok@gatech.edu (K. Ok), dellaert@cc.gatech.edu (F. Dellaert).

http://dx.doi.org/10.1016/j.robot.2014.03.010
0921-8890/Published by Elsevier B.V.

Such capability could not be achieved in previous vision-based MAVs, without dedicating additional sensors for this purpose (i.e. frontal and side-facing sonars [3]). We infer the scene structure by building a map of Wall–Floor Features, a special type of landmarks, that can cope with lack of corner-type features in man-made environments. Our map building system also overcomes the aforementioned limitations of monocular SLAM by fusing visual features with odometry inputs. We do this in a parallel tracking and mapping framework to achieve fast response to changes in the environment, while building a high-quality map of the scene. Contributions of our system, which is an extension of our previous work [4], are as follows (see Fig. 1):

Our first contribution is using vistas to determine the robot steering direction, enabling robust navigation. Our vistas are derived from first principles of what it means to be *distant*; hence, they are not hallway-specific like the previous work that depends on vanishing points detected from spurious edges [3] or hallway-specific cues [5]. Moreover, vistas are also derived from scale-space features and inherit the properties such that they are easily and reliably detected and tracked in many types of environments.

Our second contribution is a special type of landmarks, called *Wall–Floor Features*, that are suitable for mapping indoor environments and enabling autonomous exploration capabilities. In addition to vistas, for intelligent exploration schemes, the MAV needs some knowledge of the scene structure. We infer the structure from a map of compact and low-dimensional landmarks that are informative enough to capture the most important geometric information of the scene. Our novel landmarks lie on the perpendicular intersection of vertical lines on the wall and the horizontal floor plane. They encode the direction of the wall and can capture any type of corners whether straight, convex or concave.

Our third contribution is a parallel tracking and mapping framework that adopts the parallel-style execution of tracking and mapping introduced in the Parallel Tracking and Mapping (PTAM) system in [1], and extends its capabilities to deal with the lack of features in man-made environments and typical robot motions that fail monocular SLAM. While PTAM requires a large number of visual features to localize the camera, our system fuses visual features with odometry/IMU measurements to deal with periods of featureless scenes in the environment. Utilizing sensor fusion, our system can also deal with rotational motions towards unknown regions, where map building commonly fails due to the lack of motion parallax to triangulate new landmarks for the map. Unlike other PTAM–IMU fusion work that uses the actual PTAM system as a black box [6,7] and fails to prevent breakdowns for such cases, we incorporate the parallel-style execution in our own factor graph framework [8] to fuse sensor information in a principled manner; hence, our system gains robustness in these difficult situations.

We demonstrate our results on an inexpensive AR.Drone 2.0[1] quad-rotor. Our system successfully flies autonomously towards vistas while avoiding lateral collisions with wall inference. We evaluate our mapping results with loop closures on a pre-captured offline dataset using the ground-truth map, and compare its accuracy against PTAM and the quad-rotor's own estimate provided by the manufacturer's own firmware.

2. Related work

Successful MAV navigation systems neglect to address the power and payload limitations by using heavy and power-hungry sensors. For example, [9–14] use laser scanners to build a map of indoor environments for MAV navigation and exploration. Scherer et al. [15] use a laser scanner on a helicopter to detect and avoid different types of objects such as buildings, trees, and 6 mm wires in the city. Recently, Bry et al. [16] also use a laser scanner and a prebuilt map to enable aggressive flight and obstacle avoidance on a small fixed-wing airplane. These systems, however, are severely limited to short-term operations due to their heavy payload and high power usage. Furthermore, active sensors such as laser scanners are undesirable in many applications (e.g., military), due to the risk of cross-talk and ineligibility for covert operations. Therefore, we preclude the use of laser scanners and other heavy and power-hungry sensors.

Fig. 1. We use *vistas* (bottom left) to steer the robot and rely on a map of *Wall–Floor Features* (bottom right) in a parallel tracking and mapping framework to infer the scene structure and avoid lateral collisions. We present our results in an indoor setting using an AR.Drone (top).

Recent work in vision-based autonomous navigation neglects to provide exploration and planning capabilities achieved by building a map of the environment. While the system in [17] can navigate autonomously with a single camera, it relies on a prebuilt map. Bills et al. [3] enable autonomous navigation towards the end of the hallway by detecting the vanishing point from intersections of long lines along the corridors, but their system relies on supplementary sonar sensors to detect openings to the sides for autonomous exploration. Moreover, Murali and Birchfield [5] enable vision-based autonomous exploration on a ground robot by fusing many hallway-specific properties such as high entropy, symmetry, self-similarity, etc. to detect new hallway directions; however, their method is reactive and cannot support high-level planning algorithms. Based on map-building, our method can support exploration and planning algorithms to efficiently navigate towards unexplored regions.

On the other hand, state-of-the-art vision-based map-building methods are insufficient for usage in indoor navigation. Many methods build 3D point-cloud maps [18,7,19,20] but in textureless indoor environments, the point-clouds are too sparse to reveal the 3D structure needed for path/motion planning. Although some [21,22] build a map from the edges in the environment, they neglect to infer the environment structure crucial for robot navigation. Moreover, state-of-the-art vision-based methods that reconstruct the indoor scene [23–25] either rely on the indoor Manhattan world assumption or require expensive multi-hypothesis inference methods [24,25]. Our method based on

[1] http://ardrone.parrot.com.