Robotics and Autonomous Systems 100 (2018) 78–94

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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

Optical-flow based self-supervised learning of obstacle appearance applied to MAV landing

H.W. Ho [a,b,*], C. De Wagter [a], B.D.W. Remes [a], G.C.H.E. de Croon [a]

[a] *Micro Air Vehicle Laboratory of the Faculty of Aerospace Engineering, Delft University of Technology, The Netherlands*
[b] *School of Aerospace Engineering, Universiti Sains Malaysia, Malaysia*

ARTICLE INFO

Article history:
Received 17 August 2017
Accepted 6 October 2017
Available online 16 November 2017

Keywords:
Self-supervised learning
Aerial robotics
Bio-inspiration
Optical flow
Obstacle appearance
Autonomous landing

ABSTRACT

Monocular optical flow has been widely used to detect obstacles in Micro Air Vehicles (MAVs) during visual navigation. However, this approach requires significant movement, which reduces the efficiency of navigation and may even introduce risks in narrow spaces. In this paper, we introduce a novel setup of self-supervised learning (SSL), in which optical flow cues serve as a scaffold to learn the visual appearance of obstacles in the environment. We apply it to a landing task, in which initially 'surface roughness' is estimated from the optical flow field in order to detect obstacles. Subsequently, a linear regression function is learned that maps appearance features represented by texton distributions to the roughness estimate. After learning, the MAV can detect obstacles by just analyzing a still image. This allows the MAV to search for a landing spot without moving. We first demonstrate this principle to work with offline tests involving images captured from an on-board camera and then demonstrate the principle in flight. Although surface roughness is a property of the entire flow field in the global image, the appearance learning even allows for the pixel-wise segmentation of obstacles.

1. Introduction

To reduce the risk and cost of human intervention, autonomous flight of Micro Air Vehicles (MAVs) is highly desired in many circumstances. In particular, autonomous landing is a challenging, but essential task of the flight, as it needs to be done in a limited space and time [1]. Hence, quickly searching for a safe landing spot is required during landing in autonomous flights [2].

Many existing methods for finding a suitable landing spot use multiple cameras [3–5] or active sensors such as a laser range finder [6,7] to estimate the distance to many points on the landing surface. While both methods can provide accurate measurements, their perception range is limited, and they are heavy and costly for small MAVs. Therefore, use of a single camera is preferable as it is light-weight and has low power consumption [8].

State-of-the-art algorithms for autonomous landing purely rely on motion cues. There are two main approaches: (1) visual Simultaneous Localization and Mapping (SLAM) and (2) bio-inspired optical flow control.

The first approach can be categorized into feature-based and direct methods to solve the SLAM problems, i.e., to determine the vehicle's location and 3D-structure of the surrounding environment. The feature-based method [9,10] decouples these SLAM problems into the extraction of features and computation of camera pose and scene geometry based on tracking these features [11–13]. However, this approach is not sufficiently robust to challenging scenes where the features are hardly detected. The direct method tries to avoid this limitation by using image intensities directly to generate (semi-) dense maps [14–16]. Although the computational efficiency and accuracy of visual SLAM have been improved over the years [10,17–20], this approach still uses more computational resources than are strictly necessary.

The second approach is inspired by flying insects, which heavily rely on optical flow for navigation. Biologists first found that honeybees perform a grazing landing by keeping the ventral flow (lateral velocities divided by height) constant [21–24]. This approach guarantees a soft landing but does not control its vertical dynamics. To deal with that, recent studies proposed using time-to-contact (height divided by vertical velocity) [25–27]. The optical flow field can also be used during landing to identify and avoid obstacles [28,29].

For sensing obstacles with motion cues, either the vehicle or the obstacle obviously needs to move. This requirement is a drawback of motion cue approaches because it would be both safer and more

* Corresponding author at: Micro Air Vehicle laboratory of the Faculty of Aerospace Engineering, Delft University of Technology, The Netherlands.
E-mail addresses: H.W.Ho@tudelft.nl (H.W. Ho), C.deWagter@TUDelft.nl (C. De Wagter), B.D.W.Remes@TUDelft.nl (B.D.W. Remes), G.C.H.E.deCroon@TUDelft.nl (G.C.H.E. de Croon).

https://doi.org/10.1016/j.robot.2017.10.004
0921-8890/

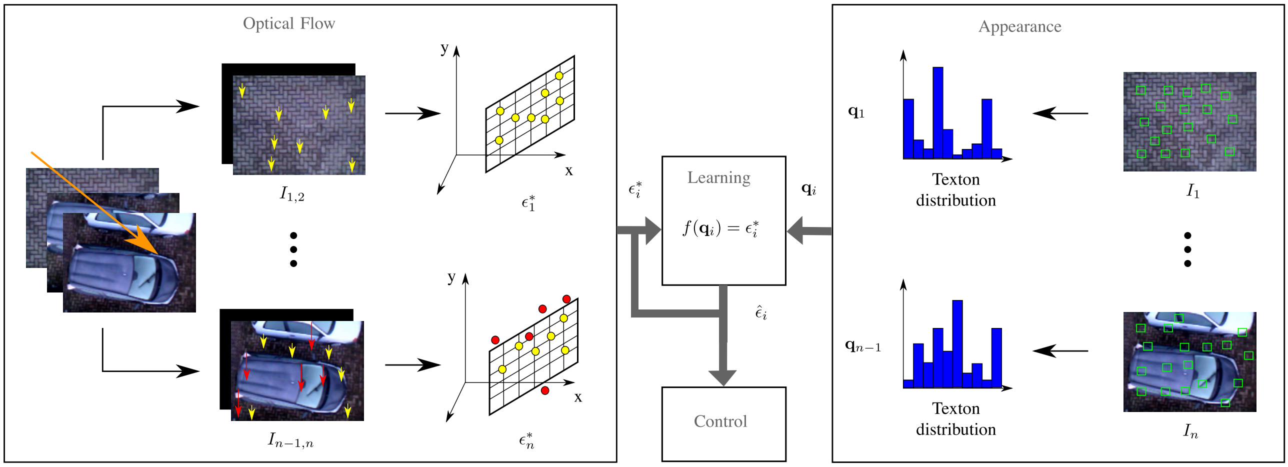

Fig. 1. Overview of the novel self-supervised learning (SSL) setup. The MAV starts flying using a roughness measure ϵ^* extracted from optical flow algorithm (left). A function is learned that maps appearance features, **q** from a *still* image to an estimate of the roughness measure $\hat{\epsilon}$ (right). After learning, the MAV can determine whether there are obstacles below based on a still image, allowing landing site selection in hover.

efficient for MAVs that have hovering capability to detect the obstacles underneath the vehicle in hover. This would require MAVs to exploit currently unused *appearance* cues — the information contained in still images. Human pilots are very able to identify potential landing sites in still images, by recognizing obstacles and flat landing areas in view. This is based on years of experience, to learn what obstacles look like.

In this paper, we propose an approach that allows MAVs to learn about the appearance of obstacles and suitable landing sites completely by themselves. The approach involves a novel setup of Self-Supervised Learning (SSL), in which motion cues provide the targets for the supervised learning of a function that maps appearance features to a surface roughness measure (see Fig. 1 for an overview). We showed the feasibility of the proposed SSL concept in [30]. The current article takes into account the influence of MAV height on obstacle detection with optical flow and significantly extends upon [30] by means of a systematic analysis with experiments. This includes threshold selection based on height information and experiments in which the learned appearance is used in the control loop and a study of the generalization of the learned appearance to various indoor and outdoor environments.

The remainder of the article is set up as follows: In Section 2, we discuss related work on self-supervised learning in more detail. In Section 3, we describe our proposed concept to learn the visual appearance of obstacles based on *surface roughness*, ϵ^* from optical flow. Section 4 presents the results of both optical flow- and appearance-based obstacle detection, and Section 5 explains generalization of our SSL approach to different environments. Then, Section 6 demonstrates landing experiments where the proposed algorithms run on-board an MAV. Finally, we draw conclusions in Section 7.

2. Related work

There are several remarkable achievements with SSL on autonomous driving cars, where stereo vision, laser scanners, or bumpers provided supervised outputs to learn the appearance of obstacles on the road [31–34]. In [31,32], a close-range map is generated by the laser scanners to identify a nearby patch of drivable surface. This patch is used to train appearance models to extend the road detection range. In [33], terrain classification obtained from 3D information with a stereo camera is used for training a convolutional neural network (ConvNet). The input image patches to the trained ConvNet need to be normalized based on the estimated distance so that the obstacles always have the same size. Then, the trained model can be used to detect obstacles beyond the range of a stereo camera. In [34], ground and non-ground regions in images can be segmented using radar and monocular vision. This is done by serving the estimate of the ground region from radar as the supervised output to learn visual appearance of the ground. After learning, the unmanned ground robot can detect drivable surface using monocular vision.

In [35], optical flow was used for tracing back the obstacles in time when they are far away. The supervised outputs were still provided by stereo vision and bumpers. A weakly-supervised approach was used to segment drivable paths of a road vehicle [36]. This approach first labeled the training images by estimating the vehicle motion using a stereo camera and detecting 3D objects using a laser scanner and then used these labeled images to train a deep semantic segmentation network. The trained network can provide path segmentations using images from a single camera for autonomous driving of a road vehicle. Another drivable path segmentations were achieved by using a ConvNet [37] or a fully convolutional network (FCN) [38] which was trained using images labeled with a stereo camera. A nonlinear regression based depth estimation method using only a monocular camera was used for MAVs to learn a deliberate scheme for navigating through a cluttered environment [39]. The training set was made using a stereo vision system on a ground robot and offboard image processing was done in the flight.

A major difference of the approach we propose and previous work on SSL is that optical flow from monocular vision is used for generating the supervised outputs. To the best of our knowledge, there is only one other SSL study that also used optical flow to provide the supervised outputs. The study in [40] used optical flow from a camera mounted on a car to learn a ground color model, assuming knowledge of the camera position relative to the ground. The learned ground color model aided in filtering optical flow vectors in order to improve the accuracy of the optical-flow based visual odometry.

In this article, we use the optical flow from a downward-looking camera mounted on an MAV to learn the appearance of obstacles. In contrast to [40], we intend for the MAV to be able to use the learned appearance of obstacles even in the absence of supervisory cue of optical flow. This leads to a fascinating extension of the MAV's autonomous flight capabilities: while the robot initially only uses motion cues, and hence needs to move significantly in order to see obstacles, after learning it is able to see obstacles without moving.