Robotics and Autonomous Systems 80 (2016) 1–10

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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

# Stereo visual odometry in urban environments based on detecting ground features

Arturo de la Escalera [a,*], Ebroul Izquierdo [b], David Martín [a], Basam Musleh [a], Fernando García [a], José María Armingol [a]

[a] *Universidad Carlos III de Madrid, Leganés, Spain*
[b] *Queen Mary University London, London, UK*

## HIGHLIGHTS

- A Visual Odometry method is proposed, which detects and tracks features on the surface of the ground.
- The use of a virtual bird-image assures a uniform 3D distribution of the features.
- Results in real urban environments show ability to follow the real trajectory drove by the vehicle.

## ARTICLE INFO

*Article history:*
Available online 25 March 2016

*Keywords:*
Autonomous vehicles
Visual odometry
Kalman filter

## ABSTRACT

Autonomous vehicles rely on the accurate estimation of their pose, speed and direction of travel to perform basic navigation tasks. Although GPSs are very useful, they have some drawbacks in urban applications that affect their accuracy. Visual odometry is an alternative or complementary method because provides the ego motion of the vehicle with enough accuracy and uses a sensor already available in some vehicles for other tasks, so no extra sensor is needed. In this paper, a new method is proposed that detects and tracks features available on the surface of the ground, due to the texture of the road or street and road markings. This way it is assured only static points are taken into account in order to obtain the relative movement between images. A Kalman filter improves the estimations and the Ackermann steering restriction is applied so the vehicle follows a constrained trajectory, which improves the camera displacement estimation obtained from a PnP algorithm. Some results and comparisons in real urban environments are shown in order to demonstrate the good performance of the algorithm. They show the method is able to estimate the linear and angular speeds of the vehicle with high accuracy as well as its ability to follow the real trajectory drove by the vehicle along long paths within a minimum error.

## 1. Introduction

Vehicle localization is a fundamental task in autonomous vehicle navigation. It relies on accurate estimation of pose, speed and direction of travel to achieve basic tasks including mapping, obstacles avoidance and path following. Nowadays, many autonomous vehicles rely on GPS-based systems for estimating their ego motion. Although GPSs are very useful, they have some drawbacks. The price of the equipment is still high for the centimeter accuracy needed for autonomous applications. Moreover, above all for urban applications, the shortcomings of the GPSs are clearer because there are some situations that affect their accuracy. For example, there may not be a direct line of sight to one or several satellites because of the presence of a building or a tree canopy. The urban canyon effect is very frequent within cities due to building heights. Finally, the vehicle has not available the GPS signal for an important task as driving along tunnels. Other sensors available are low-cost IMUs; however, although they are fast, they have a measurement bias and therefore need frequent corrections. Several solutions can be proposed to solve this problem, such as the use of maps or odometry provided by the vehicle wheels. The first one needs a continuous updating of the maps to be useful and the second lacks enough precision for several applications. That is why another sensor is needed and here is where digital cameras can play an important role. On one hand because, as it will be shown, they are useful for obtaining the vehicle's ego motion and, on the other hand, because nowadays they are already used for other tasks such

---

[*] Corresponding author.
*E-mail address:* escalera@ing.uc3m.es (A. de la Escalera).

http://dx.doi.org/10.1016/j.robot.2016.03.004
0921-8890/

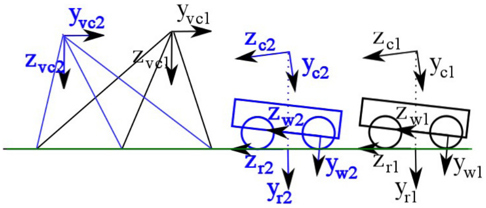

**Fig. 1.** Different reference axes used by the visual odometry system (best view in color).

as pedestrian, traffic sign or road lane detection [1], accordingly it is a sensor that can be applied for multiple assignments. Visual Odometry (VO) estimates the ego motion of a camera or a set of cameras mounted on a vehicle using only the visual information provided by it or them. The term is related to the wheel odometry used in robotics and was formulated in 2004 by Nister [2]. Usually, VO algorithms have three steps:

1. Detect features or points of interest (POI) in every image and match the ones found in two consecutive ones.
2. Find and remove the wrong matches.
3. Estimate the relative movement of the cameras.

This can be done using monocular or stereo cameras and assuming planar or non-planar motion models. A tutorial on VO can be found in [3,4]. In [5] a stereo system is presented, where it estimates the rigid body motion that best describes the transformation among the sets of 3D points acquired in consecutive frames. Optical flow and stereo disparity are computed to minimize the re-projection error of tracked feature points. Instead of performing this task using only consecutive frames, they use the whole history of the tracked features to compute the motion of the camera. The camera motion is estimated in [6] using a quaternion and RANSAC [7] for outlier removal and a Two-stage Local Binocular Bundle Adjustment for optimizing the results. In [8] the rotation and translation between consecutive poses are obtained, minimizing the distance of the correspondent point projections. They take into account that farther 3D points have higher uncertainty, RANSAC for outliers and they constrain pose estimation taking temporal flow into account. A persistent map containing 3D landmarks localized in a global frame is presented in [9]. They automatically distinguish some frames, used to update the landmark map, which serves for ego-localization. The other frames are used to track the landmarks and to localize the camera with respect to the map. In [10], they apply some monocular techniques to stereo visual odometry system. The features are detected using FAST [11], described with BRIEF [12] and tracked during the image sequence. A P3P algorithm is used for the pose estimation and local bundle adjustment is used for result refinement. Other sensors, like lasers, has been used in [13–15]. In [16], the authors combine visual and lidar odometry. Visual odometry is useful to estimate the ego-motion and as a help to register point clouds from a scanning lidar, which refines the motion estimation.

Urban environments are highly dynamic, so the case of a static scene cannot be assumed. Moreover, these surroundings are highly cluttered with frequent occlusions. Consequently, there are some specific difficulties any method has to face:

- The detected POI can belong to moving objects and, as a consequence, the camera motion estimation would be erroneous if they are used for obtaining the camera displacement.
- Due to ego motion and occlusions, some detected POI in an image are not detected in the next one, and vice versa, but this can lead to an erroneous matching and, again, to an erroneous motion estimation.

The novelty of the proposed algorithm is related to the previous difficulties. This paper is an extended version of the one presented at the Second Iberian Robotics Conference, Robot2015, in Lisbon, Portugal [17]. Here, a more detailed explanation of the algorithm is given, a comparison with other methods is presented as well as the experiments, which number has been increased. The overall of the proposed approach can be seen in Fig. 1. Due to the two difficulties for urban environments mentioned before, in this proposal, points of interests belonging to the road are going to be detected and matched, as they belong to the static part of the scene. In order to do so, first, the road ahead of the vehicle, which is assumed flat up to 20 m, is obtained. Unlike other approaches that assume that only the yaw angle changes, estimations of the camera roll, pitch and height are obtained for every image. This way the extrinsic parameters of the stereo camera for every image are found. As the road is flat and the pose and orientation of the cameras are known, any virtual image of the road can be obtained. In this case, a virtual bird-view image, perpendicular to the road, where the features are going to be detected, is created. The POI are detected due to the texture of the road or street and to the presence of road markings available on the surface of the ground. Matching the features of two consecutive images, the relative movement of the vehicles is found. A Kalman filter improves the estimations and the Ackermann steering restriction is applied so the vehicle follows a constrained trajectory.

The rest of the paper describes the algorithm. The features are going to be detected in a virtual bird-view image. In order to do this, Section 2 explains how the extrinsic parameters of the stereo camera are obtained for every image. Section 3 explains how the features are matched and the relative movement of consecutive images is found. The Kalman filter is explained in Section 4 and the results in real driving situations are shown in Section 5. Finally, the conclusions are presented.

The results are based on sequences of the KITTI Vision Benchmark Suite [18,19]. The stereo cameras for this benchmark were placed on the vehicle roof and parallel to the ground. Because of the camera placement, the minimum distance that the cameras capture is a bit far for this method, 6 m. Because the presented algorithm looks for features on the road, the cameras are not placed on the best place, on the vehicle's wind-shield and looking at the road. Another limitation is the resolution, 1344 by 391 pixels, so the images are a bit narrow. But as the sequences have a very good ground truth obtained from a centimeter GPS, they are very useful to show if the method is valid or not. Others specification of the cameras are the stereo baseline is 60 cm and the frame rate is 10 Hz.

## 2. Continuous extrinsic parameters estimation

The first step of the algorithm is to find the extrinsic parameters of the stereo camera. Other approaches find the initial position of the cameras and assume that only the yaw angle changes. Although this is valid for several domains, it is not practical in urban applications due to the change in the extrinsic parameters because of the vehicle movements, the effect of the shock absorbers and the presence of uneven road surfaces. The road is assumed to be flat up to a near distance, 20 m, and the plane of the road is found. From the plane coefficients, the values of the pitch and roll angles and the height of the camera are obtained. Besides the application for visual odometry, finding the plane is also useful for other tasks of the vehicle as obstacle and drivable area detection.

### 2.1. Obtaining the 3D point cloud

As shown in Fig. 2, the changes in illumination inside the images, the lack of texture in many objects and the presence of repetitive patterns in others are the three main problems in