Available online at www.sciencedirect.com
ScienceDirect
IFAC-PapersOnLine 49-15 (2016) 094–099

Prediction of the scene quality for stereo vision-based autonomous navigation

Hélène Roggeman * **Julien Marzat** *
Anthelme Bernard-Brunel * **Guy Le Besnerais** *

* *ONERA – The French Aerospace Lab, F-91123, Palaiseau, France.*
e-mail: firstname.lastname@onera.fr.

Abstract: This paper presents an autonomous navigation architecture for a robot using stereo vision-based localisation. The main contribution is the prediction of the quality of future localisation of the system in order to detect and avoid areas where vision-based localisation may fail, due to lack of texture in the scene. A criterion based on the estimation of future visible landmarks, considering uncertainties on landmarks and camera positions, is integrated in a Model Predictive Control loop to compute safe trajectories with respect to the visual localisation. The system was tested on a mobile robot and the obtained results demonstrate the effectiveness of our method.

Keywords:
Autonomous vehicles, Robot navigation, Stereo vision, Model Predictive Control

1. INTRODUCTION

This work takes place in the context of autonomous navigation in unknown indoor environments with Micro Air Vehicles or mobile robots. In this kind of environment, there are no global localisation systems available such as GPS. Vision sensors are then a usual solution for estimating the localisation of a vehicle. Using visual landmarks in the images, a visual odometry algorithm can be used to compute the position and orientation of the camera and thus of the system. However, if the scene contains too few landmarks, for instance when the robot faces a white wall, the algorithm cannot extract enough landmarks to compute the localisation accurately. In this case, the mission is likely to fail. In this work, we aim at ensuring that the system remains well-localised during the mission time. To reach this objective, we have developed a criterion that predicts the quality of the scene for localisation, using already known landmarks. We have then integrated this criterion into a Model Predictive Control (MPC) loop so as to demonstrate its efficiency on a real robot.

1.1 Related work

In the literature, some authors were interested in the ability to predict the quality of future measurements to improve navigation algorithms, (see Makarenko and et al. (2002), Bourgaul et al. (2002), Vidal-Calleja et al. (2006), Sim and Roy (2005)) or environment reconstruction algorithms (Forster et al. (2014), Dunn et al. (2009)). Some authors propose a criterion based on the Shannon entropy (see Bourgaul et al. (2002), Bachrach et al. (2012), Sim and Roy (2005)). Vidal-Calleja et al. (2006) choose the optimal control to reduce uncertainties on the camera and landmark positions using a criterion based on the mutual information. Forster et al. (2014) use the information gain to choose the trajectory that maximizes the precision of the environment reconstruction. Other approaches rely on a criterion based on image or scene geometry. Dunn et al. (2009) look for a Next-Best-View, considering uncertainties on measurements and scene appearance. Sadat et al. (2014) compute a texture criterion from the local density of triangles in the 3D mesh used for environment reconstruction. This criterion is used for planning trajectories with RRT* in real time towards a desired goal. Mostegel et al. (2014) propose a criterion which accounts for the geometric quality of landmarks (triangulation angle, proof of existence) and for the ability of recognizing each point in order to improve waypoint navigation. In all these references, the authors use a monocular camera, whereas, in our case, we use a stereo rig. It allows us to directly access the depth information for each point. Moreover, unlike most of these references, we take into account uncertainties on the landmark positions and the camera position.

1.2 Overview

The basic mission considered in this work is to perform autonomous navigation between waypoints with a mobile robot. The environment can have relatively textureless areas that are a problem for the visual localisation. This is why the robot has to detect these zones in order to choose safe trajectories to navigate. To locate itself, the robot is equipped with a stereo rig, composed of two fixed cameras with known calibration. The stereo images received from the rig are used to compute the pose with a visual odometry algorithm. The main steps of such an algorithm are:

- Interest points, like Harris (Harris and Stephens (1988)), are extracted from both images and matched.
- The 3D positions of the corresponding points in the scene are computed by triangulation.

2405-8963 © 2016, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2016.07.715

- The motion of the left camera is estimated from the apparent displacement of the projections of the 3D points in the images.

In this work, we have choosen to use eVO, a visual odometry algorithm described in Sanfourche et al. (2013). Other stereo algorithms could be considered as well, e.g. Klein and Murray (2007). A criterion is proposed to evaluate the quality of the scene that the robot will encounter in a future position. It uses the 3D points sent by the visual odometry algorithm to predict the visible points in the images in the future, considering the uncertainties on the landmark positions and on the movement of the system. The criterion is explained in Section 2. We have set up experiments to demonstrate the relevance of this criterion, which are described in Section 2.5. A navigation strategy based on MPC is presented in Section 3. It integrates our localisation quality criterion with waypoint navigation and an obstacle avoidance criterion. Experimental results of navigation are shown in Section 4.

2. LOCALISATION QUALITY CRITERION

Let us first introduce some notations. $Y = (x, y, z)^T$ is a 3D point, expressed in a global frame $(\mathcal{W})$, unless otherwise stated, defined by the camera position and orientation at the beginning of the mission. The camera frame $(\mathcal{C}_n)$ is the frame defined by the current camera position and orientation. At t_0, $(\mathcal{W})$ and $(\mathcal{C}_n)$ are coincident. For a vector v, $\tilde{v} = (v, 1)^T$ is the augmented vector.

2.1 Projection of a 3D point

Given a 3D point computed by the visual odometry algorithm and a desired camera position, the 2D projection of the 3D point in the corresponding image is computed. First, the 3D point is expressed in the camera frame, then, the projection is computed.

Change of basis T is the translation vector and R is the rotation matrix from camera frame to global frame. The new coordinates of the 3D point Y, denoted Y', are

$$\tilde{Y'} = P^{-1}\tilde{Y} = \begin{pmatrix} R & T \\ 0 & 1 \end{pmatrix}^{-1} \tilde{Y} \quad (1)$$

Camera model Let $p = (u, v)^T$ denote the image projection point of a 3D point $Y = (x, y, z)^T$, expressed in the camera frame. α is the focal length and (u_0, v_0) are the optical center coordinates.

$$\tilde{p} = z^{-1}KY = z^{-1} \begin{pmatrix} \alpha & 0 & u_0 \\ 0 & \alpha & v_0 \\ 0 & 0 & 1 \end{pmatrix} Y \quad (2)$$

2.2 Uncertainty computation

The goal is to estimate the uncertainty on the position of a projected point on the image after a camera displacement, denoted with rotation and translation (R, T). The rotation matrix is written as

$$R = R_z(\theta_z)R_y(\theta_y)R_x(\theta_x) \quad (3)$$

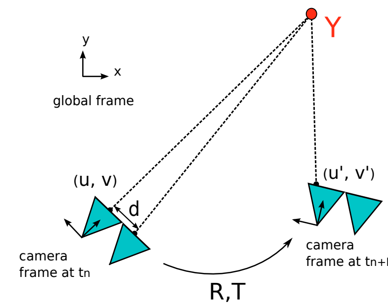

Fig. 1. Computing the 2D projection of a point in a future camera image: (u, v, d) and (R, T) are known. Y is triangulated using (u, v, d). (u', v') is deduced from Y and (R, T).

where $R_\bullet$ is the rotation about the $\bullet$ axis by an angle $\theta_\bullet$, and the translation vector is written

$$T = (t_x, t_y, t_z)^T \quad (4)$$

$\Theta = (\theta_x, \theta_y, \theta_z, t_x, t_y, t_z)$ is the vector of the displacement parameters and (u, v, d), the projected point position in the left camera and disparity. The uncertainties on the triangulation of the 3D point and on the estimation of the camera position after the displacement are taken into account. We model the uncertainties on Θ and (u, v, d) by a normal distribution with zero mean. The covariance matrices are denoted by Σ_Θ and $\Sigma_{u,v,d}$. Moreover, the uncertainties between Θ and (u, v, d) are assumed to be independent. As shown in Figure 1, $p = (u, v)$ is the projected point in the image in the first position and p' the projected point in the image after the displacement of the camera.

Expression of p' To express p' as a function f of (Θ, u, v, d), we use a triangulation function denoted by Π^{-1}, the change of basis (R, T) and a projection function denoted by Π:

$$p' = f(\Theta, u, v, d)$$
$$p' = \Pi(Y'(\Theta, u, v, d)) = \Pi\left(R(\Theta) \cdot \Pi^{-1}(u, v, d) + T(\Theta)\right) \quad (5)$$

Covariance of p' As the uncertainties between Θ and (u, v, d) are independent, the covariance on the position of p' can be written as

$$\Sigma_{p'} = J_{f_\Theta} \cdot \Sigma_\Theta \cdot J_{f_\Theta}^T + J_{f_{u,v,d}} \cdot \Sigma_{u,v,d} \cdot J_{f_{u,v,d}}^T \quad (6)$$

with J_{f_Θ} and $J_{f_{u,v,d}}$ the Jacobian matrices of f with respect to Θ and (u, v, d), respectively.

$$J_{f_\Theta} = \frac{\partial f}{\partial \Theta} = J_\Pi(Y') \cdot J_{Y'}(Y) \quad (7)$$

$$J_{f_{u,v,d}} = \frac{\partial f}{\partial u, v, d} = J_\Pi(Y') \cdot R \cdot J_{\Pi^{-1}}(u, v, d) \quad (8)$$

J_Π, $J_{\Pi^{-1}}$ and $J_{Y'}$ are the Jacobian matrices of the projection function, the triangulation function and the change of basis function, described in the following paragraphs.

Jacobian matrix of the triangulation function The 3D position of a point is computed with the coordinates of the projected point in the image and the disparity.