Visual path following using only monocular vision for urban environments

Albert Diosi, Fabien Spindler, Anthony Remazeilles, Siniša Šegvić and François Chaumette

***Abstract*— This document provides a summary to a short video with the same title. The video shows the French intelligent transportation vehicle CyCab performing visual path following using only monocular vision. All phases of the process are shown with a spoken commentary. In the teaching phase, the user drives the robot manually while images from the camera are stored. Key images with corresponding images features are stored as a map together with 2D and 3D local information. In the navigation phase, CyCab follows the learned path by tracking the images features projected from the map and with a simple visual servoing control law.**

I. INTRODUCTION

Autonomous transportation vehicles will likely play an important role in the future. The video corresponding to this summarizing text shows a system capable of learning and executing a path using only one perspective camera.

In the video a CyCab, a French-made 4 wheel drive, 4 wheel steered intelligent vehicle designed to carry 2 passengers is shown. In our CyCab all computations except the low-level control are carried out on a laptop with a 2GHz Centrino processor. A 70° field of view, forward looking, B&W Allied Vision Marlin (F-131B) camera is mounted on the robot at a 65cm height. The camera is used in its auto shutter mode, while the image are scaled down to 320x240.

II. DETAILS ON VISUAL NAVIGATION

This section briefly describes the implemented visual navigation framework. The teaching of the robot i.e. the mapping of the environment is described first, followed by the description of the navigation process consisting of localization and robot control.

A. Mapping

Learning a path (i.e. mapping) starts with the manual driving of the robot on a reference path while processing (or storing for off-line mapping) the images from the robot's camera. From the images an internal representation of the path is created, as summarized in fig. 1. The mapping starts with finding Harris points in the first image, initializing a Kanade-Lucas-Tomasi (KLT) feature tracker and by saving the first image as the first reference image. The KLT tracker was modified to compensate for changes in the illumination. In the next step a new image is acquired and the tracked features are updated. The tracking of features which appear different than in the previous reference image is abandoned.

The presented work has been performed within the French national project Predit Mobivip and project Robea Bodega.

The authors are with IRISA/INRIA Rennes, Campus Beaulieu, 35042 Rennes cedex, France. Email: firstname.lastname@irisa.fr

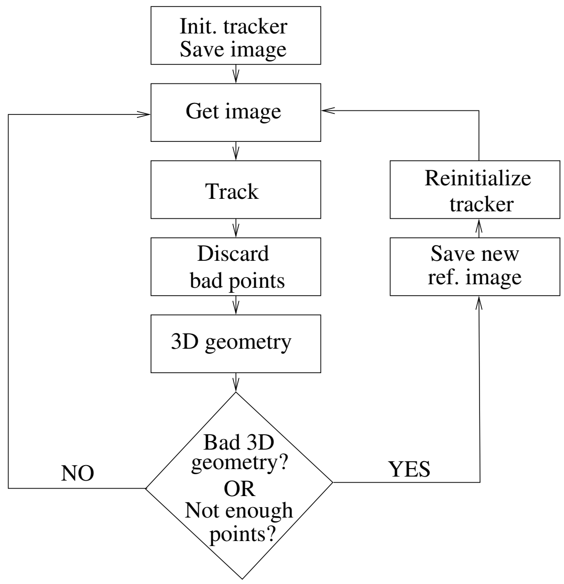

Fig. 1. The steps involved in building a representation of a path from a sequence of images, i.e. mapping.

The rest of the features are then used to estimate the 3D geometry between the previous reference and the current image. In the 3D geometry estimation, the essential matrix is recovered using a calibrated algorithm in a random sampling framework. If the 3D reconstruction error is low and there are enough tracked features a new image is acquired. Otherwise the current image is saved as the next reference image. The relative pose of the current image with respect to the previous reference image and the 2D and 3D coordinates of the point features shared with the previous reference image are also saved. Then the tracker is reinitialized with new Harris points added to the old ones and the processing loop continues with acquiring a new image.

The resulting map (fig. 2) is used during autonomous navigation in the localization module to provide stable image points for image-based visual servoing.

B. Localization

The localization process during navigation is depicted in fig. 3. The navigation process is started with the user selecting a reference image close to the robot's current location. Then an image is acquired and matched to the selected reference image. The matching is done using Lowe's SIFT descriptors. The estimation of the camera pose using the matched points enables to project map points from the reference image into the current image. The projected points

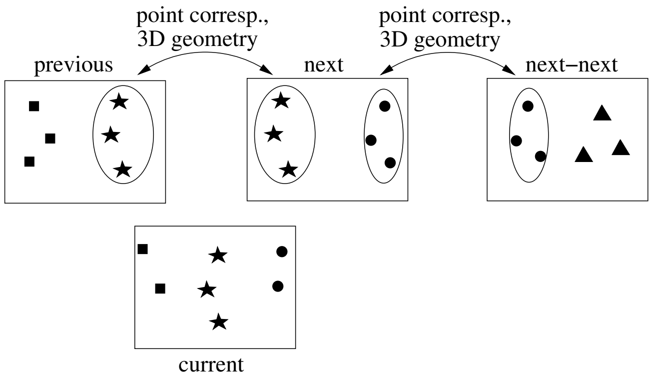

Fig. 2. The map consists of reference images, 2D and 3D information. During navigation, the point features from the map are projected into the current image and tracked.

are then used to initialize a KLT tracker. Next, a new image is acquired and the point positions are updated by the tracker. Using the tracked points a three-view geometry calculation is performed between the previous reference, current and next reference image (fig. 2). If the current image is found to precede the next reference image, then points from the map are reprojected into the current image. The projected points are used to resume the tracking of points currently not tracked and to stop the tracking of points which are far from their projections. A new image is acquired next and the whole cycle continues with tracking. However, if it is found that the current image comes after the next reference image, a topological transition is made i.e. the next-next reference image becomes the next reference image. The tracker is then reinitialized with points from the map and the process continues with acquiring a new image.

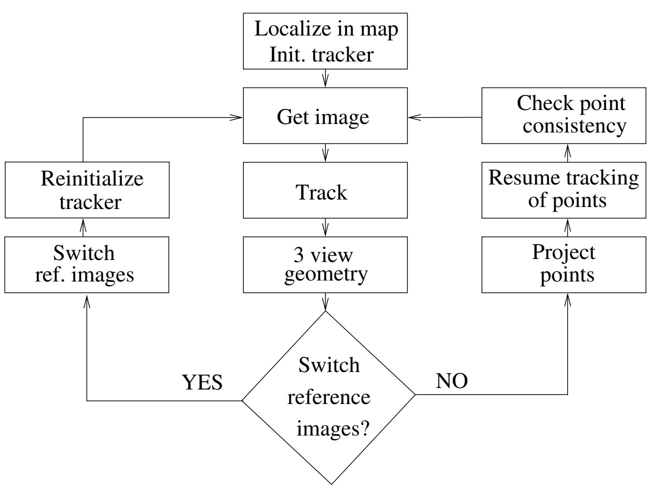

Fig. 3. Visual localization during navigation.

C. Motion Control

In the motion control scheme the robot is not required to accurately reach each reference image of the path, nor to follow accurately the learned path since that may not be useful during navigation. In practice, the exact motion of the robot should be controlled by an obstacle avoidance module which we plan to implement soon. Therefore a simple control algorithm was implemented where the difference in the x-coordinates (assuming the forward facing camera's horizontal axis is orthogonal with the axis of robot rotation) of the centroid of features in the current and next reference image are fed back into the motion controller of the robot as steering angle.

The translational velocity is set to a constant value, except during turns, where it is reduced (to a smaller constant value) to ease the tracking of quickly moving features in the image.

III. ADDITIONAL INFORMATION

The concept of the framework has been evaluated using simulations in [4], while the feature tracker and the complete vision subsystem have been described in [6], [7]. An experimental evaluation of the navigation framework can be found in [2].

IV. RELATED WORK

In [5] a CyCab follows a prerecorded trajectory using a camera with a fisheye lens. Unlike in our work, an accurate 3D model of the path is created using a dense set of reference images with global using bundle adjustment. After the scale is corrected using odometry measurements the robot accurately follows the reference path using pose based control.

In [1] a robot navigated outdoors with 2D image information only. During mapping, image features were tracked and their image patches together with their x image coordinates were saved. During navigation, the robot control was based on simple rules applied to the tracked feature coordinates in the next reference and current image.

The work described in [3] aimed at indoor navigation, can deal with occlusion of features using 3D information. A local 3D reconstruction is done between two reference omnidirectional images. During navigation, tracked features which have been occluded get reprojected back into the current image. The recovered pose of the robot is used to guide the robot towards the target image.

REFERENCES

[1] Z. Chen and S. T. Birchfield. Qualitative vision-based mobile robot navigation. In *ICRA*, Orlando, 2006.
[2] A. Diosi, S Remazeilles, A.and Segvic, and F. Chaumette. Outdoor visual path following experiments. Submitted to IROS'07.
[3] T. Goedeme, T. Tuytelaars, G. Vanacker, M. Nuttin, and L. Van Gool. Feature based omnidirectional sparse visual path following. In *IROS*, Edmonton, Canada, August 2005.
[4] A. Remazeilles, P. Gros, and F. Chaumette. 3D navigation based on a visual memory. In *ICRA'06*, 2006.
[5] E. Royer, J. Bom, M. Dhome, B Thuillot, M. Lhuillier, and F. Marmoiton. Outdoor autonomous navigation using monocular vision. In *IROS*, pages 3395–3400, Edmonton, Canada, August 2005.
[6] S. Segvic, A. Remazeilles, and F. Chaumette. Enhancing the point feature tracker by adaptive modelling of the feature support. In *ECCV*, Graz, Austria, 2006.
[7] S. Segvic, A. Remazeilles, A. Diosi, and F. Chaumette. Large scale vision based navigation without an accurate global reconstruction. In *Accepted to CVPR'07*, Minneapolis, June 2007.