VISUAL ODOMETRY CORRECTION BASED ON LOOP CLOSURE DETECTION

L. CARAMAZANA, R. ARROYO and L. M. BERGASA

Robesafe Research Group, Department of Electronics, University of Alcalá (UAH). E-mail contact: lidia.caramazana.zarzosa@gmail.com

An essential requirement in the fields of robotics and intelligent transportation systems is to know the position of a mobile robot along the time, as well as the trajectory that it describes by using on-board sensors. In this paper, we propose a novel approach focused on the use of cameras as perception sensors for visual localization in unknown environments. Our system allows to perform a robust visual odometry, where correction algorithms based on loop closure detection are applied for correctly identifying the location of a robot in long-term situations. In order to satisfy the previous conditions, we carry out a methodological improvement of some classic computer vision techniques. In addition, new algorithms are implemented with the aim of compensating the drift produced in the visual odometry calculation along the traversed path. According to this, our main goal is to obtain an accurate estimation of the position, orientation and trajectory followed by an autonomous vehicle. Sequences of images acquired by an on-board stereo camera system are analyzed without any previous knowledge about the real environment. Several results obtained from these sequences are presented to demonstrate the benefits of our proposal.

1 Introduction

In recent years, the estimation of the pose of an autonomous robot or vehicle using computer vision techniques has become a topic of a great interest in the robotics community. This is due to the improvements in cameras features and to their reduced costs with respect to other sensors traditionally used for localization tasks, such as GPS, IMU, range-based or ultrasounds, among others. Besides, the proliferation of visual SLAM systems (Bailey et al., 2006) has extended the application of camera-based approaches for determining the global location of a mobile robot in an unknown environment.

In this context, visual odometry (Nister et al., 2004) has the goal of estimating the position and orientation of a robot or vehicle by analyzing an image sequence acquired by cameras without previous information about locations. Each pair of images is considered to match their keypoints and calculate the translation and rotation between two poses of the vehicle. Unfortunately, visual odometry typically accumulates a drift when long periods of time are taken into account. This problem makes that the localization tasks could not be completely reliable in these cases. For this reason, in extended trajectories the information provided by standard visual odometry algorithms gives errors in long-term conditions.

According to the previous considerations, in this work we propose a novel approach based on loop closure detection using ABLE (Arroyo et al., 2014) for correcting the drift in visual odometry, which is initially processed by means of the LIBVISO library (Kitt et al., 2010). With the aim of solving the problems related to the drift, our system recognizes revisited places and recalculates a corrected pose. We contribute a method that uses this information to estimate the deviation between the revisited pose and the previous one. In order to validate our proposal, image sequences from the publicly available KITTI dataset (Geiger et al., 2013) are employed.

2 Method for Visual Odometry: The LIBVISO Algorithm

The visual odometry algorithm provided by LIBVISO allows to determine the six degrees of freedom (rotation and translation) in a visual localization system. In our work, stereo cameras are employed in image acquisition. Due to this, intrinsic and extrinsic camera parameters are needed to correctly perform the matching between the stereo images. In our case, the tests performed in the KITTI dataset are carried out using the specific camera parameters defined in (Geiger et al., 2013). The application of a stereo camera approach provides a higher robustness to our global system, because it avoids the scale ambiguities that are common when monocular cameras are used for visual odometry computation.

The methodology behind our implementation derived from LIBVISO is mainly based on a trifocal geometry between the images. Initially, some keypoints are detected and their main features are extracted and matched for each two consecutive pair of images, as shown in the example presented in Fig. 1. Taking into account the obtained matches, the movement of the autonomous robot or vehicle is estimated by processing a trifocal tensor that associates the keypoints between three frames of a same static scene.

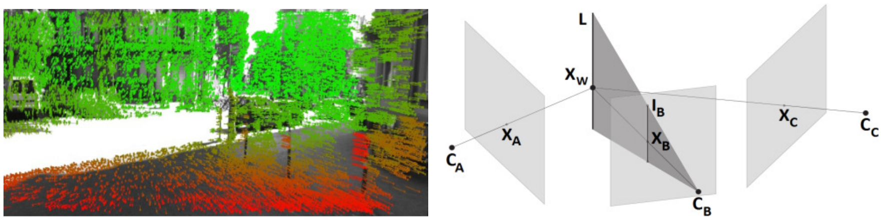


Fig. 1. A representation of the movement estimated over an example image using visual odometry, jointly with a diagram of the applied trifocal tensor.

In addition, the implementation of LIBVISO detects outliers using RANSAC (Scaramuzza et al., 2009), which allows to reject the atypical values obtained by erroneous matches and to improve the odometry results with respect to schemes without this filtering technique. However, this is not sufficient to avoid the drift along the time, as will be evidenced in the section of results. For this reason, we contribute a more robust approach based on a refined correction of the poses using loop closure information.

3 Method for Loop Closure Detection: The ABLE Algorithm

Some previous studies recently carried out by our research group in the topic of visual loop closure detection (Arroyo et al., 2014) are now applied to correct the drift derived from the previous visual odometry computation stage. The developed method for identifying when a place is revisited is named ABLE (Able for Binary-appearance Loop-closure Evaluation).

The main goal of this algorithm is to visually describe places in order to give similarity measurements between them for elucidating if a loop closure exists or not. Typically, ABLE computes global LDB binary features (Yang et al., 2012) for image description. In this case, disparity information obtained from the stereo images is also added to the descriptor. Apart from this, a variant of the description method initially designed in

ABLE is contributed in this paper, where the recently proposed AKAZE features (Alcantarilla et al., 2013) are tested as core of the global description approach instead of LDB. We implement it to evaluate the robustness and efficiency of AKAZE, which adds gradient information in a nonlinear space to obtain a description invariant to scale and rotation.

After describing the images, the binary features ($\mathbf{d}$) computed for each frame are matched to see if they are similar enough to consider a revisited place. In the case of binary descriptors such as LDB or AKAZE, the Hamming distance can be applied in matching, which provides a great efficiency, because it consists on a simple XOR operation ($\oplus$) followed by a basic sum of bits, as formulated in Equation (1). The obtained similarity values are stored on a distance matrix (M). These values are used to detect the loop closures when high similarity measurements are obtained.

$$M(i,j) = M(j,i) = \text{bitsum}(\mathbf{d}(i) \oplus \mathbf{d}(j)) \tag{1}$$

4 Our Proposal for Visual Odometry Correction

The information about the loops identified in the previous system stage is now used to correct the visual odometry. Here, we contribute the formulation of our method to perform these corrections. After a revisited place is detected in a specific frame, the drift of the pose currently estimated by the visual odometry algorithm is compensated by taking into account the pose obtained when the place was previously traversed. In this case, we consider corrections for the plane *xz*, where the deviation (Δ) between the current pose (i) and the previous one (j) is calculated as follows:

$$\Delta x(i) = |x(i) - x(j)| \tag{2}$$

$$\Delta z(i) = |z(i) - z(j)| \tag{3}$$

Then, the current poses are updated ($x(i)'$, $z(i)'$) in the x and z axes using the previously estimated deviation:

$$x(i)' = x(i) + \Delta x(i) \tag{4}$$

$$z(i)' = z(i) + \Delta z(i) \tag{5}$$

Besides, an average deviation (Δx_{avg}, Δz_{avg}) is subsequently computed after detecting the first pose corresponding to a loop closure. This information is employed to correct the poses in the rest of the trajectory, where m is the number of processed frames:

$$\Delta x_{avg} = \frac{\sum_{i=1}^{m} \Delta x(i)}{m} \tag{6}$$

$$\Delta z_{avg} = \frac{\sum_{i=1}^{m} \Delta z(i)}{m} \tag{7}$$

After calculating the average deviations in the loop zone, the poses in the remaining frames are updated according to the following equations:

$$x(i)' = x(i) + \Delta x_{avg} \tag{8}$$

$$z(i)' = z(i) + \Delta z_{avg} \tag{9}$$

The application of the formulated corrections in poses improves the accuracy initially obtained by only using a visual odometry without consider the progressive drift, as corroborated in the next section of results.

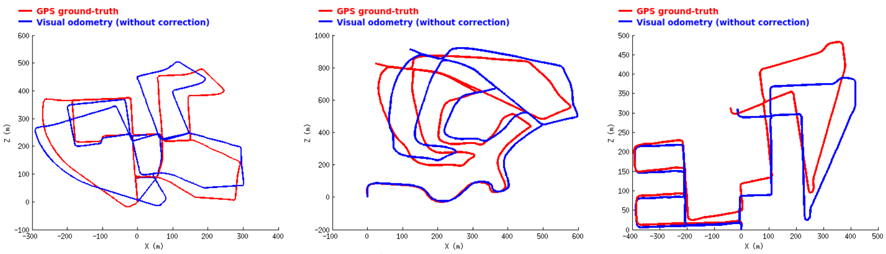
5 Evaluation and Main Results

Our proposal is evaluated in the KITTI Odometry dataset (Geiger et al., 2013). It contains 22 sequences recorded on different car routes around Karlsruhe (Germany). GPS ground-truth measurements are available. A ground-truth for loop closure was also defined in (Arroyo et al., 2014).

In Fig. 2, it can be seen how the visual odometry measurements obtained by LIBVISO without correction have a considerable drift with respect to the GPS ground-truth. The maps presented correspond to some significant sequences from the KITTI dataset, which are also presented in Fig. 3 to show the matches of the loop closures detected using ABLE.

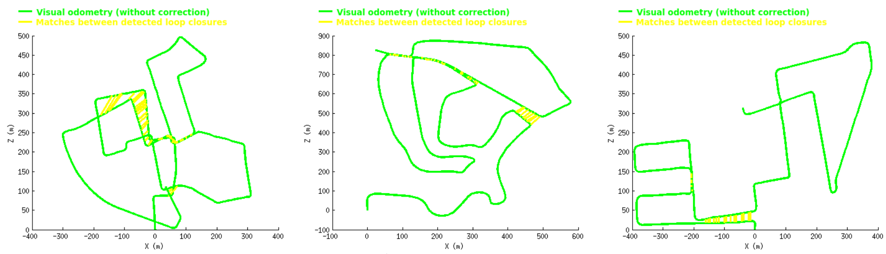
In addition, Fig. 4 depicts some examples of distance matrices processed by means of ABLE using LDB and AKAZE descriptors as core. The detected loop closures correspond to the diagonals in the matrices. Besides, Fig. 5 introduces precision-recall results about ABLE performance in loop closure detection depending on the descriptor used as core. Apart from LDB and AKAZE, we also test other typical descriptors such as HOG (Dalal et al., 2005), SURF (Bay et al., 2008), BRIEF (Calonder et al., 2010) and ORB (Rublee et al., 2011). These results demonstrate the better performance of LDB and the new approach based on AKAZE.

Finally, Fig. 6 evidences the better accuracy of our proposal based on a visual odometry with loop closure corrections, where it can be seen how the drift is reduced with respect to the original visual odometry.



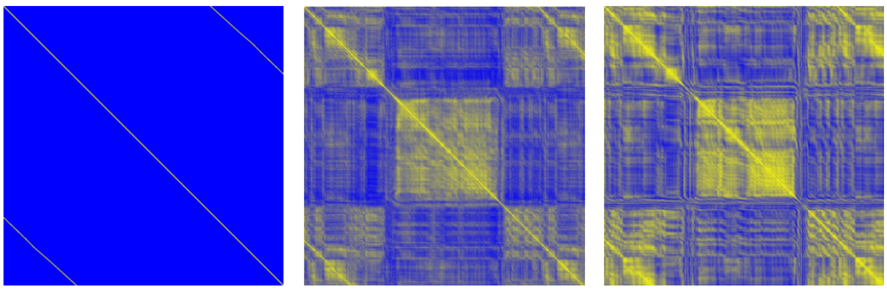
(a) Sequence 00 (b) Sequence 02 (c) Sequence 08

Fig. 2. Results for visual odometry without correction in the KITTI dataset.



(a) Sequence 00 (b) Sequence 02 (c) Sequence 08

Fig. 3. Results for loop closure detection in the KITTI dataset.



(a) M ground-truth (b) M using LDB (c) M using AKAZE

Fig. 4. Examples of distance matrices from the Sequence 06 of the KITTI dataset.

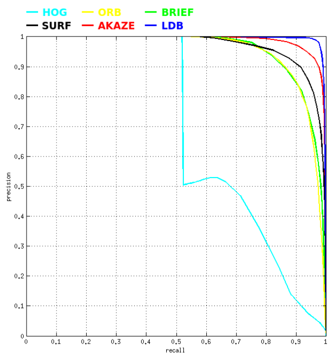
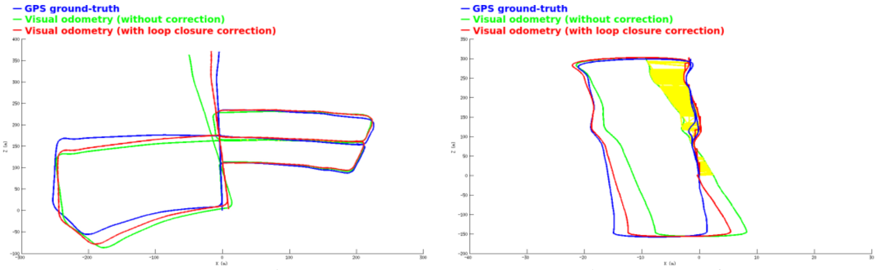


Fig. 5. Precision-recall curves obtained for the Sequence 00 of the KITTI dataset.



(a) Sequence 05 (b) Sequence 06

Fig. 6. Results for visual odometry with loop correction in the KITTI dataset.

6 Conclusions and Future Works

In this work, we have defined and validated our system based on a robust visual odometry estimation using loop closure corrections. The results exposed along the paper demonstrate the benefits of this proposal, such as the visible reduction of the progressive drift accumulated along the time.

The contributions presented can be divided into three main areas. First of all, the implementation of the initial stereo visual odometry system based on LIBVISO. Secondly, the application of ABLE for loop closure detection, including a new approach based on AKAZE features. And finally, the formulation of a complete method for correcting the visual odometry estimations using the information about the loop closures detected.

In future works, we plan to improve our visual odometry model using optimizations based on algorithms such as Levenberg-Marquardt.

Acknowledgements

This research has received funding from the RoboCity2030-III-CM project (S2013/MIT-2748), supported by the CAM and Structural Funds of the EU.

References

(Alcantarilla et al., 2013) Alcantarilla, P. F., Nuevo, J., and Bartoli, A., 2013, “Fast explicit difusion for accelerated features in nonlinear scale spaces," in British Machine Vision Conference (BMVC), pp. 1-13.

(Arroyo et al., 2014) Arroyo, R., Alcantarilla, P. F., Bergasa, L. M., Yebes, J. J., and Bronte, S., 2014, “Fast and effective visual place recognition using binary codes and disparity information," in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 3089-3094.

(Bailey et al., 2006) Bailey, T., and Durrant-Whyte, H., 2006, “Simultaneous Localisation And Mapping (SLAM): Part II State of the art,” IEEE Robotics and Automation Magazine (RAM), 13(3): 108-117.

(Bay et al., 2008) Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L., 2008, "Speeded-Up Robust Features (SURF)," Computer Vision and Image Understanding (CVIU), 110(3): 346-359.

(Calonder et al., 2010) Calonder, M., Lepetit, V., Strecha, C., and Fua, P., 2010, “BRIEF: Binary Robust Independent Elementary Features," in European Conference on Computer Vision (ECCV), pp. 778-792.

(Dalal et al., 2005) Dalal, N., and Triggs, B., 2005, "Histograms of Oriented Gradients for human detection," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2, pp. 886-893.

(Geiger et al., 2013) Geiger, A., Lenz, P., Stiller, C., and Urtasun, R., 2013, "Vision meets robotics: The KITTI dataset," International Journal of Robotics Research (IJRR), 32(11): 1231-1237.

(Kitt et al., 2010) Kitt, B., Geiger, A., and Lategahn, H., 2010, “Visual odometry based on stereo image sequences with RANSAC-based outlier rejection scheme," in IEEE Intelligent Vehicles Symp. (IV), pp. 486-492.

(Nister et al., 2004) Nister, D., Naroditsky, O., and Bergen, J., 2004, "Visual odometry," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 652-659.

(Rublee et al., 2011) Rublee, E., Rabaud, V., Konolige, K., and Bradski, G., 2011, "ORB: An efficient alternative to SIFT or SURF," in International Conference on Computer Vision (ICCV), pp. 2564-2571.

(Scaramuzza et al., 2009) Scaramuzza, D., Fraundorfer, F., and Siegwart, R., 2009, “Real-time monocular visual odometry for on-road vehicles with 1-point RANSAC,” in IEEE International Conference on Robotics and Automation (ICRA), pp. 4293-4299.

(Yang et al., 2012) Yang, X., and Cheng, K. T., 2012, “LDB: An ultra-fast feature for scalable augmented reality on mobile devices," in International Symposium on Mixed and Augmented Reality (ISMAR), pp. 49-57.