Journal of Zhejiang University SCIENCE A
ISSN 1673-565X (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Science Letters:

Visual odometry for road vehicles—feasibility analysis

SOTELO Miguel-angel[†1], GARCÍA Roberto[1], PARRA Ignacio[1], FERNÁNDEZ David[1], GAVILÁN Miguel[1], ÁLVAREZ Sergio[1], NARANJO José-eugenio[2]

([1]*Department of Electronics, University of Alcalá, Alcalá de Henares 28871, Spain*)

([2]*Department of Informatics, Industrial Automation Institute, CSIC, Madrid, Spain*)

[†]E-mail: sotelo@depeca.uah.es

Received Aug. 24, 2007; revision accepted Oct. 30, 2007

Abstract: Estimating the global position of a road vehicle without using GPS is a challenge that many scientists look forward to solving in the near future. Normally, inertial and odometry sensors are used to complement GPS measures in an attempt to provide a means for maintaining vehicle odometry during GPS outage. Nonetheless, recent experiments have demonstrated that computer vision can also be used as a valuable source to provide what can be denoted as visual odometry. For this purpose, vehicle motion can be estimated using a non-linear, photogrametric approach based on RAndom SAmple Consensus (RANSAC). The results prove that the detection and selection of relevant feature points is a crucial factor in the global performance of the visual odometry algorithm. The key issues for further improvement are discussed in this letter.

Key words: 3D visual odometry, Ego-motion estimation, RAndom SAmple Consensus (RANSAC), Photogrametric approach
doi:10.1631/jzus.2007.A2017 **Document code:** A **CLC number:** TP391.4

INTRODUCTION

Accurate estimation of the vehicle global position is a key issue, not only for developing useful driver assistance systems, but also for achieving autonomous driving. The proposed challenge is to use stereo-vision capabilities to provide accurate estimation of the vehicle ego-motion with regard to the road, and thus to compute the vehicle global position. This is becoming more and more tractable to implement on standard PC-based systems nowadays. However, there are still open issues that constitute a challenge in achieving highly robust ego-motion estimation in real traffic conditions. These are discussed in the following lines:

(1) There must exist stationary reference objects that can be seen from the cameras position.

(2) Information contained on road scenes can be divided into road feature points and background feature points.

(3) Typical road scenes may contain a large amount of outlier information. This includes non-stationary objects such as moving vehicles, pedestrians, and car wipers. All these artifacts contribute to false measurements for ego-motion estimation.

In this paper, we investigated a method for ego-motion computing based on stereo-vision. The idea of estimating displacements from two 3D frames using stereo vision has been previously used in (Nister *et al.*, 2004; Hagnelius, 2005). A common factor of these works is the use of robust estimation and outliers rejection using RAndom SAmple Consensus (RANSAC) (Fischler and Bolles, 1981). In (Nister *et al.*, 2004), a so-called firewall mechanism is implemented in order to reset the system to remove cumulative error. Both monocular and stereo-based versions of visual odometry were developed in (Nister *et al.*, 2004). In (Agrawal and Konolige, 2006) a stereo system composed of two wide Field of View cameras was installed on a mobile robot together with a GPS receiver and classical encoders. The system was tested in outdoor scenarios on different runs under 150 m. In the present work, the solution of the non-linear system equations describing the vehicle motion at each frame is computed under the non-linear, photogrametric approach using RANSAC. A clear contribution of this work is the optimization of the RANSAC parameters. Exhaustive experimentation was conducted in this aspect in order to yield the really optimal RANSAC parameters. Indeed, a genetic algorithm was off-line run to set a comparison between the optimized RANSAC parameters achieved on-line by our method and the same parameters obtained off-line by an evolutionary algorithm performing exhaustive global search. The results were extremely similar.

FEATURES DETECTION AND MATCHING

In each frame, Harris corners are detected, since this type of point feature has been found to yield detections that are relatively stable under small to moderate image distortions (Schmid *et al.*, 2000). In order to reduce the computation time and to remove irrelevant features that move in the central part of the image, the method is only run in the lower left and right parts of the image, where significant features are most frequently located. The feature points are matched at each frame, using the left and right images of the stereo-vision arrangement, and between pairs of frames. Features are detected in all frames and matches are allowed only between features. The details about the matching method can be found in (Alonso *et al.*, 2007), where a similar method was used for pedestrian detection purpose.

PROBLEM STATEMENT

The problem of estimating the trajectory followed by a moving vehicle can be defined as that of determining at frame i the rotation matrix $\boldsymbol{R}_{i-1,i}$ and the translational vector $\boldsymbol{T}_{i-1,i}$ that characterize the relative vehicle movement between two consecutive frames. The use of non-linear methods becomes necessary since the 9 elements of the rotation matrix cannot be considered individually (the rotation matrix has to be orthonormal). Indeed, there are only 3 unconstrained, independent parameters, i.e., the three rotation angles θ_x, θ_y, and θ_z, respectively. The system's rotation can be expressed by means of the rotation matrix $\boldsymbol{R}$ given by

$$\boldsymbol{R}=\begin{pmatrix} cycz & sxsycz+cxsz & -cxsycz+sxsz \\ -cysz & -sxsysz+cxcz & cxsysz+sxcz \\ sy & -sxcy & cxcy \end{pmatrix}, \quad (1)$$

where $ci=\cos\theta_i$ and $si=\sin\theta_i$ for $i=x, y, z$. The estimation of the rotation angles must be undertaken by using an iterative, least squares-based algorithm. Otherwise, the linear approach can lead to a non-realistic solution where the rotation matrix is not orthonormal.

RANSAC

RANSAC (RAndom SAmple Consensus) is an alternative to modifying the generative model to have heavier tails to search the collection of data points S for good points that reject points containing large errors, namely "outliers". RANSAC is used in this work to estimate the rotation matrix $\boldsymbol{R}$ and the translational vector $\boldsymbol{T}$ that characterize the relative movement of a vehicle between two consecutive frames. The input data to the algorithm are the 3D coordinates of the selected points at times t and $t+1$. After drawing samples from three points, the models that best fit to the input data are estimated using non-linear least squares. Then, a distance function is defined to classify the rest of points as inliers or outliers depending on threshold D_t. In this case, the distance function is the square error e between the sample and the predicted model. In this work, the following parameters were chosen: $D_t=0.005$ m^2, $N=10$ (maximum number of trials), $TH=0.8$ (consensus threshold). Detailed explanations about how these parameters are obtained can be found in (García *et al.*, 2007).

NON-LINEAR LEAST SQUARES

Given a system of n non-linear equations containing p variables the solution for $n<p$ does not form a vectorial subspace in general. Its structure depends on the nature of the equations. For $n=p$ a finite set of solutions exists instead of a unique solution as in the linear case. To solve this problem, an underdetermined system is built ($n>p$) in which a given error function $E(x)$ must be minimized. The error function can exhibit several local minima, although in general there is a single global minimum. Unfortunately, there is no numerical method that can assure the obtaining of such global minimum, except for the case of polynomial functions. Iterative methods based on the gradient descent can find a global minimum whenever the starting point meets certain conditions. By using non-linear least squares the process is in reality linearized following the tangent linearization approach. Formally, the functions describing the system equations can be approximated using the first term of Taylor's series expansion. After linearization, an overdetermined linear system of n equations and p variables has been constructed ($n<p$). The system can be solved using iterative non-linear least squares. At each iteration k of the regression method the linearized system is solved (given the 3D coordinates of N points in two consecutive frames). The vector of independent terms $\boldsymbol{c}$ is computed at iteration k. Once the Jacobean matrix $\boldsymbol{J}$ of the non-linear equations and vector $\boldsymbol{c}$ have been computed at iteration k, the non-linear system is finally solved. On completion of the process the algorithm yields the final solution $\boldsymbol{w}=[\theta_x, \theta_y, \theta_z, t_x, t_y, t_z]^{\mathrm{T}}$ that describes the relative vehicle movement between two consecutive iterations at t_0 and t_1, respectively. The termination condition is given by a minimum value of error or a maximum number of iterations. The complete mathematical details of this operation are provided in (García *et al.*, 2007).

DATA POST-PROCESSING

This is the last stage of the algorithm. Some partial estimations are discarded, in an attempt to remove as many outliers as possible. For this purpose, high root mean square error e estimations are removed and meaningless rotation angles estimations (non physically feasible) are discarded. Accordingly, a maximum value of e has been set to 0.5. Similarly, a maximum rotation angle threshold is used to discard meaningless rotation estimations. In such cases, the estimated vehicle motion is maintained according to motion estimated in the previous frame. Removing false rotation estimations is a key aspect in visual odometry systems since false rotation estimations lead to high cumulative errors.

IMPLEMENTATION AND RESULTS

The visual odometry system described in this paper has been implemented on a Pentium IV at 1.7 GHz running Linux Knoppix 3.7. A stereo vision platform based on Fire-i cameras (IEEE1394) was installed on a prototype vehicle. Several sequences were recorded in different locations including Alcalá de Henares and Arganda del Rey in Madrid (Spain). All sequences correspond to real traffic conditions in urban environments. In the experiments, the vehicle was driven bellow the maximum allowed velocity in cities, i.e., 50 km/h. A real experiment is graphically documented in this section as illustrated in Fig.1. The

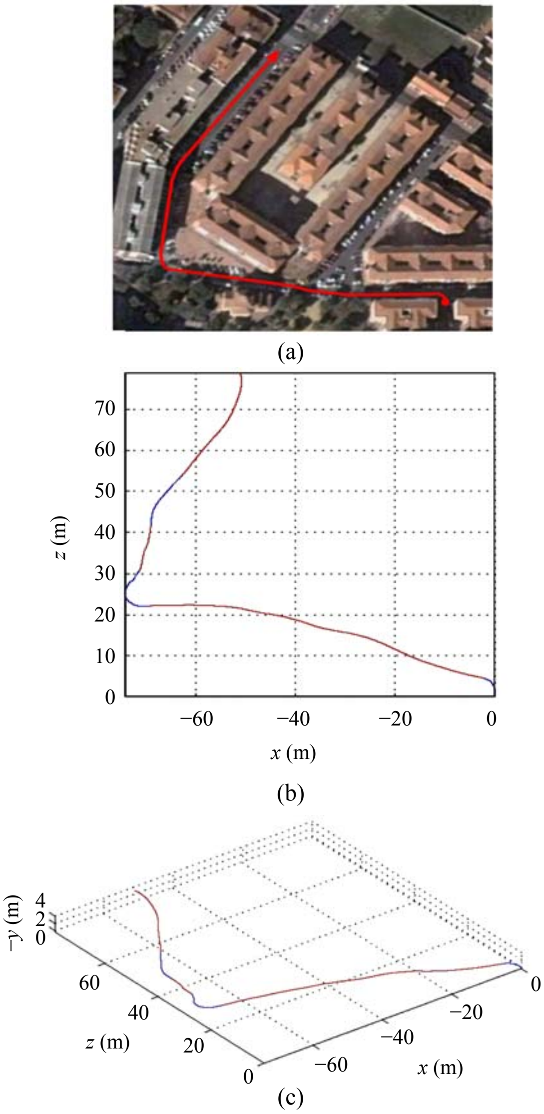

Fig.1 (a) Aerial view of the trajectory followed in the experiment; (b) Estimated 2D trajectory; (c) Estimated 3D trajectory

car ran a real distance of 276.79 m while the system estimated a distance of 272.52 m, which can be considered as quite an accurate visual estimation. Nonetheless, the 3D visual odometry method yields an altitude change of 2 m during the car's run, which is not a realistic figure since the trajectory described by the vehicle is almost planar. The whole sequence lasted 50.83 s and was analyzed by the system in 53.19 s including acquisition time. It can then be stated that algorithm execution at frame rate is practically preserved.

CONCLUSION AND FUTURE WORK

After observation of the results obtained in this work, it can be stated that the 3D visual odometry method described in this paper provides approximate trajectory estimations that can be useful for enhancing GPS accuracy, or even for substituting GPS in short outage periods. Nonetheless, the system provides estimations that exhibit cumulative errors. A major conclusion of this feasibility analysis is that it cannot be realistically expected that a 3D visual odometry system be used as a standalone method for global positioning applications. Apart from this obvious fact, other problems arise especially in altitude estimation. The reason for this stems from the fact that estimations of pitch and roll angles become complex using visual means, since variations of these angles in usual car displacements are really small and difficult to measure in the 2D image plane. These difficulties produce a non-real altitude change in estimated 3D trajectories. Another problem arises when features corresponding to non-stationary objects are detected and used by the system. As part of our future work we envision to develop a method for discriminating stationary points from those which are moving in the scene. Moving points can correspond to pedestrians or other vehicles circulating in the same area. Vehicle motion estimation will mainly rely on stationary points. Another important aspect for future improvement is the feature detection stage. For this purpose, the robustness of the process will be enhanced by incorporating temporal tracking of invariant-to-rotation features. This will allow for removing false features, i.e. glitches and features that correspond to unreal moving objects, such as, for instance, the edge caused by the intersection of a moving car and a tree. Finally, we will try to minimize the drift using sparse bundle adjustment.

References

Agrawal, M., Konolige, K., 2006. Real-time Localization in Outdoor Environments Using Stereo Vision and Inexpensive GPS. 18th Int. Conf. on Pattern Recognition, p.1063-1068. [doi:10.1109/ICPR.2006.962]

Alonso, I.P., Llorca, D.F., Sotelo, M.A., Bergasa, L.M., de Toro, P.R., Nuevo, J., Ocana, M., Garrido, M.A.G., 2007. Combination of feature extraction methods for SVM pedestrian detection. *IEEE Trans. on Intell. Transport. Syst.*, **8**(2):292-307. [doi:10.1109/TITS.2007.894194]

Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, **24**(6):381-395. [doi:10.1145/358669.358692]

García, R., Sotelo, M., Parra, I., Fernández, D., Gavilán, M., Álvarez, S., Naranjo, J.E., 2007. Visual odometry for road vehicles. *J. Intell. Rob. Syst.*, in press. [10.1007/s10846-007-9182-5]

Hagnelius, A., 2005. Visual Odometry. Master Thesis, Umea University.

Nister, D., Naroditsky, O., Beren, J., 2004. Visual Odometry. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, **1**:652-659. [doi:10.1109/CVPR.2004.1315094]

Schmid, C., Mohr, R., Bauckhage, C., 2000. Evaluation of interest point detectors. *Int. J. Computer Vision*, **37**(2): 151-172. [doi:10.1023/A:1008199403446]