15th IAPR International Conference on Machine Vision Applications (MVA)
Nagoya University, Nagoya, Japan, May 8-12, 2017.

Field Tests on Flat Ground of an Intensity-Difference Based Monocular Visual Odometry Algorithm for Planetary Rovers

Geovanni Martinez
Image Processing and Computer Vision Research Laboratory (IPCV-LAB)
Escuela de Ingeniería Eléctrica, Universidad de Costa Rica
11501-2060 San José, Costa Rica
geovanni.martinez@ucr.ac.cr

Abstract

In this contribution, the experimental results of testing a monocular visual odometry algorithm in a real rover platform over flat terrain for localization in outdoor sunlit conditions are presented. The algorithm computes the three-dimensional (3D) position of the rover by integrating its motion over time. The motion is directly estimated by maximizing a likelihood function that is the natural logarithm of the conditional probability of intensity differences measured at different observation points between consecutive images. It does not requiere as an intermediate step to determine the optical flow or establish correspondences. The images are captured by a monocular video camera that has been mounted on the rover looking to one side tilted downwards to the planet's surface. Most of the experiments were conducted under severe global illumination changes. Comparisons with ground truth data have shown an average absolute position error of 0.9% *of distance traveled with an average processing time per image of 0.06 seconds.*

1 Introduction

In order to improve the safety and autonomous navigation accuracy of planetary rovers [1], such as the Mars Exploration rover Opportunity and the Mars Science Laboratory's rover Curiosity, in slippery environments, after moving a small amount, the rover is often commanded to perform the correction of any error, which occurred because of wheel slippage, by using the rover's position estimate that is determined by a feature based stereo visual odometry algorithm [2]. This algorithm estimates the rover's motion tracking feature points over a sequence of image pairs, which are captured by a stereo camera, and integrating the estimated motion over time to obtain the rover's position. It was initially described in [3], then it was further developed in [4], until a real-time version of it was implemented and incorporated in the rovers Spirit and Opportunity of the Mars Exploration Rover Mission [2]. A more robust and faster updated version of it is currently being used in the Curiosity rover [5]. There are other similar algorithms in the scientific literature [6, 7, 8], which have even been adapted for to operate with a monocular [8] or an omnidirectional video camera [9], and recently, extended to Simultaneous Localization and Mapping (SLAM) [10]. Refer to [11] for a comprehensive tutorial on visual odometry.

In [12], a monocular visual odometry algorithm based on intensity differences was proposed as an alternative to the long-established feature based stereo visual odometry algorithms, which avoids having to track feature points for motion estimation, tasks that are known to be very difficult, to consume a lot of processing time and are prone to match errors due to large motions, occlusions or ambiguities, which greatly affect the 3D motion estimation [4]. With this algorithm it is possible to estimate the 3D motion of the rover by means of the maximization of the conditional probability of the intensity differences measured at key observation points between two successive images. The images are taken by a single video camera rigidly attached to the rover. The key observation points are image points whose linear intensity gradients are found to be high.

Despite that in [12] the above intensity-difference based monocular visual odometry algorithm has been extensively tested with synthetic data, an experimental validation of the algorithm in a real rover platform in outdoor sunlit conditions is still missing. This paper's main contribution will be to provide the results of the first outdoor experiments towards validation of the algorithm, which will be obtained for now on surfaces of little geometrical complexity such as flat terrain, to help to clarify whether the algorithm really does what is intended to do in real outdoors situations under severe global illumination changes.

This contribution is organized as follows: in section 2, the monocular visual odometry algorithm is briefly described; in section 3, the experimental results are presented; and finally, in section 4, a summary and the conclusions are given.

2 Monocular visual odometry algorithm

Here the algorithm for estimating the rover's 3D motion from two consecutive intensity images I_{k-1} and I_k will be briefly presented (see [12] for a more detailed description). The images depict part of the planet's surface next to the rover and are taken by a single video camera with coordinate system (q, r, s) at time t_{k-1} and time t_k, where the camera coordinate system and robot coordinate system are supposed to be the same. The camera has been mounted on the rover looking to one side tilted downwards 37 degrees and the images are supposed to be formed through perspective projection with focal length f onto a camera plane with coordinate system (x, y), where the focal lens f is set according to pre-calibration results obtained using the Tsai algorithm [13]. The 3D shape of a rectangular portion of the surface part that is being captured by the camera is assumed to be flat and rigid and described by meshing together two triangles, forming a rectangle with coordinate system (X, Y, Z).

This 3D shape and its relative pose to the camera coordinate system (q, r, s) are supposed to known at time t_{k-1}. The pose is described by a set of six parameters: the three components of a 3D position vector and three rotation angles.

A set of N observation points are also supposed to be known at time t_{k-1}. An observation point lies on one of the two triangles at barycentric coordinates $\mathbf{A}_v$ with respect to the corresponding triangle's vertex 3D positions and carries the intensity value I and the linear intensity gradients $\mathbf{g} = (g_x, g_y)^\top$ at surface position $\mathbf{A}_v$. Let $\mathbf{A} = (A_q, A_r, A_s)^\top$ be the corresponding 3D coordinates of the observation point with respect to the camera coordinate system, where the component A_s represents its depth. These shape, pose and observation points are referred here as the surface model at time t_{k-1}. The surface model at time t_{k-1} is obtained by moving (rotating and translating) the surface model from its pose at time t_{k-2} to the corresponding pose at time t_{k-1} with the negative of the rover's 3D motion estimates from time t_{k-2} to time t_{k-1}.

The surface model at time t_0 is created and initialized in three steps. By the first step, the surface model's shape is initialized as a flat and rigid mesh of two triangles forming a rectangle with dimensions $40x30cm^2$, whose coordinate system (X, Y, Z) is placed at the upper left corner. By the second step, the pose of the initial surface model with respect to the camera coordinate system (q, r, s) is estimated by applying the Tsai's coplanar camera calibration algorithm [13]. The pattern is removed from the scene after calibration is performed. The camera calibration also ensures metric motion estimates. In the third step, after surface model's shape and pose initialization, the observation points are created and initialized. The initial observation points are selected as the image points with high linear intensity gradients ($|\mathbf{g}| > \delta_1$) in the first image I_0. This selection rule will reduce the influence of the camera noise and increase the accuracy of the estimation. The value of the threshold δ_1 was heuristically set to 12 and remains constant throughout the experiments. After an observation point at image position $\mathbf{a}$ has been selected, its 3D position $\mathbf{A}$ is computed with respect to the camera coordinate system as the intersection of the $\mathbf{a}$'s line of sight and the plane containing the corresponding triangle of the surface model's shape at time t_0. Then, its barycentric coordinates $\mathbf{A}_v$ with respect to the triangle's vertex 3D positions are computed. Finally, its position, intensity value and linear intensity gradients are set to $\mathbf{A}_v$, as well as to the corresponding intensity value I and to the linear intensity gradients $\mathbf{g}$ measured on the first image I_0 at position $\mathbf{a}$, respectively.

Due to the movement of the robot, it is possible that at time $t_{k-1} \gg t_0$ the camera will begin to lose sight of the rectangular portion of the planetary surface being described by the surface model. This situation is detected by checking if any of the vertices of the surface model at time t_{k-1} are outside of the camera's field of view. If at least one of them is outside, the surface model's pose is reinitialized with the same position and orientation used in time t_0, as well as all observation points are deleted and new ones are created by using the intensity image I_{k-1} captured at time t_{k-1} instead of the first intensity image I_0.

The rover's 3D motion from time t_{k-1} to time t_k is described by a rotation followed by a translation of its own coordinate system (q, r, s) with respect to the surface model's coordinate system (X, Y, Z). The translation is described by the 3 components of the 3D translation vector $\mathbf{\Delta T} = (\Delta T_X, \Delta T_Y, \Delta T_Z)^\top$ and the rotation is described by 3 rotation angles: $\Delta\omega_X, \Delta\omega_Y, \Delta\omega_Z$. Here, the unknown six motion parameters are represented by the vector $\Delta\mathbf{B}= (\Delta T_X, \Delta T_Y, \Delta T_Z, \Delta\omega_X, \Delta\omega_Y, \Delta\omega_Z)^\top$. The estimation is achieved by maximizing a likelihood function consisting of the natural logarithm of the conditional probability of intensity differences at the N key observation points. The conditional probability is computed by expanding the intensity signal by a Taylor series and neglecting the nonlinear terms, as well as using a linearized 3D observation point position transformation, which transforms the 3D position of an observation point before motion into its 3D position after motion given the rover's 3D motion parameters. Statistically independent zero-mean, common variance, normally distributed intensity measurement errors at the observation points are also assumed. The resulted motion estimates have the following compact form:

$$\Delta\mathbf{B} = \left(\mathbf{O}^\top\mathbf{O}\right)^{-1}\mathbf{O}^\top\mathbf{FD}, \tag{1}$$

where

$$\mathbf{O}^\top = (\mathbf{o}^{(N-1)\top}, \mathbf{o}^{(N-2)\top}, \ldots, \mathbf{o}^{(0)\top})$$

$$\mathbf{o} = \begin{bmatrix} \frac{f\ \bar{g_x}}{A_s} \\ \frac{f\ \bar{g_y}}{A_s} \\ -\frac{f\ (A_q\bar{g_x}+A_r\bar{g_y})}{A_s^2} \\ -\frac{f\ [A_q\bar{g_x}(A_r-C_r)+A_r\bar{g_y}(A_r-C_r)+A_s\bar{g_y}(A_s-C_s)]}{A_s^2} \\ \frac{f\ [A_r\bar{g_y}(A_q-C_q)+A_q\bar{g_x}(A_q-C_q)+A_s\bar{g_x}(A_s-C_s)]}{A_s^2} \\ -\frac{f\ [\bar{g_x}(A_r-C_r)-\bar{g_y}(A_q-C_q)]}{A_s} \end{bmatrix}$$

$$\mathbf{FD}^\top = (fd(\mathbf{a}^{(N-1)}), fd(\mathbf{a}^{(N-2)}), \ldots, fd(\mathbf{a}^{(0)}))$$

$\mathbf{O}$ is the observation matrix; $\mathbf{A}$ is the 3D position of an observation point with respect to the camera coordinate system at time t_{k-1}, which in turn is computed from its barycentric coordinates $\mathbf{A}_v$ and the triangle's vertex 3D positions at time t_{k-1} with respect to the camera coordinate system; $\mathbf{a}$ is the 2D position of the projection of $\mathbf{A}$ into the camera plane; $\bar{\mathbf{g}}$ is the average of the linear intensity gradients $\mathbf{g}$ of the observation point and the linear intensity gradients of the current intensity image I_k at position $\mathbf{a}$; $\mathbf{C} = (C_q, C_r, C_s)^\top$ is the 3D position of the planet's surface model at time t_{k-1}; f the focal length of the camera; and $\mathbf{FD}$ is a vector with the intensity differences $fd(\mathbf{a}^{(n)}) = I_k(\mathbf{a}^{(n)}) - I^{(n)}$ measured at the projections $\mathbf{a}^{(n)}$, $n = 0, 1, \cdots, N-1$, of the N observation points into the image plane.

Since the observation matrix $\mathbf{O}$ resulted from several truncated Taylor series expansions (i.e. approximations), the Eq. (1) needs to be applied iteratively to improve the reliability and accuracy of the estimation.

Assuming that the rover's coordinate system (q, r, s) coincides with the fixed surface coordinate system (α, β, γ) at time t_0, the rover's 3D position with respect to that fixed coordinate system is computed by integrating the estimated frame to frame rover's 3D motion over time.

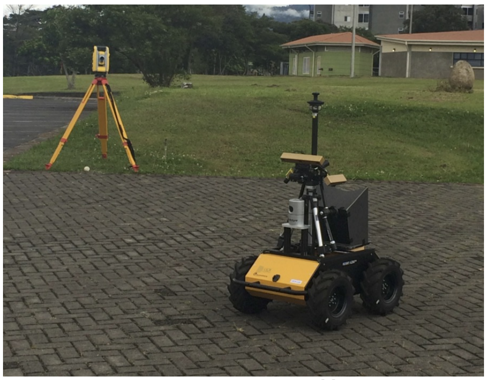


Figure 1. Clearpath Robotics$^{\text{TM}}$ Husky A200$^{\text{TM}}$ rover platform and Trimble® S3 robotic total station used for experimental validation.

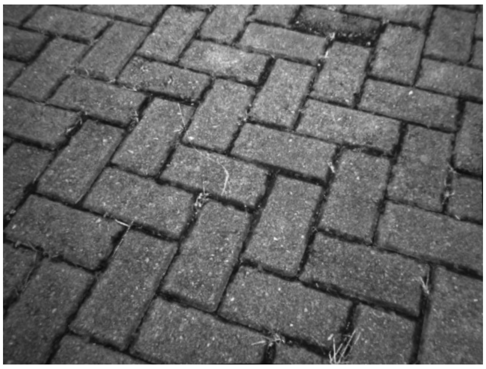


Figure 2. Example of an image with resolution 640x480 $pixel^2$ captured during experiment number 334. The camera is located at 77 cm above the ground looking to the left side of the rover tilted downwards 37 degrees.

3 Experimental Results

The intensity-difference based monocular visual odometry algorithm has been implemented in the programing language C and tested in a Clearpath Robotics$^{\text{TM}}$ Husky A200$^{\text{TM}}$ rover platform (see Fig. 1). In this contribution, our efforts were concentrated on measuring its performance in rover localization on flat ground in real outdoor sunlit conditions, where the absolute position error of distance traveled was used as a performance measure. In total 343 experiments were carried out over flat paver sidewalks only (see Fig. 1), under severe global illumination changes due to cumulus clouds passing fast across the sun. As it has been done on Mars [2], special care was taken to avoid the rover's own shadow in the scene, because the intensity differences due to moving shadows can confuse the motion estimation algorithm. The processing time per image was also measured.

During each experiment, the rover is commanded to drive on a predefined path at a constant velocity of 3 cm/sec over a paver sidewalk, usually a straight segment from 1 to 12 m in length or a clockwise arc from 45 to 280 degrees with 2.5 m radius, while a single camera with a real time image acquisition system captures images at 15 fps and stores them in the onboard computer (see Fig. 2). Although the rover's real time image acquisition system consists of three IEEE-1394 cameras—a 6 mm Grey Point Bumblebee®2 stereo camera, a Grey Point 6 mm Bumblebee® XB3 stereo camera and a 6 mm Basler A601f monocular camera, rigidly attached to the rover by a mast built in its cargo area—only the right camera of the Bumblebee®2 stereo camera was used in all experiments. This camera has an image resolution of 640x480 $pixel^2$ and a horizontal field of view of 43 degrees (see Fig. 2). It is located at 77 cm above the ground looking to the left side of the rover tilted downwards 37 degrees. Because during experiments with arc paths the rover is commanded to rotate clockwise only, only images of the ground outside the arcs can be captured with this camera setup. The radial and tangential distortions due to the camera lens are also corrected in real time by the image acquisition system. This image acquisition software was developed under Ubuntu, ROS and the programing language C.

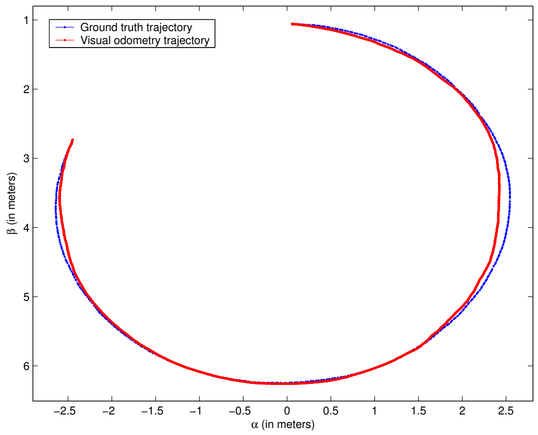
Table 1. Summary of experimental results.

	mean	standard deviation	min	max
Observation points per image	15906	67.74	15775	15999
Iterations per image	14.88	1.89	12.33	19.09
Processing time (in seconds) per image	0.06	0.006	0.05	0.08
Absolute position error	0.9%	0.45%	0.31%	2.12%

Simultaneously, a Trimble® S3 robotic total station (robotic theodolite with a laser range sensor) tracks a prism rigidly attached to the rover and measures its 3D position with high precision ($\leq$ 5 mm) every second (see Fig. 1), where the position and orientation of the local coordinate system of the robotic total station with respect to the planet's surface model coordinate system at time t_0 is precisely known.

After that, the intensity-difference based monocular visual odometry algorithm is applied to the captured image sequence. Then, the prism trajectory is computed from the rover's estimated 3D motion. Finally, it is compared with the ground truth prism trajectory delivered by the robotic total station.

All the experiments were performed on an Intel® Core$^{\text{TM}}$ i5 at 3.1 GHz with 12.0 GB RAM. In Table 1, the main experimental results are summarized. The number of observation points N per image was 15906 on average with a standard deviation of 67.74, a minimum of 15775 and a maximum of 15999 observation points. The average number of motion estimation iterations per image was 14.88 with a standard deviation of 1.89, as well as a minimum and maximum of 12.33 and 19.09 iterations, respectively. The processing time per image was 0.06 seconds on average with a standard deviation of 0.006, a minimum of 0.05 and a maximum of 0.08 seconds. The absolute position error was 0.9% of the distance traveled on average with a standard



(b)

Figure 3. Trajectory obtained by visual odometry (in red) and corresponding ground truth trajectory (in blue) for the experiment number 334. In the experiment the rover was commanded to drive a clockwise arc of 280 degrees with radius of 2.5 m over paver sidewalk.

deviation of 0.45%. The minimum and the maximum absolute position error was 0.31% and 2.12%, respectively. The tracking was not lost in any of the experiments. As an example, Fig. 3 depicts the visual odometry trajectory and the robotic total station trajectory for the path number 334 forming an arc segment of the 343 different paths driven by the rover during the experiments. Although the experiments so far have been only on flat ground, these results closely resembles those achieved by known traditional feature based stereo visual odometry algorithms [7, 8, 2, 9], whose absolute position errors of distance traveled are within the range of 0.15% and 2.5%. Although it is difficult to draw any conclusions from this comparison, since our experiments were carried out in different surface environments and driving modes, we believe that these results are still relevant because they reveal the potential of the algorithm for obtaining the rover's position in real outdoors situations, even under severe global illumination changes, in a non-traditional way, without establishing correspondences between features or solving the optical flow as an intermediate step, just directly evaluating the intensity differences between successive frames delivered by a monocular camera.

4 Conclusion

After testing the monocular visual odometry algorithm proposed in [12] in a real rover platform for localization in outdoor sunlit conditions, even under severe global illumination changes, over flat terrain, along straight lines and gentle arcs at a constant velocity, without the presence of shadows, and comparing the results with the corresponding ground truth data, we concluded that the algorithm is able to deliver the rover's position in average of 0.06 seconds after an image has been captured and with an average absolute position error of 0.9% of distance traveled. Although experiments are still missing over different types of terrain and geometries, particularly over rough terrain, we believe that these results represent an important step towards the validation of the algorithm and that it may be an excellent candidate to be used as an alternative when wheel odometry and traditionally stereo visual odometry have failed. It may also be a great candidate to be merged with other visual odometry algorithms and/or with sensors such as IMUs, laser rangefinders, etc., to improve autonomous navigation of current and future Moon and Mars rovers.

5 Future Work

In the future, the algorithm will be tested over different types of terrain and geometries, and also it will be made robust to shadows.

References

[1] A. Ellery, *Planetary Rovers*, 1st ed., Springer-Verlag, 2016.

[2] M. Maimone, Y. Cheng, L. Matthies, "Two Years of Visual Odometry on the Mars Exploration Rovers", *J. of Field Robotics*, vol. 24, no. 3, pp. 169–186, Mar. 2007.

[3] H. Moravec, "Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover", Ph.D. thesis, Stanford University, USA, 1980.

[4] C. Olson, L. Matthies, M. Schoppers, M. Maimone, "Rover Navigation Using Stereo Ego-Motion", *Robotics and Autonomous Systems*, vol. 43, no. 4, pp. 215–229, June 2003.

[5] A. Johnson, S. Goldberg, Y. Cheng, L. Matthies, "Robust and Efficient Stereo Feature Tracking for Visual Odometry", in *IEEE Int. Conf. on Robotics and Automation*, Pasadena, California, 2008 May 19-23, pp. 39–46.

[6] D. Scaramuzza, "Performance Evaluation of 1-Point-RANSAC Visual Odometry", *J. of Field Robotics*, vol. 28, no. 5, pp. 792–811, Sept./Oct. 2011.

[7] A. Howard, "Real-time Stereo Visual Odometry for Autonomous Ground Vehicles", in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, Nice, France, 2008 Sept. 22-26, pp. 3946–3952.

[8] D. Nister, O. Naroditsky, J. Bergen, "Visual Odometry for Ground Vehicle Applications", *J. of Field Robotics*, vol. 23, no. 1, pp. 3–20, Jan. 2006.

[9] P. Corke, D. Strelow, S. Singh, "Omnidirectional Visual Odometry for a Planetary Rover", in *IEEE Int. Conf. on Intelligent Robots and Systems*, Sendai, Japan, 2004 28 Sept.-2 Oct., pp. 4007–4012

[10] R. Mur-Artal, J. Montiel, J. Tardós, "ORB-SLAM: A Versatile and Accurate Monocular SLAM System", *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, October 2015.

[11] D. Scaramuzza, F. Fraundorfer, "Visual Odometry: Part I: The First 30 Years and Fundamentals", *IEEE Robot. Autom. Mag.*, vol. 18, no. 4, pp. 80–92, Dec. 2011.

[12] G. Martinez, "Intensity-Difference Based Monocular Visual Odometry for Planetary Rovers" in *New Development in Robot Vision*, vol. 23 of the series Cognitive Systems Monographs, Berlin, Heidelberg: Springer Verlag, 2014, ch. 10, pp. 181–198.

[13] R. Tsai, "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses", *IEEE J. of Robotics and Automation*, vol. 3, no. 4, pp. 323–344, Aug. 1987.