

# Uncertainty analysis of a stereo system performing ego-motion measurements in a simulated planetary environment

**M. Pertile, S. Debei, E. Lorenzini**

CISAS G. Colombo, University of Padova, Italy

E-mail: marco.pertile@unipd.it

**Abstract.** In planetary exploration space missions, motion measurement of a vehicle on the surface of a planet is a very important task. In this work, a non linear vision-based algorithm for ego-motion measurement is described and calibrated using telephoto lenses. Several motion types, including displacement, rotation and their combination, are considered and the evaluated uncertainties are compared, pointing out strengths and weaknesses of employing telephoto lenses for motion measurement.

## 1. Introduction

In planetary exploration, motion measurement of a vehicle on the surface of a planet should be very accurate in order to track the vehicle also for long paths. The odometric evaluation of vehicle position and attitude performed measuring the rotation of wheels has wide uncertainty due to slippage of wheels on a natural, often sandy or slippery, surface. Moreover, on extraterrestrial planets GPS-like positioning systems are not yet available and inertial navigation sensors exhibit unacceptable drifts. Thus, the need of a reliable and accurate motion instrument is particularly relevant. In this work a vision-based instrument for displacement and rotation is described and calibrated using a simulated rocky scene. Displacement and rotation is measured through the images taken by a vision system. The use of stereo systems for visual-odometry is well known, e.g. see [1] - [9]. Stereo-processing allows estimation of the three dimensional (3D) location of landmarks observed by a stereo-camera. If the same landmarks are acquired and detected by a stereo-system that moves from an initial position to a final one, the two 3D point clouds allow to evaluate the position and orientation of the stereo-system in the final position with reference to the initial one. [5] describes a method for visual odometry based on a stereo camera, providing experimental results gathered during development and flight phases of the NASA's twin Mars Exploration Rovers Spirit and Opportunity landed on the surface of Mars in January 2004. [8] emphasizes the importance of a detailed and correct uncertainty evaluation, and describes a method for visual-odometry that allows to reduce the final measurement uncertainty. However, [8] does not take explicitly into account the uncertainty contribution of the feature detector that allows to find the projection of landmarks in the image plane. In the papers [1]- [8], a detailed calibration of the vision system as a measurement instrument is not present, while in our analysis it is performed according to [10]. Moreover, in the cited literature, the focus is the development of a frame to frame visual odometry, but how the measurement uncertainty changes if one or more frames are skipped is not analyzed. In this work, we calculate the motion of a stereo system and its uncertainty always taking as reference the images acquired in the initial position; thus, we directly analyze what happens when the motion step is increased and the intermediate frames are not used by the algorithm. The methods described in [1] - [9] require that the same 3D landmarks are observed in the initial and



final position; this requirement limits the maximum translation and, even more, the maximum rotation that can be measured in one step. For wider ranges, several incremental measurements can be performed and summed together, with the warning that uncertainty may significantly increase with the number of steps, i.e. small errors in each measurement step can eventually cause large errors in the estimated trajectory. This drawback can be solved using more global approaches (e.g. Simultaneous Localization And Mapping, loop-closing, semi-global optimization). However, in the present work the attention is focused on a single step measurement (the same 3D landmarks must be observed in the first and in the final position) and more global approaches, which are beyond the purpose of this work, are not taken into account. Once the intrinsic and extrinsic parameters of a stereo system are carefully determined, one of the main uncertainty sources is the position on the image plane of the features detected and matched in corresponding images. Several feature detectors and descriptors were considered in the preliminary phases of this work, see [11] – [18]. Particularly, [15] compares several different detectors invariant to scale and affine transformations and finds out that the best results are obtained by the Hessian-Affine detector and the Maximally Stable Extremal Regions (MSER) approach. The last one performs well on images containing homogeneous regions with distinctive boundaries. [18] makes a comparison among different descriptors and concludes that the Scale Invariant Feature Transform SIFT and the Gradient Location and Orientation Histogram GLOH descriptors are the best. For the reasons described above, the selected detector is the Hessian-Affine (MSER was discarded since the simulated rocky scenes do not contain homogeneous regions with distinctive boundaries), while SIFT descriptor is chosen.

The present paper gives four main contributions, in terms of novelty and originality: first, it performs a detailed calibration and uncertainty evaluation according to [10] and [19] using an experimental set-up which comprises a simulated rocky scene and reference instruments for both translation and rotation measurements; second, a particularly tricky case is taken into account: the two cameras of the stereo system are equipped with a telephoto lens with a small angle of view; the performed analysis allows to evaluate how the whole measurement algorithm behaves in this difficult case; third, the present work directly analyses what happens when the motion step is increased between two acquired couples of images, i.e. when the whole number of analyzed images is reduced; fourth, the paper describes a slightly new measurement method obtained combining together subroutines and procedures recently published and described in different literature sources. In the following section 2 the measurement algorithm is presented, section 3 describes the calibration procedure and the uncertainty analysis, and in section 5 the experimental results are discussed.

## 2. Measurement algorithm

In this section, the procedure employed to perform the displacement and rotation measurement of the stereo system is described. The goal is to calculate the displacement and rotation of a calibrated stereo system using the images acquired in an initial position and in a second one. Thus, the input quantities to be measured are a displacement and/or a rotation of the vision system and the output of the indirect measurement is a numerical evaluation of the displacement and rotation vectors. Rotation is described by a sequence of Euler angles around axes X, Y, Z. The procedure begins with the detection of image features (keypoints) that is performed using the Hessian-Affine detector, see [9], [15], [18] for details. After the corresponding 2D features are detected, a triangulation phase allows to compute the 3D coordinates of the physical landmarks acquired by the cameras. The middle point algorithm is used for triangulation, as in [5], for more details see [9].

The uncertainties of 3D points are needed for the evaluation of stereo system displacement and rotation as described in subsection 2.1.2. Uncertainty evaluation for triangulated 3D points becomes an uncertainty propagation task, which is performed by the Kline-McClintock formula, see GUM [10]. This method is selected, instead of the Monte Carlo propagation approach, since the calculation is embedded in the algorithm that calculate the displacement and rotation of the stereo camera, and is performed for all detected features (see subsection 2.1.2). Thus, the use of a Monte Carlo simulation could lead to unacceptable time delays in the position and attitude evaluation of the stereo camera.

### 2.1. Stereo system displacement and rotation

When the stereo system is calibrated (intrinsic parameters of both cameras and the position and orientation of camera 2 with reference to camera 1 are known), the 3D points  ${}^{P1}\mathbf{X}$  can be calculated for all features detected by both cameras when the vision system is in an initial position P1. The notation  ${}^{P1}\mathbf{X}$  means that these points are expressed in the reference frame attached to the first camera, when the vision system is in the initial position P1. When the vision system is moved (cameras are rigidly connected) from the initial position P1 to a second position P2, the same procedure can be used to compute the 3D vectors  ${}^{P2}\mathbf{X}$  of the same features detected by both cameras in the second position P2 and expressed in the new frame 1 attached to the first camera. For each feature that is detected by both cameras in both positions P1, P2, the following equation can be written:

$${}^{P1}\mathbf{X}_i = {}^{P1}_{P2}\mathbf{R} \cdot {}^{P2}\mathbf{X}_i + {}^{P1}\mathbf{P}_{P2,P1} \quad (1)$$

Where  ${}^{P1}_{P2}\mathbf{R}$  is the rotation matrix from frame 1 in the second position P2 to frame 1 in the initial position P1;  ${}^{P1}\mathbf{P}_{P2,P1}$  is the origin of frame 1 in P2 with reference to the origin of frame 1 in P1 and expressed in P1;  $i = 1, \dots, n$ , with  $n$  being the number of common detected and matched features.

Vector  ${}^{P1}\mathbf{P}_{P2,P1}$  and the Euler angles that define matrix  ${}^{P1}_{P2}\mathbf{R}$  are the numerical output values of the whole measurement procedure and, in this work are evaluated in two steps in a similar way as in [5]: first, a less accurate motion is estimated by least squares estimation embedded within a random sample consensus (RANSAC) process to remove outliers; then, a maximum likelihood motion estimation is performed minimizing a non linear problem. The main differences with [5] are: the feature detection and matching algorithms employed in this work are more advanced; the linear least squares approach is derived from [6]; the non linear minimization procedure comprises the Levenberg-Marquardt algorithm; the analysis is performed for motion steps of different sizes and not in a frame by frame way, i.e. we progressively increased the distance (displacement and rotation) between acquired images.

**2.1.1. Linear least squares step.** In the linear motion evaluation an error vector  $\mathbf{e}_i$  is defined for each couple of 3D points  ${}^{P1}\mathbf{X}_i$ ,  ${}^{P2}\mathbf{X}_i$  detected in both the first P1 and the second P2 position and then the cost function  $E$  to be minimized is calculated:

$$\mathbf{e}_i = {}^{P1}\mathbf{X}_i - {}^{P1}_{P2}\mathbf{R} \cdot {}^{P2}\mathbf{X}_i - {}^{P1}\mathbf{P}_{P2,P1} \quad E = \sum_{i=1}^n \|\mathbf{e}_i\|^2 \quad (2)$$

In order to separate the evaluation of rotation and translation, the centers of the two point clouds are subtracted from the 3D points. Using the formulas explained in [6], first the rotation matrix is evaluated and then the translation  ${}^{P1}\mathbf{P}_{P2,P1}$  of the stereo system is calculated rearranging equation (2).

**2.1.2. Non linear step.** In the non linear step also the covariance matrices of the 3D landmarks, calculated as explained above, are taken into account. For the non linear analysis, the errors defined for each 3D feature in equation (2) is weighted using the inverse of a combined covariance matrix. For each 3D point evaluated with the stereo system in the first position, the corresponding covariance matrix is evaluated; then, this evaluation is repeated for the same feature in the second position to obtain the covariance matrix  ${}^{P2}\mathbf{C}_i$ .  ${}^{P1}\mathbf{C}_i$  and  ${}^{P2}\mathbf{C}_i$  of the same 3D landmark are combined together and the new cost function  $E_{nl}$  (minimized by the Levenberg-Marquardt algorithm) is:

$$\mathbf{C}_i = {}^{P1}\mathbf{C}_i + {}^{P1}_{P2}\mathbf{R} \cdot {}^{P2}\mathbf{C}_i \cdot {}^{P1}_{P2}\mathbf{R}^T \quad E_{nl} = \sum_{i=1}^n \left( \mathbf{e}_i^T \cdot \mathbf{C}_i^{-1} \cdot \mathbf{e}_i \right) \quad (3)$$

In this way each component of the error vector  $\mathbf{e}_i$  of the feature  $i$  can have a different weight that takes into account the uncertainty of feature  $i$  along the considered direction. This heteroscedastic modeling of uncertainty, i.e. inhomogeneous (it may be different from point to point) and anisotropic (it may be different in each direction) is particularly useful in stereo systems whose baseline (i.e. distance between cameras) is small and the uncertainty along a direction roughly parallel to the optical axes

may be much greater than in the other directions. A properly tuned RANSAC algorithm used in the linear phase of motion estimation allows to identify and exclude possible outlier landmarks. Thus, the non linear phase is performed taking into account only 3D points that passed RANSAC algorithm.

### 3. Calibration procedure and Uncertainty analysis

The first step to calibrate the whole measurement system is the determination of intrinsic and extrinsic parameters of the stereo system. The selected method is described in [19]. After the intrinsic and extrinsic parameters of the stereo system are evaluated, the calibration of the measurement system can be performed. To calibrate the whole measurement system, the stereo cameras are mounted on a rotary motorized stage, and/or on a linear motorized stage. The stereo system position and orientation are changed by known quantities and measured by the encoders of the linear and rotary stages which are used as reference instruments; in each position both cameras acquire an image; the imposed movements are evaluated analyzing images in different positions and orientations. Both cameras are aimed to a simulated planetary scene obtained with crumpled brown paper. In experimental tests the types of movements that are taken into account are: longitudinal displacement along the optical axes of the cameras; transverse displacement orthogonal to the optical axes of cameras; pure rotation; rotation combined with transverse displacement.

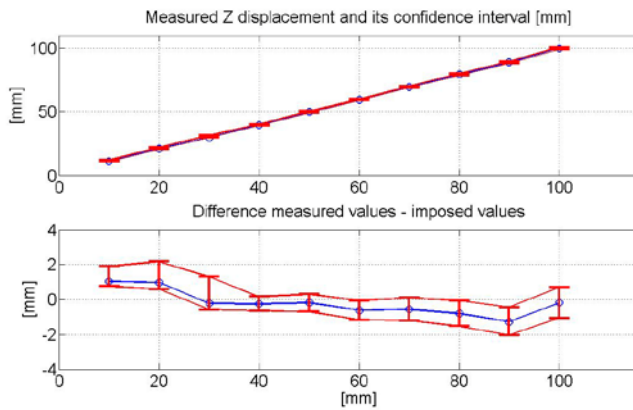
Besides the experimental calibration tests, a detailed uncertainty analysis is carried out according to the metrological procedures described in [10] and [19]. A Monte Carlo propagation is used to evaluate the uncertainty of linear displacement and rotation, which are treated as the output quantities of an indirect measurement. In this analysis phase, the aim is to evaluate uncertainties of position and attitude of the stereo system as precisely as possible, and not to calculate displacements and rotations in a fast way. Moreover, the whole measurement algorithm is highly non linear. Thus, a Monte Carlo simulation is employed instead of the propagation formula embedded in the calculation algorithm as described in section 2 and that is an approximated method.

Several uncertainty sources are analyzed and evaluated using experimental tests also with the simulated planetary scene. Particularly, the uncertainty associated with the following quantities are taken into account: intrinsic and extrinsic parameters of the stereo system, whose uncertainties are evaluated using the camera calibration procedure described in [20], and the positions of image features, whose uncertainty is affected by several contributions. In this paper two main contributions are evaluated for image features: the reading uncertainty (image noise) of the cameras, and the lighting variations of the simulated scene. According to [10] and [19], all uncertainty sources are expressed by a probability density function (PDF) and are then propagated to the output displacement of the stereo system using a Monte Carlo simulation. In this way, both the calibration curve and its uncertainty are evaluated for the output measured displacement and/or rotation.

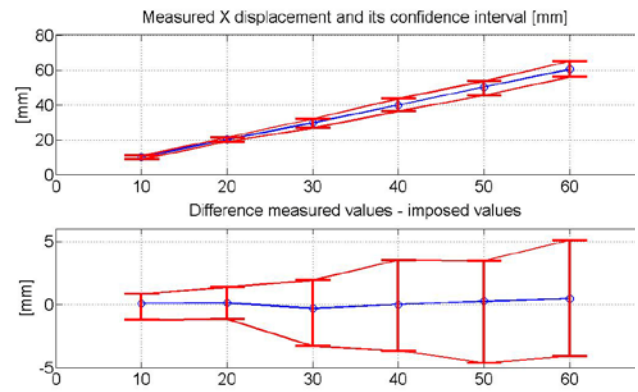
### 4. Results

The results are the obtained calibration curves of the instrument (figures 1-5). Figure 1 shows the curve and its evaluated uncertainty for longitudinal displacement (substantially parallel to the optical axes), figure 2 for transverse displacement (substantially orthogonal to the optical axes), Figure 3 for rotation, figures 4 and 5 for combined rotation and transverse displacement. For all figures 1-5, uncertainty is depicted with level of confidence (l.o.c.) 95.5% and in the upper graph the vision system measurement is depicted vs. the motion measured by encoders; in the bottom graph the difference between the values measured by the stereo system and the imposed ones (measured by encoders) is depicted for each step amplitude that is analyzed. Comparing figures 1 and 2, the uncertainty evaluated for transverse displacement is larger than that obtained for longitudinal movements, except for the first two motion steps of figure 2. This is an interesting result of the analyzed approach, since the uncertainty of 3D points acquired by the stereo system is wider along the direction parallel to the optical axis than along a transverse direction, due to the small distance between the two cameras. However, this disadvantage along longitudinal direction is compensated by the greater number of

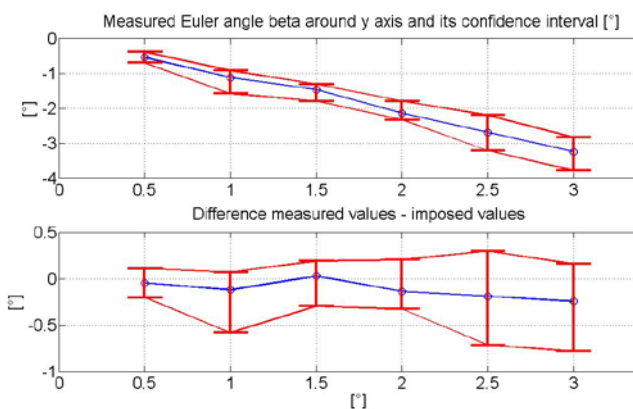
image features correctly matched in the case of axial displacement. Figure 6 depicts the number of matched features vs. motion step amplitude for all the four considered types of motion.



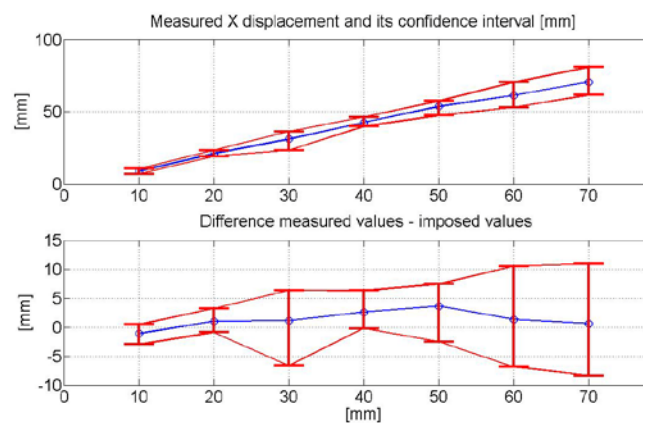
**Figure 1.** Longitudinal displacement (l.o.c. 95.5%).



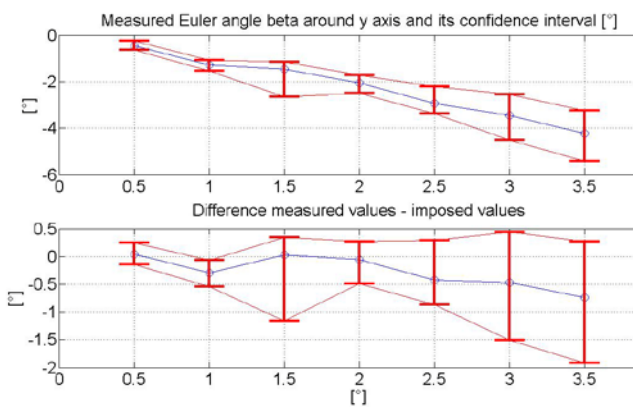
**Figure 2.** Transverse displacement (l.o.c. 95.5%).



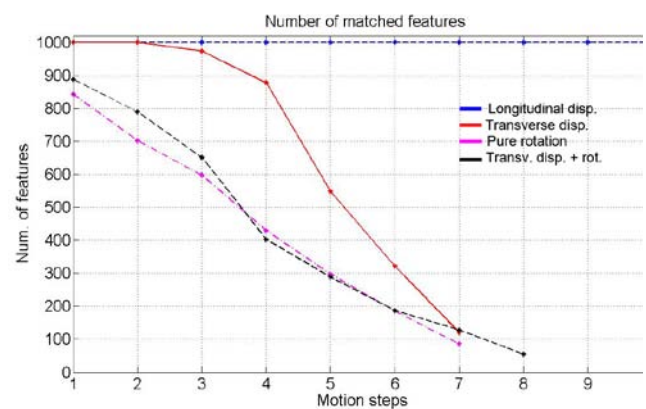
**Figure 3.** Pure rotation (l.o.c. 95.5%).



**Figure 4.** Transverse displacement, with rotation.



**Figure 5.** Rotation combined with transverse displacement (l.o.c. 95.5%).



**Figure 6.** Number of matched features vs. motion step amplitude.

From figure 6 it is clear that in the first two motion steps of the transverse displacement case (figure 2) the algorithm employs the same number of 3D points than in the longitudinal displacement case. This fact may explain the comparable uncertainty obtained. In the other motion steps of the transverse displacement case, the number of matched features rapidly decreases and uncertainty gets wider. Thus, with a photo lens, due to its small angle of view, the variation of the number of matched features seems to prevail over the differences in the uncertainty of 3D points.

Figure 3 shows that the evaluation of small rotation angles with a telephoto lens is not an easy task. There are some physical limiting factors inherent in the stereo system with two telephoto lenses that was employed: the angle of view of the cameras is small ( $9.80^\circ \times 7.85^\circ$  in our case); the two cameras have substantially parallel optical axes; the depth of field of the cameras is narrow, thus, the acquired 3D points exhibit a small distance along the optical axes of the cameras. The physical factors listed above make the maximum measurable angular step very limited, as confirmed by figure 3. The fourth type of motion that was taken into account is particularly difficult to tackle using a telephoto lens.

In all cases, except for the first one of longitudinal displacement (figure 1), the evaluated uncertainty increases with motion step. A possible reason of this behavior is that the larger the motion step and the fewer the correctly matched features among images; thus, the averaging effect associated with a large number of matched features decreases with step amplitude. This reduction of matched features is not present for movements substantially parallel to the optical axis, see figure 6. Thus, in the first case the uncertainty does not increase with the motion step and longer distances can be measured in this direction than in the transverse one, as it can be seen comparing figures 1 and 2.

## 5. Conclusion

The calibration of a stereo system as an instrument for displacement and rotation measurement was described. The calibration was carried out using a simulated rocky scene and different types of motion. This procedure allowed to highlight advantages and drawbacks of the considered system which is provided with telephoto lenses. The variation of uncertainty due to a progressive increase of the motion step between acquired images was analyzed.

## References

- [1] Arun K. S., Huang T. S., Blostein S. D., *IEEE Trans on PAMI*, Vol.9, No.5, 1987, pp. 698–700.
- [2] Eggert D., Lorusso A., Fisher R., *Machine Vision and App.*, Vol.9, No.5-6, 1997, pp.272–290.
- [3] Umeyama S., *IEEE Trans on PAMI*, Vol. 13, No.4, 1991, pp. 376 – 380.
- [4] Olson C. F., et Al., *Robotics and Autonomous Systems*, Vol. 43, 2003, pp. 215–229.
- [5] Cheng Y., Maimone M. W., Matthies L., *IEEE Robotics & Automation Magazine*, 2006.
- [6] Horn B.K.P., et Al., *Jou. of the Optical Soc. of Am.*, July 1988, Vol.5, pp.1127.
- [7] Ma Y., Soatto S., Kosecka J., Sastry S. S., *An invitation to 3-D Vision*, Springer, 2004.
- [8] Dubbelman G., Groen F. C. A., *IEEE CVPR*, 2009.
- [9] M Pertile, M Magnabosco, S Debei, *Journal of Physics: Conference Series* 238 (2010) 012031.
- [10] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML, *Evaluation of measurement data—Guide to the Expression of Uncertainty in Measurement*, Geneva, Switzerland: ISO, 2008.
- [11] Harris C., Stephens M., *Proc. Alvey Vision Conf.*, 1988, pp. 147-151.
- [12] Mikolajczyk K., Schmid C., *Proc. 8th Int. Conf. Computer Vision*, 2001, pp. 525-531.
- [13] Lowe D., *Proc. 7th Int. Conf. Computer Vision*, 1999, pp. 1150-1157.
- [14] Lowe D., *Int. J. Computer Vision*, Vol. 2, No. 60, 2004, pp. 91-110.
- [15] Mikolajczyk K., Schmid C., *Int. J. Computer Vision*, Vol. 1, No. 60, 2004, pp. 63-86.
- [16] Mikolajczyk K., et Al., *Int. J. Computer Vision*, No. 65, 2005, pp. 43–72.
- [17] Matas J., et Al., *Image and Vision Computing*, Vol. 22, 2004, pp.761–767.
- [18] Mikolajczyk K., Schmid C., *IEEE Trans On PAMI*, Vol. 27, No. 10, 2005.
- [19] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML, *Evaluation of measurement data—Supplement 1 to the Guide to the Expression of Uncertainty in Measurement*.
- [20] Zhang Z., *IEEE Trans on PAMI*, vol. 22, n.11, 2000.