Chapter 10
Intensity-Difference Based Monocular Visual Odometry for Planetary Rovers

Geovanni Martinez

Abstract. A monocular visual odometry algorithm is presented that is able to estimate the rover's 3D motion by maximizing the conditional probability of the intensity differences between two consecutive images, which were captured by a monocular video camera before and after the rover's motion. The camera is supposed to be rigidly attached to the rover. The intensity differences are measured at observation points only that are points with high linear intensity gradients. It represents an alternative to traditionally stereo visual odometry algorithms, where the rover's 3D motion is estimated by maximizing the conditional probability of the 3D correspondences between two sets of 3D feature point positions, which were obtained from two consecutive stereo image pairs that were captured by a stereo video camera before and after the rover's motion. Experimental results with synthetic and real image sequences revealed highly accurate and reliable estimates, respectively. Additionally, it seems to be an excellent candidate for mobile robot missions where space, weight and power supply are really very limited.

10.1 Introduction

Over the past two decades, robotic rovers have been extensively used for planetary surface exploration and have demonstrated that unmanned missions are very practical and productive, as well as much cheaper and less risky than manned ones. Up to date, the most successful planetary rovers have been the following six-wheel rocker-bogie rovers developed at NASA Jet Propulsion Laboratory: the Mars Pathfinder mission rover Sojourner [1], the Mars Exploration Rovers (MER) Spirit and Opportunity [2] and the Mars Science Laboratory (MSL) rover Curiosity [3]. In order to increase the maximum exploration range from few tens of kilometers to hundreds of

Geovanni Martinez
Image Processing and Computer Vision Research Laboratory (IPCV-LAB),
Escuela de Ingeniería Eléctrica, Universidad de Costa Rica, 2060 San José, Costa Rica
e-mail: gmartin@eie.ucr.ac.cr

Y. Sun et al. (eds.), *New Development in Robot Vision*,
Cognitive Systems Monographs 23, DOI: 10.1007/978-3-662-43859-6_10

Kilometers, a Mars Airplane is also being investigated [4]. As opposite to the Mars Airplane, flapping insect robots (entomopters) are also being studied due to their potential to fly slow as well as safety land and take off on rocky planets with low atmospheric pressure like Mars [5].

Since communication between Earth and Mars rovers only occurs once or twice a day, and there is a significant delay from the time a command is sent to when the rover receives it, they must be able to autonomously navigate to science targets and to place instruments precisely against these targets, where any navigation error could cause the loss of the entire day of scientific activity.

For reliable and precise autonomous navigation, the rovers need to be able to determine its position and orientation at any time instant. Usually the current rover's position and orientation are estimated by integrating the rover's motion (rover's change of position and orientation) from the time the motion began to the current time, assuming that the initial rover's position and orientation are known or previously estimated. In the MER rovers Spirit and Opportunity the rover's change of orientation (rover's rotation) is estimated from measurements of three-axis angular rate sensors (gyros) provided by an Inertial Measurement Unit (IMU) onboard the rover [6]. The rover's change of position (rover's translation) is estimated from encoder readings of how much the wheels turned (wheel odometry). The initial rover's orientation is estimated from measurements of three-axis accelerometers provided by the IMU, as well as a sun position vector provided by a sun sensor which is also onboard the rover [7]. The initial position is reset by command at the beginning of the rover's motion.

One limitation of the rocker-bogie mobile rovers as observed on Mars is the excessive wheel slippage on steep slopes and soft soils, which causes large errors particularly on the estimated rover's position from wheel odometry. These position errors can even bring the rover to get stuck, digging itself deeper and deeper into a sand hole, from where it is difficult to get the rover back up on solid ground without incurring some significant damage in the process.

To correct any position error due to wheel slippage, the rover's 3D motion is also estimated by using a stereo visual odometry algorithm. It estimates the rover's 3D motion by maximizing the conditional probability of the 3D correspondences between two sets of 3D feature point positions, which were previously obtained from two consecutive stereo image pairs that were captured by a stereo video camera before and after the rover's motion, respectively. This feature based stereo visual odometry algorithm was first proposed by Moravec in [8] and then improved in [9, 10]. Afterwards, it evolved to become more robust [11] until it was finally implemented in real time to be used in the Mars Exploration Rover Mission [12]. Similar feature based stereo visual odometry algorithms have been also presented in [13, 14, 15]. In [14] and [16] feature based visual odometry algorithms are also described using a monocular standard video camera and an omnidirectional video camera, respectively. In [17, 18] feature based monocular visual odometry is extended to a Simultaneous Localization and Mapping (SLAM).

In the stereo visual odometry algorithm used in the Mars Exploration Rover Mission [12], feature points are first selected on the left frame (image) of the first stereo

pair. Next, the feature points are also found in the corresponding right frame by correlation based template matching. Then, the 3D positions of the feature points before rover's motion are estimated from the 2D positions of the feature points in both frames of the first stereo pair by stereo triangulation. Next, the feature points are moved using an initial motion estimate provided by the onboard wheel odometry, then they are projected into the left frame of the second stereo pair and their positions refined by correlation based template matching. Then, the feature points are also found in the corresponding right frame by correlation based template matching. Next, the 3D positions of the feature points after rover's motion are estimated from the 2D positions of the feature points in both frames of the second stereo pair by stereo triangulation. Next, the 3D correspondences (3D offsets) between the two sets of 3D feature point positions before and after the rover's motion are established. Finally, those 3D motion parameters which maximize the conditional probability of the established 3D correspondences are considered to be the stereo visual odometry 3D motion estimates. The conditional probability is computed by modeling the 3D position error at each feature point with Gaussian distributions and using a linearized 3D feature point position transformation, which transforms the 3D position of a feature point before motion into its 3D position after motion given the rover's 3D motion parameters.

After evaluating the performance of the above stereo visual odometry algorithm in both MER rovers Spirit and Opporttuniy on Mars, it was further improved in [19] resulting in a more robust and at least four time more computationally efficient algorithm, which can also operate with no initial motion estimate from wheel odometry. This last updated version of the stereo visual odometry algorithm was planed to be used in the MSL rover Curiosity.

In this paper, as an alternative to the traditional feature based stereo visual odometry, an intensity-difference based monocular visual odometry algorithm is described, which will be able to estimate the rover's 3D motion evaluating the intensity differences at different observation points between two intensity frames captured by a monocular video camera before and after the robot's motion, where an observation point is an image point with high linear intensity gradient. This avoids establishing feature points correspondences for rover's 3D motion estimation and the problems associate with it. The rover's 3D motion will be estimated by maximizing the conditional probability of the frame to frame intensity differences at the observation points. The conditional probability is computed by expanding the intensity signal by a Taylor series and neglecting the nonlinear terms, resulting the well known optical flow constraint [20, 21], 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. Perspective projection of the observation points into the image plane and zero-mean Gaussian stochastic intensity errors at the observation points are also assumed. Similar approaches have been already implemented successfully in applications such as video compression [22] and teleoperation of space robots [23].

Our approach differs from traditional optical flow approaches such as described in [24, 25] because we do not follow the typical two-stage algorithm, where the

optical flow vector field is first estimated and then the rover's 3D motion is estimated from the previously estimated optical flow vector field, instead we follow a similar approach such as described in [26] by developing an one-stage estimation algorithm, which is able to directly deliver the rover's 3D motion parameters without estimating the optical flow vector field, avoiding in this way solving the optical flow as an intermediate step and the associate problems with the flow estimation.

This contribution is organized as follows. In section 10.2, the proposed monocular visual odometry algorithm is presented. In section 10.3, the experimental results are presented. In section 10.4, a brief summary and the conclusions are given. Finally, in section 10.5, a short description of the future work is given.

10.2 Monocular Visual Odometry Algorithm

In this section, an algorithm to estimate the 3D motion of a rover between two arbitrary time instants t_{k-1} and t_k is presented. This is done by maximizing the conditional probability of the intensity differences between two consecutive frames I_{k-1} and I_k, which were captured at the same time instants by a monocular camera rigidly attached to the rover. The intensity differences are measured at selected points called observation points. The conditional probability is a function of the rover's 3D motion, the frame to frame intensity differences and the covariance matrix of the intensity errors at the observation points. To compute this conditional probability, a mathematical relationship between the rover's 3D motion and the frame to frame intensity differences at the observation points is used. This relationship is based on a number of assumptions about the world and how it is projected into the image plane of the camera. This assumptions are reviewed in subsection 10.2.1. Subsection 10.2.2 explains what is an observation point and how the intensity differences are measured at the observation points. In subsection 10.2.3, the conditional probability of the intensity differences at the observation points is computed. In subsection

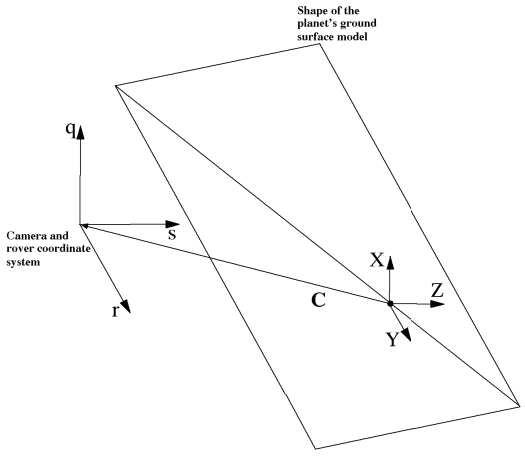


Fig. 10.1 Shape of the planet's ground surface model. Currently, it is described by a planar and rigid mesh of two triangles with coordinate system (X,Y,Z). The camera coordinate system (q,r,s) is supposed to coincide with the robot coordinate system.

10.2.4, the method to maximize the conditional probability for getting the rover's 3D motion estimates is described. Finally, subsection 10.2.5 explains how the algorithm is initialized at the beginning of the image sequence.

10.2.1 Planet's Ground Surface Model

For rover's 3D motion estimation from time t_{k-1} to time t_k, it is assumed that the monocular camera is looking down toward the planet's ground surface, that the planet's ground surface is visible by the camera and that it covers the whole area of the captured intensity images. Furthermore, it is assumed that at time t_{k-1} a model of a visible portion of the planet's ground surface is available or has been previously computed as described in subsection 10.2.5, whose shape with coordinate system (X,Y,Z), texture and relative pose to the camera coordinate system (q,r,s) correspond with those of the real portion of the planet's ground surface at time t_{k-1}. For simplicity, from now on, this model will be referred as planet's ground surface model. The model's shape is represented by a rigid mesh of triangles. Currently, the mesh consists of only two triangles forming a rectangle (see Fig. 10.1). The model's texture is described by the intensity and chrominance values being reflecting from it. The model's pose is given by six parameters: the three components of a 3D position vector and three rotation angles. In addition, the camera coordinate system (q,r,s) is supposed to coincide with the rover coordinate system and an image is supposed to be generated by perspective projection of the planet's ground surface into the image plane of the camera. Moreover, it is assumed that there are no moving objects on the planet's ground surface and that the illumination is diffuse as well as spatial and time invariant. Thus, the frame to frame intensity differences are due to the frame to frame rover's 3D motion only.

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) respect to the ground surface coordinate system (X,Y,Z). The translation is described by the 3 components of the 3D translation vector $\Delta\mathbf{T} = (\Delta T_X, \Delta T_Y, \Delta T_Z)^\top$. The rotation is described by 3 rotation angles: $\Delta\omega_X$, $\Delta\omega_Y$, $\Delta\omega_Z$. These six motion parameters are represented by the vector $B = (\Delta T_X, \Delta T_Y, \Delta T_Z, \Delta\omega_X, \Delta\omega_Y, \Delta\omega_Z)^\top$ and estimated by maximizing the conditional probability of the frame to frame intensity differences measured at the observation points.

10.2.2 Observation Points

For estimating the rover's 3D motion from time t_{k-1} to time t_k, it is also assumed that at time t_{k-1} a set of observation points is available. An observation point lies on one of the triangles of the mesh of the planet's ground surface model at barycentric coordinates $\mathbf{A}_v$ and carries the intensity value I, as well as the linear intensity gradients $g = (g_x, g_y)^\top$ at position A_v. The observation points are created and initialized at the beginning of the image sequence according to the method that is described in subsection 10.2.5. Let $A = (A_q, A_r, A_s)^\top$ be the corresponding position of

an observation point with respect to the rover coordinate system at time t_{k-1} and let $a = (a_x, a_y)^\top$ be the position of its perspective projection into the camera plane (see Fig. 10.2). Assuming that the shape, position and orientation of the planet's ground surface model correspond with those of the real portion of the planet's ground surface at time t_{k-1}, the frame to frame intensity difference fd at the observation point a is approximated as follows:

$$fd(\mathbf{a}) = I_k(\mathbf{a}) - I_{k-1}(\mathbf{a}) \approx I_k(\mathbf{a}) - I \tag{10.1}$$

where $I_{k-1}(a)$ and $I_k(a)$ represent the intensity value of the images I_{k-1} and I_k at the position a, respectively. Since in general a lies outside of the image raster, the intensity value $I_k(a)$ is computed by bilinear interpolation of the intensity values of the nearest four pixels of the intensity image I_k. Then, the frame to frame intensity difference at N observation points is represented as follows:

$$FD = (fd(a^{(N-1)}), fd(a^{(N-2)}), \ldots, fd(a^{(0)}))^\top \tag{10.2}$$

Moreover, the mean squared frame to frame intensity difference at the observation points is given by:

$$msd \;=\; \frac{1}{N}\sum_{n=0}^{N-1} fd(a^{(n)})^2 \tag{10.3}$$

10.2.3 Conditional Probability of the Intensity Differences

Let's consider an arbitrary observation point at barycentric coordinates A_v on the surface of one of the triangles of the planet's ground surface model. Let I and $g = (g_x, g_y)^\top$ be its intensity value and linear intensity gradients, respectively. Due

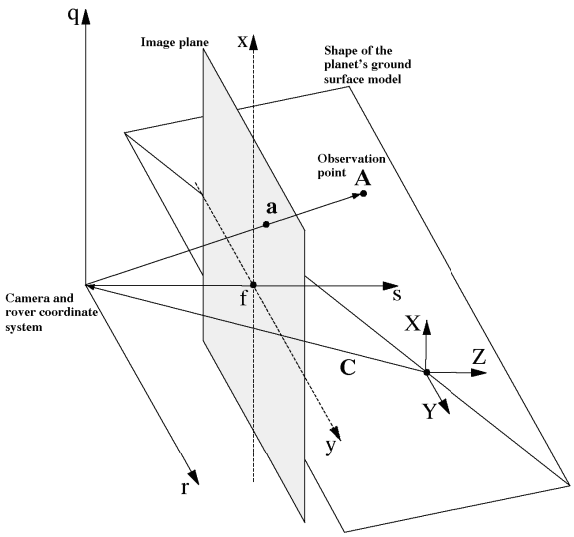


Fig. 10.2 Observation point A on the planet's ground surface model with respect to the camera coordinate system and its perspective projection a into the image plane

to the robot's motion from time t_{k-1} to time t_k it moves from A to A' with respect to the camera coordinate system (see Fig. 10.3). The corresponding perspective projections into the image plane are a and a', respectively. Expanding the intensity signal I_{k-1} at image position a by a Taylor series and neglecting the nonlinear terms as proposed in [21], the following relationship between the unknown position a' and the frame to frame intensity difference is obtained:

$$fd(a) = I_k(a) - I_{k-1}(a) \approx -g^\top (a' - a) \tag{10.4}$$

which is known in the professional literature as Horn and Schunck optical flow constraint equation [20]. Expressing a with their corresponding coordinates at the camera coordinate system using a perspective camera model with known focal distance f results:

$$a = \begin{bmatrix} \frac{f\,A_q}{A_s} \\ \frac{f\,A_r}{A_s} \end{bmatrix} \tag{10.5}$$

Approximating the Eq. 10.5 at position A by using a Taylor series and neglecting the nonlinear terms the following transformation for the unknown position a' can be obtained:

$$a' \approx a + \begin{bmatrix} \frac{f}{A_s} & 0 & \frac{f\,A_q}{A_s^2} \\ 0 & \frac{f}{A_s} & \frac{f\,A_r}{A_s^2} \end{bmatrix} (A' - A) \tag{10.6}$$

where the known position $A = (A_q, A_r, A_s)^\top$ is related with the unknown position $A' = (A'_q, A'_r, A'_s)^\top$ according to:

$$A' = \Delta\mathbf{R}\,(A + C) - C - \Delta\mathbf{T} \tag{10.7}$$

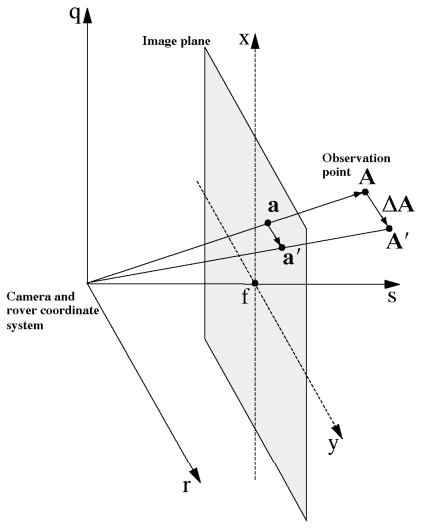


Fig. 10.3 Translation ΔA of an observation point due to the rover's 3D motion with respect to the camera coordinate system and its perspective projection into the camera plane

In the previous equation $C = (C_X, C_Y, C_Z)^\top$ represents the origin of the robot coordinate system respect to the ground surface coordinate system (see Fig. 10.1) and $\Delta\mathbf{R}$ a rotation matrix computed with the rotation angles $-\Delta\omega_X$, $-\Delta\omega_Y$, $-\Delta\omega_Z$. Assuming that the rotation angles $-\Delta\omega_X$, $-\Delta\omega_Y$, $-\Delta\omega_Z$ are small, thus $\cos(-\Delta\omega) \approx 1$ and $\sin(-\Delta\omega) \approx -\Delta\omega$, the Eq. 10.7 can be transformed into:

$$A' \approx \begin{bmatrix} A_q' \\ A_r' \\ A_s' \end{bmatrix} = \begin{bmatrix} 1 & \Delta\omega_Z & -\Delta\omega_Y \\ -\Delta\omega_Z & 1 & \Delta\omega_X \\ \Delta\omega_Y & -\Delta\omega_X & 1 \end{bmatrix} \begin{bmatrix} A_q + C_X \\ A_r + C_Y \\ A_s + C_Z \end{bmatrix} - \begin{bmatrix} C_X - \Delta T_X \\ C_Y - \Delta T_Y \\ C_Z - \Delta T_Z \end{bmatrix} \tag{10.8}$$

Substituing Eq. 10.8 in Eq. 10.6, we obtain:

$$a' \approx \begin{bmatrix} \frac{f\,A_q}{A_s} \\ \frac{f\,A_r}{A_s} \end{bmatrix} + \begin{bmatrix} \frac{f}{A_s} & 0 & \frac{f\,A_q}{A_s{}^2} \\ 0 & \frac{f}{A_s} & \frac{f\,A_r}{A_s{}^2} \end{bmatrix} \begin{bmatrix} \Delta\omega_Z(A_r + C_X) - \Delta\omega_Y(A_s + C_Z) - \Delta T_X \\ -\Delta\omega_Z(A_q + C_X) + \Delta\omega_X(A_s + C_Z) - \Delta T_Y \\ \Delta\omega_Y(A_q + C_X) - \Delta\omega_X(A_r + C_Y) - \Delta T_Z \end{bmatrix} \tag{10.9}$$

Finally, substituting Eqs. 10.9 and 10.5 in Eq. 10.4, the following linear equation that relates the unknown motion parameters and the frame to frame intensity difference measured at the observation point position **a** is obtained:

$$\begin{aligned} fd(a) = {} & \frac{f\,g_x}{A_s}\,\Delta T_X + \frac{f\,g_y}{A_s}\,\Delta T_Y - \frac{f\,(A_q g_x + A_r g_y)}{A_s^2}\,\Delta T_Z + \\ & - \frac{f\,[A_q g_x (A_r + C_Y) + A_r g_y (A_r + C_Y) + A_s g_y (A_s + C_Z)]}{A_z^2}\,\Delta\omega_X + \\ & + \frac{f\,[A_r g_y (A_q + C_X) + A_q g_x (A_q + C_X) + A_s g_x (A_s + C_Z)]}{A_z^2}\,\Delta\omega_Y + \\ & - \frac{f\,[g_x (A_r + C_Y) - g_y (A_q + C_X)]}{A_s}\,\Delta\omega_Z \end{aligned} \tag{10.10}$$

Eq. 10.10 can also be written in vector form as:

$$fd(a) = o^\top B + \Delta I \tag{10.11}$$

where

$$o = \begin{bmatrix} \frac{f\,g_x}{A_s} \\ \frac{f\,g_y}{A_s} \\ -\frac{f\,(A_q g_x + A_r g_y)}{A_s^2} \\ -\frac{f\,[A_q g_x (A_r + C_Y) + A_r g_y (A_r + C_Y) + A_s g_y (A_s + C_Z)]}{A_s^2} \\ \frac{f\,[A_r g_y (A_q + C_X) + A_q g_x (A_q + C_X) + A_s g_x (A_s + C_Z)]}{A_s^2} \\ -\frac{f\,[g_x (A_r + C_Y) - g_y (A_q + C_X)]}{A_s} \end{bmatrix} \tag{10.12}$$

and the term ΔI represents the frame to frame intensity error caused by the camera noise, the local or global temporal illumination changes, the shape error of the planet's ground surface model and the accumulated position error due to the motion estimation errors occurred by the motion analysis of previous frames.

Evaluating Eq. 10.11 at $N > 6$ observation points the following system of linear equations is obtained:

$$FD = O\,B + V \tag{10.13}$$

where O is the observation matrix:

$$O = \left[{o^{(N-1)}}^\top, {o^{(N-2)}}^\top, \ldots, {o^{(0)}}^\top \right]^\top \tag{10.14}$$

and V is the vector with the N intensity errors:

$$V = \left[\Delta I^{(N-1)}, \Delta I^{(N-2)}, \ldots, \Delta I^{(0)} \right]^\top \tag{10.15}$$

The latter can be computed solving for V in Eq. 10.13:

$$V = FD - O\,B \tag{10.16}$$

Modeling the intensity error $\Delta I^{(n)}$ with image coordinates $a^{(n)}$ by a stationary zero-mean Gaussian stochastic process, the joint probability density of the intensity errors at N observation points with image coordinates $a^{(n)}$, $n = 0,..,N-1$, can be computed as:

$$p(V) = \frac{1}{\sqrt{(2\pi)^N\,|U|}}\, e^{-\frac{1}{2}\left(V^\top\, U^{-1}\, V\right)} \tag{10.17}$$

where $|U|$ is the determinant of the covariance matrix U of the intensity errors at the N observation points. Assuming that the variance of the intensity error $\Delta I^{(n)}$ with image coordinates $a^{(n)}$ is 1 and that the intensity errors are statistically independent, this covariance matrix becomes the identity matrix:

$$U = E[V\,V^\top] = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{10.18}$$

Substituting Eq. 10.16 in Eq. 10.17 the conditional probability of the frame to frame intensity differences at the N observation points can be written as follows:

$$p(FD|B) = \frac{1}{\sqrt{(2\pi)^N\,|U|}} e^{-\frac{1}{2}\left((FD-O\,B)^\top U^{-1}(FD-O\,B)\right)} \tag{10.19}$$

10.2.4 Maximizing the Conditional Probability

The 3D motion parameters $B = (\Delta T_X, \Delta T_Y, \Delta T_Z, \Delta \omega_X, \Delta \omega_Y, \Delta \omega_Z)^\top$, which describe the rover's 3D motion from time t_{k-1} to time t_k, are estimated by maximizing the Eq. 10.19. To this end, the derivative of the natural logarithm of the Eq. 10.19 is first computed and then set to 0 as follows:

$$\frac{\partial \ln p(FD|B)}{\partial B} = \frac{\partial((FD - O\,B)^\top U^{-1}(FD - O\,B))}{\partial B} = 0 \tag{10.20}$$

Finally, the Maximum-Likelihood motion estimates are obtained by solving for B in the above equation:

$$\widehat{B} = \left(O^\top U^{-1} O\right)^{-1} O^\top U^{-1} FD \tag{10.21}$$

Due to the truncation errors caused by neglecting the nonlinear terms in all approximations done to obtain Eq. 10.10, the algorithm needs to be applied iteratively to improve its reliability and accuracy [22, 23]. After each iteration i, the resulting estimates ${}^i\widehat{B}$ are used to compensate the motion of the planet's ground surface model, as well as to update the motion estimates $\widehat{B}$ found by previous iterations. Due to the motion compensation, the frame to frame intensity differences at the observation points decreases. The iteration ends when after two consecutive iterations the mean square frame to frame intensity difference at the observation does not decrease significantly. In each iteration i the following steps are carried out:

1. Evaluate Eq. 10.10 at each observation point
2. Compute the intensity differences ${}^iFD$ and observation matrix ${}^iO$ according to Eq. 10.2 and Eq. 10.14, respectively
3. Obtain the motion estimates ${}^i\widehat{B}$ using Eq. 10.21
4. Compensate the motion of the mesh of triangles by moving its vertices according to Eq. 10.7 with the estimates ${}^i\widehat{B}$
5. Compute the mean squared intensity difference ${}^imsd$ using Eq. 10.3
6. Update the rotation matrix: $\widehat{\Delta\mathbf{R}} \leftarrow {}^i\widehat{\Delta\mathbf{R}}\,\widehat{\Delta\mathbf{R}}$
7. Update the translation vector: $\widehat{\Delta T} \leftarrow \widehat{\Delta T} + {}^i\widehat{\Delta T}$
8. If $|{}^imsd - {}^{i-1}msd| \geq \delta_2$ goto step 1

10.2.5 Planet's Ground Surface Model Initialization

As explained in subsections 10.2.1 and 10.2.2, to estimate the rover's 3D motion from time t_{k-1} to time t_k, the shape, the pose and the observation points of a visible portion of the planet's ground surface need to be known at time t_{k-1}. Here, they are computed by compensating the motion of a planet's ground surface model with the accumulated motion estimates from time t_0 to time t_{k-1}. The planet's ground surface model is created and initialized at the beginning of the image sequence. The motion compensation is gradually performed in the fourth step of each iteration of the motion estimation algorithm described in the previous subsection.

The initialization of the shape, the pose and the texture of the model along with the observation points is carried out only once at the beginning of the image sequence, one after the other, as follows:

1. *Shape initialization*: The model's shape is currently initialized as a flat and rigid mesh of two triangles forming a rectangle. Thus, at present, it is only possible to model the shape of a flat portion of the visible planet's ground surface. Here, the image area of that portion is at least 20% of the total image area.
2. *Pose initialization*: Currently, the orientation of the model's shape is initialized manually so that it fairly corresponds to that of the surface portion being modeled. The position is also set manually so that the perspective projection of the mesh into the image plane covers the corresponding image region of the surface portion being modeled. The focal lens f is assume to be 1. In this uncalibrated case, the robot translation can only be estimated up to scale factor.
3. *Texture initialization*: The model's texture is initialized by projecting the intensity and chrominance values of the first image of the sequence onto its mesh of triangles by using texture mapping.
4. *Observation points initialization*: The observation points are created and initialized as follows. First, the gradient images G_{0x} and G_{0y} are computed by convolving the first intensity image I_0 with the Sobel operator. Then, the 3D vertex positions of all visible triangles of the ground surface model are perspectively projected into the camera plane. In order to reduce the influence of the camera noise and to increase the accuracy of the estimation, an arbitrary image point $\mathbf{a}$ inside the image region of a projected triangle will be selected as an observation point only if the linear intensity gradient at position $\mathbf{a}$ satisfies $|\mathbf{G}_0(\mathbf{a})| > \delta_1$. Next, the 3D position vector $\mathbf{A}$ with respect to the camera coordinate system of each selected observation point is computed as the intersection of the $\mathbf{a}$'s line of sight and the plane containing the corresponding triangle's vertex 3D positions. Then, the corresponding barycentric coordinates $\mathbf{A}_v$ with respect to the vertex 3D positions are also computed. Finally, each selected observation point is rigidly attached to the triangle's surface. Its position, intensity value I and linear intensity gradient $\mathbf{g} = (g_x, g_y)^\top$ are set to $\mathbf{A}_v$, $I_0(\mathbf{a})$ and $(G_{0x}(\mathbf{a}), G_{0y}(\mathbf{a}))^\top$, respectively. Since it is assumed that the illumination is diffuse as well as spatial and time invariant, the intensity value I and the linear intensity gradient $\mathbf{g} = (g_x, g_y)^\top$ of at observation point at position $\mathbf{A}_v$ remain constant during the image sequence.

10.3 Experimental Results

In order to evaluate the estimation accuracy of the proposed monocular visual odometry algorithm, we applied it to each image pair of a set of 1000 synthetically generated image pairs with the following image dimensions: 688 pixel $\times$544 pixel. Each image pair consists of two images captured by a rover at time t_{k-1} and time t_k, where the camera is rigidly attached to the rover and its coordinate system coincides with the rover's coordinate system. For each image pair, the shape, the texture and the

observation points of the ground surface, as well as its position and orientation at time t_{k-1} with respect to the camera, including the rotation and translation parameters which describes the rovert's 3D motion from time t_{k-1} to time t_k are exactly known. The rover's rotation and the rover's translation from time t_{k-1} to time t_k are different in all image pairs and their parameters are uniformly distributed in interval (-10.0 to +10.0) m for the translation and in interval (-5.0 to +5.0)° for the rotation. The experiment was performed on a iMac with an Intel Core i5 at 3.1 GHz and 12.0 GB RAM.

At a camera noise variance of 10, the Table 10.1 depicts the absolute estimation error of the translation parameter along the X axis $|E_{\Delta T_X}|$, along the Y axis $|E_{\Delta T_Y}|$ and along the Z axis$|E_{\Delta T_Z}|$. This gives an average of 0.00432 m and 0.00111 m for the mean and the standard deviation of the absolute estimation error of the translation parameters, respectively. Taking into account that the rover's translation along each axis was uniformly distributed in interval (-10.0 to +10.0) m, an average of the absolute estimation error of the translation parameters of approximately 0.005 m is an excellent indicative of the high accuracy achieved by the proposed algorithm in the estimation of the rover's translation parameters.

At a camera noise of 10, Table 10.2 depicts the absolute estimation error of the rotation parameter around the X axis $|E_{\Delta\omega_X}|$, around the Y axis $|E_{\Delta\omega_Y}|$ and around the Z axis $|E_{\Delta\omega_Z}|$. This gives an average of 0.00337° and 0.00198° for the mean and the standard deviation of the absolute estimation error of the rotation parameters, respectively. Taking into account that the rover's rotation around each axis was uniformly distributed in interval (-5.0 to +5.0)°, an average of the absolute estimation error of the rotation parameters of approximately 0.004° is also an excellent indica-

Table 10.1 Absolute estimation error of the translation parameter along the X axis $\left|E_{\Delta T_X}\right|$, along the Y axis $\left|E_{\Delta T_Y}\right|$ and along the Z axis $\left|E_{\Delta T_Z}\right|$ at a camera noise variance of 10

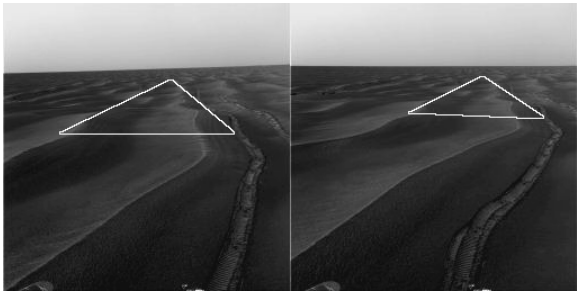
Absolute estimation error	mean (m)	standard deviation (m)
$\left\|E_{\Delta T_X}\right\|$	0.00203	0.00124
$\left\|E_{\Delta T_Y}\right\|$	0.00065	0.00033
$\left\|E_{\Delta T_Z}\right\|$	0.01027	0.00177
Average	0.00432	0.00111

Table 10.2 Absolute estimation error of the rotation parameter around the X axis $\left|E_{\Delta\omega_X}\right|$, around the Y axis $\left|E_{\Delta\omega_X}\right|$ and around the Z axis $\left|E_{\Delta\omega_X}\right|$ at a camera noise variance of 10

Absolute estimation error	mean (°)	standard deviation (°)
$\left\|E_{\Delta\omega_X}\right\|$	0.00537	0.00316
$\left\|E_{\Delta\omega_Y}\right\|$	0.00324	0.00234
$\left\|E_{\Delta\omega_Z}\right\|$	0.00150	0.00042
Average	0.00337	0.00198

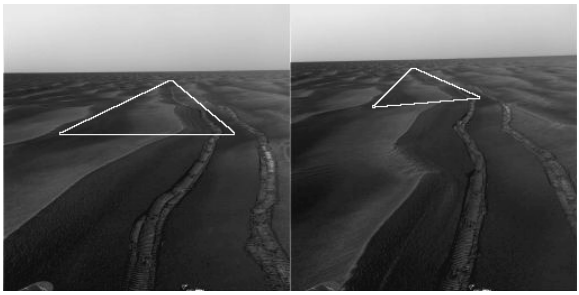
tive of the high accuracy achieved by the proposed algorithm in the estimation of the rover's rotation parameters.

In a second experiment, we applied the proposed visual odometry algorithm to 6 different real image sequences with image size of 256 pixel $\times$ 256 pixel. They were taken by the left navigation camera of the MER rover Opportunity at different dark chocolate brown flat desolate martian landscapes. All original images are courtesy of NASA/JPL-Caltech. Because in this experiment the shape, the texture and the observation points of the planet's ground surface model as well as its relative pose with respect to the camera are not known at the beginning of the image sequence, they are first initialized as described in subsection 10.2.5. The experimental results revealed an average processing time of 0.1 sec/frame, as well as that the tracking was never lost. We have also generated an image sequence where a virtual planar object (triangle) has been integrated into the original image sequence and animated in a way that it looks like it is rigidly attached to the ground surface using the



(a) (b)

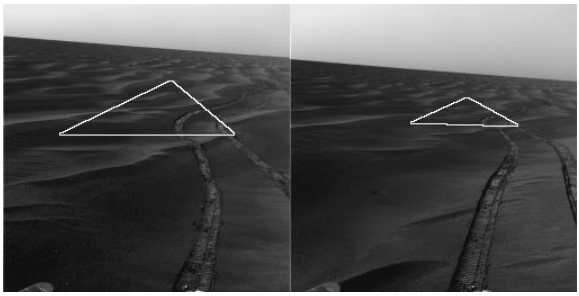
Fig. 10.4 aaPerspective projection of a triangle into (a) the first and (b) the last image of the test sequence #2 captured by the rover Opportunity on Mars. Since the triangle is motion compensated with the negative of the accumulated rover's 3D motion estimates, it seems to be glued to the ground surface. All original images are courtesy NASA/JPL-Caltech.



(a) (b)

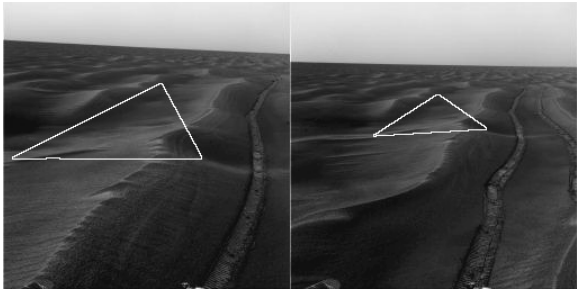
Fig. 10.5 Perspective projection of a triangle into (a) the first and (b) the last image of the test sequence #3 captured by the rover Opportunity on Mars. Since the triangle is motion compensated with the negative of the accumulated rover's 3D motion estimates, it seems to be glued to the ground surface. All original images are courtesy NASA/JPL-Caltech.

estimated frame to frame 3D motion parameters of the rover. Actually this is achieved by compensating the motion of the triangle with the negative of the accumulated rover's 3D motion estimates and then projecting the triangle into the image plane for each image of the sequence. A subjective analysis of the results revealed that the triangle really seems to be glued to the ground surface, which is also an indicative of the excellent reliability of the proposed algorithm. Figs. 10.4.(a)-10.7.(a) and Figs. 10.4.(b)-10.7.(b) depict the animated triangle projected into the first image and into the last image of the test image sequence #2, #3, #4, and #6, respectively. We have also achieved similar results by applying the proposed visual odometry algorithm to both the 400 frames of a real aerial image sequence and the 400 frames of a real infrared aerial image sequence, the two with image size of 688 pixel $\times$ 544 pixel.



(a) (b)

Fig. 10.6 Perspective projection of a triangle into (a) the first and (b) the last image of the test sequence #4 captured by the rover Opportunity on Mars. Since the triangle is motion compensated with the negative of the accumulated rover's 3D motion estimates, it seems to be glued to the ground surface. All original images are courtesy NASA/JPL-Caltech.



(a) (b)

Fig. 10.7 Perspective projection of a triangle into (a) the first and (b) the last image of the test sequence #6 captured by the rover Opportunity on Mars. Since the triangle is motion compensated with the negative of the accumulated rover's 3D motion estimates, it seems to be glued to the ground surface. All original images are courtesy NASA/JPL-Caltech.

10.4 Summary and Conclusions

In this contribution, we have showed that the rover's 3D motion can be also precisely estimated from measurements of frame to frame intensity differences provided by a monocular video camera. Although a validation of its performance in a real rover test bed is still missing, the experimental results so far revealed that it is promising candidate to be used to improve long-range autonomous navigation of rovers on a planetary surface. This because it could help when wheel odometry and traditionally stereo visual odometry have failed, as well as it could be used to validate the stereo visual odometry estimate or to generate a better estimate by statistically combining the wheel odometry estimate, the stereo visual odometry estimate and the estimate of the proposed algorithm using sensor fusion techniques. It is also an excellent candidate for lighter rover systems or entomopters where space, weight and power supply are really very limited. We even believe that the proposed algorithm could have a similar error growth than that achieved with the stereo visual odometry algorithm used by the MER rovers and the MSL rover and it could be more effective when the distance to the scene is much larger than the stereo baseline. Additionally, it has the advantage of being able to operate just with a single monocular video camera, which consumes less energy, weight less and needs less space than a stereo video camera. We are also convinced that the proposed algorithm could be computationally more efficient than the stereo visual odometry because it does not depend at all on any correlation based template matching for operation.

Our intention is not to replace the stereo visual odometry but to show that monocular visual odometry based on frame to frame intensity differences is another reliable and precise way for odometry estimation that can be merged with other sensors to improve the long rage autonomous navigation of the current and future rovers, airplanes and flapping insect robots for planetary exploration.

10.5 Future Work

In the future work, the proposed algorithm will be implemented and tested in a real rover platform Clearpath Robotics Husky A200 (see Fig.10.8), as well as a set of experiments will be performed in a simulated Martian landscape to assess its reliability, robustness, error growth, power consumption and overall size. The obtained results will be compared with those achieved by the stereo visual odometry and existing monocular SLAM algorithms.

Currently, we are working on the development of the rover's real time image acquisition system consisting of three IEEE-1394 cameras installed onto the rover's mast, each looking down toward the ground surface in three different directions: to the front, rear and left side parts of the rover, respectively (see Figs. 10.8 and 10.9). During the experiments only one of those three cameras will be used for monocular visual odometry. The system is being developed under Ubuntu 12.04.2 LTS, ROS Fuerte and the programing language C. The image acquisition system will also correct, in real time, the radial and tangential distortions due to the camera lens.

Fig. 10.8 Real rover test bed Clearpath Robotics Husky A200, which is currently used to validate the proposed monocular visual odometry algorithm. The rover's real time image acquisition system consists of three IEEE-1394 cameras installed onto the rover's mast, each looking down toward the ground surface in three different directions: to the front, rear and left side parts of the rover, respectively. During the experiments only one of those three cameras will be used for monocular visual odometry.

Fig. 10.9 First tests of the real rover test bed Husky A200 and its image acquisition system on campus of the University of Costa Rica

References

1. Shirley, D., Matijevic, J.: Mars Pathfinder Microrover. Autonomous Robots 2, 283–289 (1995)
2. NASA, Mars Exploration Rover Launches. In: Press Kits. Jet Propulsion Laboratory, California Institute of Technology (2003)
3. NASA, Mars Science Laboratory Launch. In: Press Kits. Jet Propulsion Laboratory, California Institute of Technology (2011)
4. Braum, R.: Design of the ARES Mars Airplane and Mission Architecture. Journal of Spacecraft and Rockets 43(5), 1026–1034 (2006)
5. Michelson, R.C.: Slow flight in the lower Mars Atmosphere in support of NASA science missions. In: Proceedings of International Unmanned Vehicles Workshop (2010)
6. Ali, K., Vanelli, C., Biesiadecki, J., et al.: Attitude and Position Estimation on the Mars Exploration Rovers. In: Proceedings of IEEE International Conference on Systems, Man, and Cybernetics (2005)
7. Eisenman, A., Liebe, C., Perez, R.: Sun sensing on the Mars exploration rovers. In: Proceedings of IEEE Aerospace Conference (2002)
8. Moravec, H.: Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover. PhD Thesis. Stanford University (1980)
9. Matthies, L., Shafer, S.: Error Modeling in Stereo Navigation. IEEE Journal of Robotics and Automation Ra-3(3), 239–248 (1987)
10. Matthies, L.: Dynamic stereo vision. PhD Thesis. Computer Science, Carnegie Mellon (1989)
11. Olson, C., Matthies, L., Schoppers, M., et al.: Robust stereo ego-motion for long distance navigation. In: Proceedings of IEEE Conference on Computer Vision Pattern Recognition (2000)
12. Maimone, M., Cheng, Y., Matthies, L.: Two years of visual odometry on the Mars Exploration Rovers. Journal of Field Robotics 24(3), 169–186 (2007)
13. Lacroix, S., Mallet, A., Chatila, R., et al.: Rover self localization in planetary-like environments. In: Proceedings of International Symposium on Artificial Intelligence, Robotics and Automation in Space (1999)
14. Nister, D., Naroditsky, O., Bergen, J.: Visual Odometry. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (2004)
15. Comport, A., Malis, E., Rives, P.: Accurate Quadrifocal Tracking for Robust 3D Visual Odometry. In: Proceedings of IEEE International Conference on Robotics and Automation (2007)
16. Corke, P., Strelow, D., Singh, S.: Omnidirectional Visual Odometry for a Planetary Rover. In: Proceedings of IEEE International Conference on Robots and Systems (2005)
17. Strasdat, H., Montiel, J., Davison, A.: Real time monocular SLAM: Why filter? In: Proceedings of IEEE International Conference on Robotics and Automation (2010)
18. Strasdat, H., Montiel, J., Davison, A.: Scale Drift-Aware Large Scale Monocular SLAM. In: Proceedings of Robotics: Science and Systems (2010)
19. Johnson, A., Goldberg, S., Cheng, Y., et al.: Robust and Efficient Stereo Feature Tracking for Visual Odometry. In: Proceedings of IEEE International Conference on Robotics and Automation (2008)
20. Horn, B., Schunck, B.: Determining Optical Flow. Artificial Intelligence 17, 185–203 (1981)
21. Lucas, B., Kanade, T.: An Iterative Image Registration Technique with an Application to Stereo Vision. In: Proceedings of the 7th International Joint Conference on Artificial Intelligence (1981)

22. Martinez, G.: Analyse-Synthese-Codierung basierend auf dem Modell bewegter dreidimensionaler, gegliederter Objekte. PhD Thesis. Leibniz Universitaet Hannover, Germany (1998)
23. Martinez, G., Kakadiaris, I., Magruder, D.: Teleoperating ROBONAUT: A case study. In: Proceedings of British Machine Vision Conference (2002)
24. Adiv, G.: Determining Three-Dimensional Motion and Structure from Optical Flow Generated by Several Moving Objects. IEEE Transactions on Pattern Analysis and Machine Intelligence 7(4), 384–401 (1985)
25. Heeger, D., Jepson, A.D.: Subspace methods for recovering rigid motion I: Algorithm and Implementation. International Journal of Computer Vision 7(2), 95–117 (1992)
26. Horn, B., Weldon, E.J.: Direct methods for recovering motion. International Journal of Computer Vision 2, 51–76 (1988)