STEREO PROCESSING OF HRSC MARS EXPRESS IMAGES BY SEMI-GLOBAL MATCHING

Heiko Hirschmüller[1], Helmut Mayer[2], G. Neukum[3] and the HRSC CoI-Team

[1] Institute of Robotics and Mechatronics, German Aerospace Center (DLR) Oberpfaffenhofen
[2] Institute for Photogrammetry and Cartography, Bundeswehr University Munich
[3] Institute of Geological Sciences/Planetology, Freie Universität Berlin
heiko.hirschmueller@dlr.de

Commission IV/7

KEY WORDS: HRSC, Mars Express, Stereo Matching, Semi-Global Matching, Digital Elevation Model

ABSTRACT:

The High Resolution Stereo Camera (HRSC) is used on-board the Mars Express probe for imaging the Martian surface from orbit. Furthermore, the DLR employs an airborne version HRSC-AX for capturing earth's landscape and cities. The challenge of creating accurate elevation models of terrestrial scenes, which include buildings with sharp boundaries, has lead to the Semi-Global Matching (SGM) method at the Institute of Robotics and Mechatronics of DLR. The strength of SGM is the accurate, pixelwise matching at full image resolution. The efficiency of the approach allows the processing of huge areas on a processing cluster. The success has lead to the question, whether processing of Martian data yields equally good results. This paper describes the used HRSC projection model, SGM based stereo matching, and the creation of elevation models and true ortho images. Also, specific adaptations for processing Martian HRSC data are discussed. The method has been tested on 4 orbits and compared to MOLA data. The results suggest that the method is equally suited for terrestrial and Martian HRSC data.

1 INTRODUCTION

The High Resolution Stereo Camera (HRSC) has been developed by the Institute of Planetary Research of DLR (Wewel et al., 2000) and is used on-board the Mars Express probe for imaging the Martian surface from orbit. The airborne version HRSC-AX is used at DLR for capturing earth's landscape and cities from flight altitudes between 1 500m to 5 000m, with ground resolutions of 15-20cm/pixel. The challenge of creating accurate Digital Elevation Models (DEM) through stereo matching in the presence of sharp depth discontinuities (i.e., due to buildings) has lead to the Semi-Global Matching (SGM) method (Hirschmüller, 2005) at the Institute of Robotics and Mechatronics of DLR. The advantages of applying this method to HRSC data of urban terrain has been demonstrated earlier (Hirschmüller et al., 2005). Figure 1 shows a few examples.

Preprocessing of terrestrial HRSC data is done at the Institute of Planetary Research of DLR, resulting in radiometrically and geometrically corrected 2D images. SGM based stereo matching and the creation of DEMs, true ortho and tilted ortho images, i.e., for generating facade textures of buildings, are done at the Institute of Robotics and Mechatronics of DLR on a processing cluster. Within the last year, more than 20000km^2 of terrestrial HRSC data have been processed at resolutions of 15-20cm/pixel.

The success of terrestrial HRSC processing has lead to the question whether SGM based processing of Martian HRSC data would yield equally good results.

2 PROCESSING MARS EXPRESS HRSC DATA

The following sections present the creation of Digital Elevation Models and True Ortho Images from preprocessed image and navigation data. Section 2.1 describes the data that has been used as input for the subsequent process. The projection model for reprojecting and interpreting the images is detailed in Section 2.2. The projection model is employed in Section 2.3 for calculating epipolar lines. Section 2.4 describes the Semi-Global Matching method that is used for finding corresponding points along epipolar lines. Finally, Section 2.5 discusses the creation of elevation models and true ortho images from the matching results.

2.1 Input Data

The Martian HRSC is equipped with nine sensor arrays, which are arranged orthogonally to the flight direction at different angles for permitting stereo analysis. The recorded pixels of all nine sensor arrays are combined into nine 2D images, such that each image line corresponds to the sensor line that has been recorded at a certain time. The image information is preprocessed, e.g., radiometrically corrected, at the Institute of Planetary Research of DLR.

The intrinsic camera geometry is given by pre-calibration. The extrinsic camera parameters, i.e., position and orientation, of each image line is measured by instruments on-board Mars Express and processed at the Institute of Planetary Research of DLR. An optional refinement of the extrinsic camera parameters is done through bundle adjustment (Spiegel et al., 2005) at the Institute of Photogrammetry and Geo Information of University Hannover and Photogrammetry and Remote Sensing of Technical University Munich.

The image information together with the intrinsic and extrinsic camera parameters are the only input of the following process.

2.2 Projection Model

The projection model uses the intrinsic and extrinsic camera parameters for describing the geometric projection and reconstruction of points in a Euclidean world coordinate system. The projection model has already been described in the context of processing terrestrial HRSC data (Hirschmüller et al., 2005), but it has been refined.

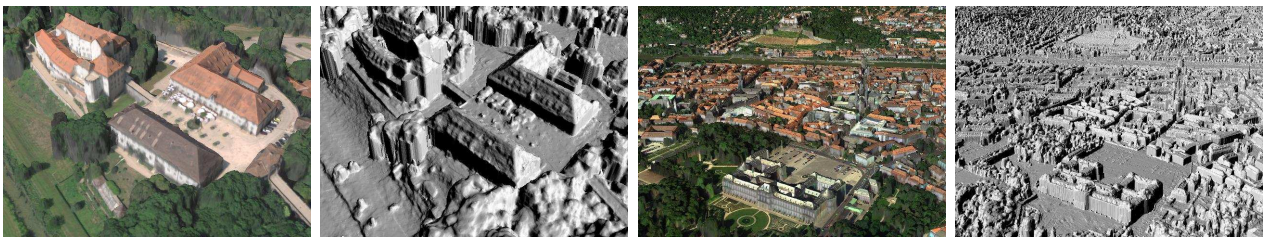

Figure 1: Textured and shaded DEM of Seefeld Castle (15cm/pixel) on the left and Würzburg (20cm/pixel) on the right.

The intrinsic camera model of each sensor array is shown in Figure 2a. Due to lens distortion, the straight sensor array usually appears curved on the focal plane. All intrinsic parameters including lens distortion are modelled by describing the positions x_k, y_k of all k pixels of the seonsor array individually on the focal plane, which is at a distance of f from the optical center. Thus, the 3D location of a pixel k is $\mathbf{S}_k = \begin{pmatrix} x_k & y_k & f \end{pmatrix}^T$ in the camera coordinate system.

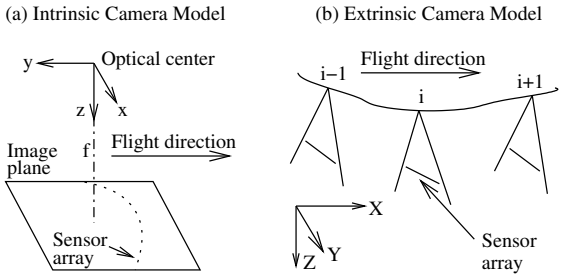

Figure 2: Geometric camera model.

A 2D image is captured line by line, while the camera moves. For terrestrial applications, a linear movement cannot be guaranteed while flying with an airplane, e.g. due to athmospheric disturbances. Therefore, it is assumed that the path is only roughly a straight line, the speed is not constant and the orientation can change slightly (Fig. 2b). This flexible model may not be necessary for modeling the movement in the orbit of Mars. Nevertheless, the general extrinsic model is used, which relies on knowing the orientations R_i and locations $\mathbf{T}_i$ at all capturing positions i with high accuracy.

This leads to the relationship (1) between a world point $\mathbf{P}$ and the kth pixel of the ith capturing position.

$$\mathbf{P} = sR_i\mathbf{S}_k + \mathbf{T}_i \tag{1}$$

Equation (1) is used for projecting all pixels $\mathbf{S}_k$ of a sensor array at all camera positions R_i, $\mathbf{T}_i$ onto a common focal plane I. For large scale, e.g., terrestrial applications, the common focal plane is often approximated by the surface of the reference ellipsoid that is extended such that it corresponds locally to the mean terrain height (Hirschmüller et al., 2005). However, the Mars Express images can cover a larger part of the surface, which cannot be treated as flat any more. Therefore, the plane that is tangential to the Martian reference ellipsoid at approximately the center of the covered area is chosen as common focal plane.

The projection is calculated as the intersection of (1) with the chosen focal plane. This leads to an irregular distribution of projected points. The values at regular grid positions are calculated as linear interpolations of nearby pixels. Orientation changes that destroy the order of projected pixels, e.g., camera position j in Fig. 3, are treated by removing disturbing camera positions and

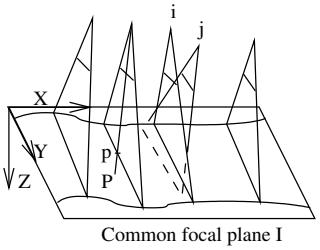

Figure 3: Reprojection of all pixels onto a common focal plane.

their projected pixels, though this does not occur for the relatively smooth Mars data.

The projected images can be seen as if they come from a new camera, with an optical center that moves over the common focal plane I. Due to reprojection of pixels, it is not necessary any more to know the original position of the pixels S_k. Instead, the shape of the sensor array is described by,

$$x'_k = k\Delta x + x_1 \text{ (with } x_1 \text{ as the first pixel in the array)} \tag{2a}$$
$$y'_k = \text{ such that it approximates the shape of } x_k, y_k. \tag{2b}$$

For simplifying subsequent calculations, the local camera coordinate system is chosen such that the height of I is 0. This constraint defines the transformations $G_R()$ and $G_T()$ for converting the viewing directions and optical center positions into the local camera coordinate system.

$$R'_i = G_R(R_i) \qquad \mathbf{T}'_i = G_T(\mathbf{T}_i) \tag{3}$$

A point P' in the camera coordinate system is projected onto the focal plane of the optical center i, by

$$\begin{pmatrix} u_i(\mathbf{P}') \\ v_i(\mathbf{P}') \\ f \end{pmatrix} = \frac{1}{s} R_i'^{-1}(\mathbf{P}' - \mathbf{T}'_i). \tag{4}$$

The distance r_i to the sensor array of the optical center i is defined using (2), by

$$r_i(\mathbf{P}') = y'_k - v_i(\mathbf{P}'), \text{ with } k = \frac{u_i(\mathbf{P}') - x_1}{\Delta x}. \tag{5}$$

The optical center i with $r_i(\mathbf{P}') = 0$ would *see* the point $\mathbf{P}'$ *exactly*. Generally, the position is in an interval $[i, i+1]$ with $r_i \geq 0$ and

$r_{i+1} < 0$. The interval is found by a binary search in $O(\log_2 n)$ steps. Removing overlapping projections, e.g., j in Fig. 3, ensures a sorted list, which is required for the binary search. The sign of r_i determines the search direction for the next step.

The position $C(\mathbf{P}')$ of the optical center for $\mathbf{P}'$ is calculated similarly to pixel values at fractional positions through linear interpolation, by

$$C(\mathbf{P}') = \mathbf{T}'_i + (\mathbf{T}'_{i+1} - \mathbf{T}'_i)\frac{r_i(\mathbf{P}')}{r_i(\mathbf{P}') - r_{i+1}(\mathbf{P}')}, \quad \text{for } i, \text{ such that } r_i \geq 0 \text{ and } r_{i+1} < 0. \tag{6}$$

The determination of the optical center permits the projection of a world point $\mathbf{P}$ onto an image point $\mathbf{p}$ by calculating the intersection of $\mathbf{P}' = G_T(\mathbf{P})$, $C(\mathbf{P}')$ at $z = 0$, i.e.,

$$\mathbf{p} = f_{proj}(\mathbf{P}') = \frac{C_z}{C_z - P'_z}\begin{pmatrix} P'_x - C_x \\ P'_y - C_y \end{pmatrix} + \begin{pmatrix} C_x \\ C_y \end{pmatrix}, \quad \text{with } \mathbf{C} = C(\mathbf{P}'). \tag{7}$$

Similarly, the world point $\mathbf{P} = G_T^{-1}(\mathbf{P}')$ is reconstructed from a given pixel $\mathbf{p}$ at the given distance z, i.e.,

$$\mathbf{P}' = f_{rec}(\mathbf{p}, z) = \frac{C_z - z}{C_z}\begin{pmatrix} p_x - C_x \\ p_y - C_y \\ -C_z \end{pmatrix} + \begin{pmatrix} C_x \\ C_y \\ C_z \end{pmatrix}, \quad \text{with } \mathbf{C} = C(\begin{pmatrix} p_x & p_y & 0 \end{pmatrix}^T). \tag{8}$$

2.3 Calculation of Epipolar Lines

A pixel $\mathbf{p}_1$ and the corresponding optical center $C_1(\mathbf{p}_1)$ define a straight line, which contains the world point $\mathbf{P}$ that is projected on $\mathbf{p}_1$. The projection of this line into a second image is called epipolar line. The projection of $\mathbf{P}$ in the second image must be on the epipolar line (Figure 4). It is formally defined as,

$$\mathbf{p}_2 = e_{12}(\mathbf{p}_1, d) = f_{proj,2}(f_{rec,1}(\mathbf{p}_1, -d\Delta z)). \tag{9}$$

The disparity d controls the position on the epipolar line. The constant Δz is set such that a disparity step of 1 causes a mean translation of 1 pixel on the epipolar line. For aerial HRSC applications, the epipolar lines are curved, similarly to the flight path that is influenced by athmospheric disturbances. However, even if the flight path is straight, which can be used as an approximation in orbital applications, epipolar lines are not straight, but

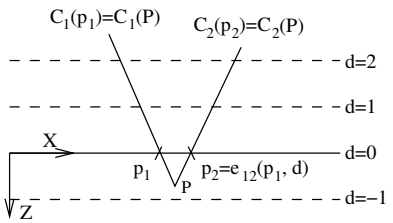

Figure 4: Calculation of epipolar lines.

in general hyperbolas (Gupta and Hartley, 1997), due to the parallel projection in flight direction and the perspective projection orthogonally to it.

The point by point calculation of epipolar lines reduces the search range for stereo matching to a minimum. The efficiency of this approach is optimized by calculating only a few points on the line and assuming piecewise linearity in between.

2.4 Semi-Global Matching

Stereo matching is conducted using the Semi-Global Matching (SGM) method (Hirschmüller, 2005), which aims to determine the disparity image D, such that the cost $E(D)$ is a minimum.

$$E(D) = \sum_{\mathbf{p}}(C(\mathbf{p}, D_\mathbf{p}) + \sum_{\mathbf{q}\in N_\mathbf{p}} P_1 T[|D_\mathbf{p} - D_\mathbf{q}| = 1] + \sum_{\mathbf{q}\in N_\mathbf{p}} P_2 T[|D_\mathbf{p} - D_\mathbf{q}| > 1]) \tag{10}$$

The first term of the cost function sums pixelwise matching costs $C(\mathbf{p}, D_\mathbf{p})$ at the pixel $\mathbf{p}$ with the disparity $D_\mathbf{p}$. It is based on hierarchically computed Mutual Information (MI) instead of intensity differences (Hirschmüller, 2005). This makes it robust against recording differences and illumination changes. The second term of equation (10) encourages piecewise smoothness of the disparity image, by adding a small cost P_1 for all small disparity changes. Similarly, the third term adds a higher cost P_2 for all larger disparity changes. Adding a *constant* cost for all larger disparity changes preserves discontinuities.

Finding the global minimum of equation (10) is an NP-complete problem (Boykov et al., 2001). The SGM algorithm approximates the global minimization by pathwise minimizations from 16 directions as shown in Figure 5. The minimum cost paths L_r for reaching a pixel $\mathbf{p}$ at disparity d are calculated using (10) and summed for all directions. For each pixel, the disparity is chosen that corresponds to the minimum sum of costs.

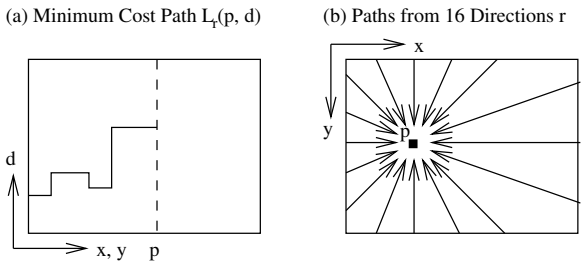

Figure 5: Calculation of disparities.

A left, right consistency check is used for eliminating wrong matches and a quadratic interpolation over neighboring costs for reaching sub-pixel accuracy. Stereo matching is done separately between nadir and one of the two stereo (s1/s2) or photometric (p1/p2) images recorded by HRSC. This results in four disparity images, which are combined pixelwise by first choosing the median and then calculating the weighted mean disparity from values within a certain distance from the median. The weights are chosen according to the angle between the Nadir and the corresponding match image, because larger angles give higher geometric accuracy, but are more difficult to match. This strategy eliminates outlier and also increases the accuracy due to the weighted mean. Finally, a filter is applied that removes isolated disparity segments below a certain number of pixel, as those segments are typically remaining outliers.

Location	HRSC Orbit number	Resolution [m/pixel]	Size [MPixel]	Projection onto plane, Time [h]	Stereo matching (SGM), Time [h]	Ortho projection Time [h]
Nanedi Valles	0894, 0905, 0927	15	894	2.5	88.0	2.0
Valles Marineris	1235	30	127	1.0	(60m/pixel) 2.0	0.3

Table 1: Processing times measured on one 2.0GHz Opteron CPU.

The complexity of SGM is just $O(ND)$, but unfortunately the memory consumption is also proportional to ND, i.e., number of pixels times the disparity range. Processing arbitrarily large images is implemented by working with image parts, i.e., tiles, that are chosen as big as possible, but small enough so that the available memory is sufficient for processing them. The tiles are defined slightly overlapping. Pixel near image borders are not taken into account, because they receive support only from one side by the global cost function. The remaining overlapping parts are linearly blended from one tile to the other for smooth transitions between tiles.

The disparity range is initially set very large and used for processing images that are downscaled by factor 16. A reduced disparity range is determined separately for each tile and used for processing higher resolution images. This is repeated until the full resolution has been reached. The disparity range determination is done during the hierarchical computation of MI. It is important to note that the only information that is used from a lower to a higher resolution scale is the refined disparity range and pixelwise cost estimation from MI. It has been purposely avoided reusing more, possibly wrong information as this could lead the matching process at a higher resolution level into a wrong direction. More details about the SGM method can be found in an earlier publication (Hirschmüller, 2005).

2.5 Creation of Elevation Models and True Ortho Images

The resulting disparity image corresponds to the nadir image of the HRSC. The geometric model of this image (Section 2.2) is complex due to a mixture of perspective and parallel projection.

The disparity image can be converted into a simple orthographic height model. Each pixel $\mathbf{p}$ of the disparity image D is reconstructed by $\mathbf{P} = f_{rec,D}(\mathbf{p}, -D_{\mathbf{p}}\Delta z)$. An orthographic projection is used, which stores each height value P_z at the pixel position P_x, P_y. Double mappings are resolved by using the height that is closest to the camera. This orthographic model also supports fusing results of different orbits, by taking the mean or median of heights, which decreases surviving outliers. Finally, gaps often appearing due to low texture are filled by inverse distance weighted interpolation from the border of each gap. The result is a Digital Elevation Model (DEM) of the scene.

The corresponding intensity or color value at P_x, P_y and the stored height value P_z are determined by bilinear interpolation in the image I at the position $\mathbf{q} = f_{proj,I}(\mathbf{P})$. Occlusions, e.g., in terrestrial data due to buildings, are detected using the DEM and interpolated using the same strategies used for the DEM. The results is a true ortho image.

3 RESULTS

The described method has been applied to 4 orbits that were provided for testing and comparing different HRSC processing strategies (Heipke et al., 2006). The extrinsic camera parameters have been refined by bundle adjustment for increasing accuracy (Spiegel et al., 2005). Only the radiometrically corrected images and the intrinsic and extrinsic camera parameters have been used as input. The DEM and ortho image have been calculated as discussed in Section 2. The parameters of the method have been adjusted for matching Martian scenes and kept constant for all orbits. The main change was to increase the penalty P_2 of equation (10) that penalizes sudden depth changes, which are common in urban terrain, but unusual in Martian scenes. Furthermore, the overlap of tiles has been increased. Both means reduce the chances of random peaks in large areas with low texture. All processing steps, after the one time adjustment of parameters, are fully automatic.

Table 1 lists the properties of all used orbits. The speed is measured on a singe CPU. However, parallelization onto a processing cluster can be done as well. The nadir, s1, and s2 images of orbit 1235 have a low contrast. These images have been adjusted by multiplying all intensity values by 2 for bringing them into the same range as p1 and p2. The remaining adjustment was left to the automatic mutual information based matching. Due to the same reason, stereo matching has been done with only half of the resolution, i.e. 60m/pixel. These were the only manual adjustments that have been applied.

3.1 Visual Inspection

The ortho images and DEMs of all four orbits are shown on the left of Figures 6 and 7. The heights of the DEMs are linearly mapped into intensity images for visualization purposes, such that the lowest height of the shown part is black and the highest is white. At this scale it can be seen that there are no big errors in the calculated DEMs.

Small parts of the ortho images and DEMs are shown on the right of Figure 6 and 7. The DEMs are mapped into intensities and shaded by an artificial light source. A visual comparison of the ortho images and the corresponding DEM shows that all features of the ortho images appear to be represented in the DEMs.

3.2 Expected accuracy and Comparison to MOLA Data

Stereo matching is performed between the nadir and s1/s2 images that are captured in an angle of $\pm 18.9^\circ$ and the p1/p2 images that are captured in an angle of $\pm 12.8^\circ$ against nadir. The resolution of the s1/s2 images is $\frac{1}{2}$ and that of p1/p2 just $\frac{1}{4}$ of the nadir resolution. Assuming a mean stereo matching error of 1 pixel in the resolution of the nadir image and the best stereo angle of 18.9°, results in a theoretical height error of 44m or 88m using a nadir resolution of 15m or 30m respectively.

For a statistical comparison of the HRSC DEMs and the Mars Orbiter Laser Altimeter (MOLA) DEM (Smith and et al., 2001), the HRSC DEM has been resampled into the grid of the MOLA data. The mean has been calculated for all points that fall into a cell and the mean difference and standard deviation to the MOLA data has been computed.

For the block of orbits 0894, 0905, 0927, the mean difference is 9m and the standard deviation is 51m, which is just a bit higher than the expected, i.e., 44m. However, comparing the orbit 1235 to the MOLA data results in a mean difference of 39m and a standard deviation of 237m. The large error can be partly due to the non-optimal image quality of that orbit, e.g., low contrast in nadir and s1/s2, as described above.

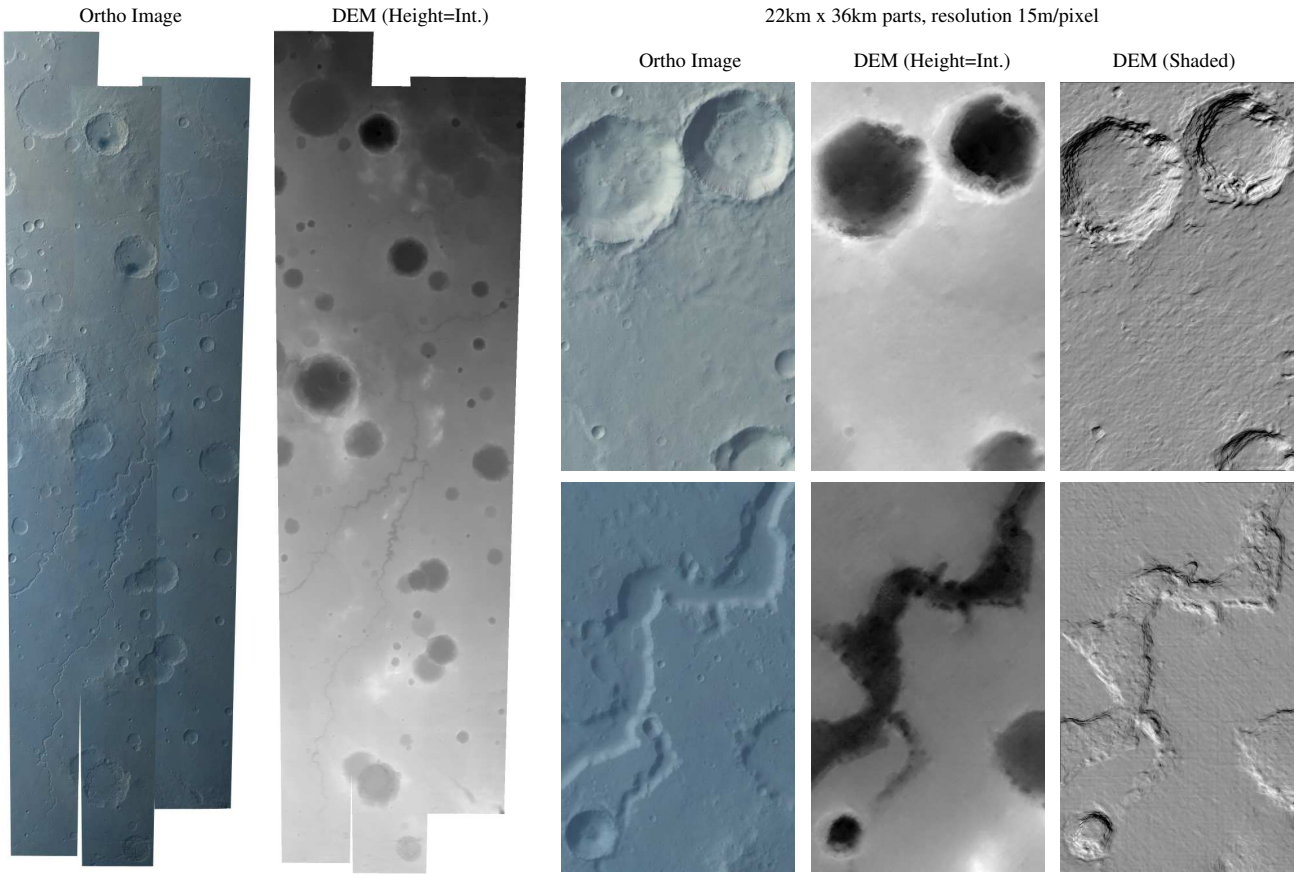

Figure 6: Mars Express orbits 0894, 0905 and 0927 sampled with 15m/pixel.

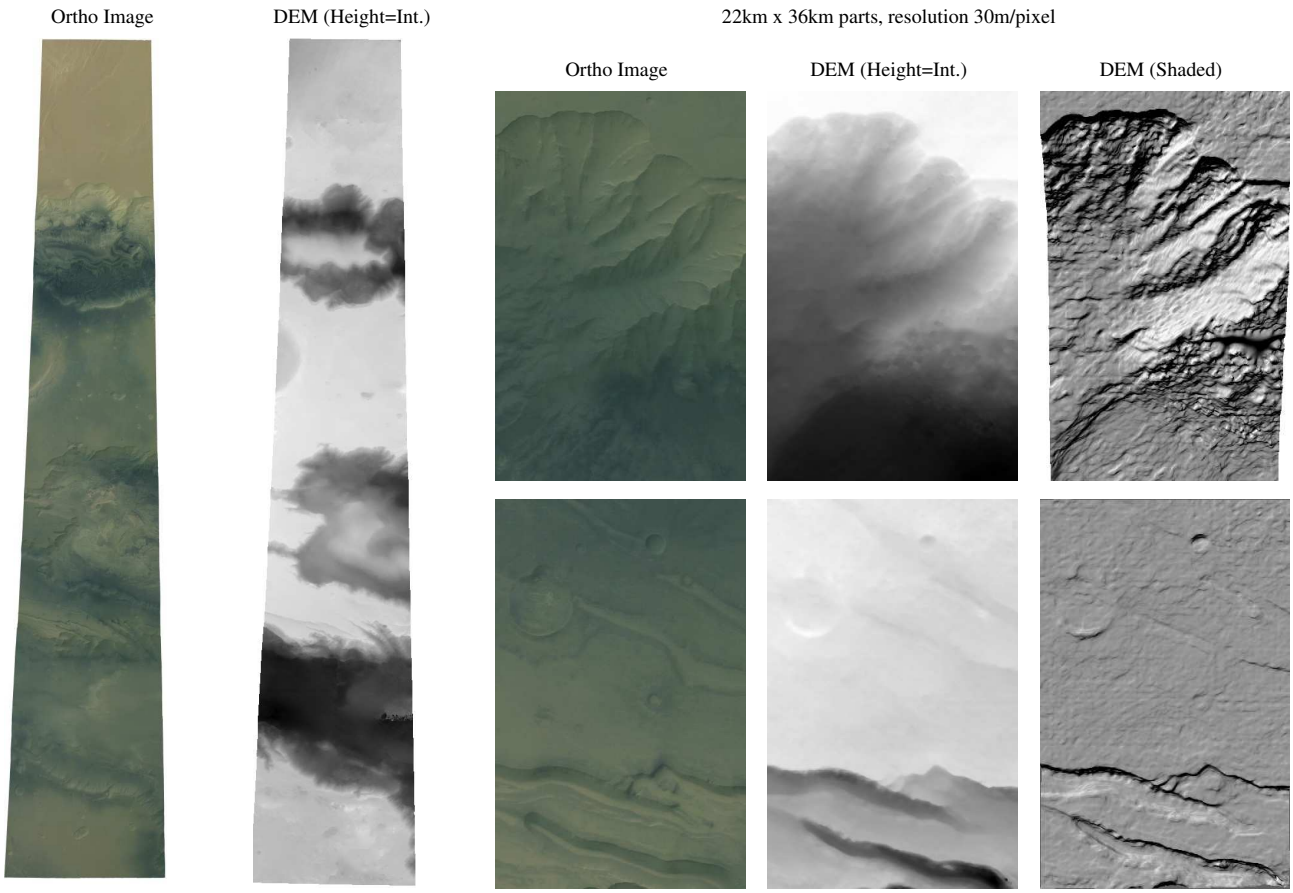

Figure 7: Mars Express orbit 1235 sampled with 30m/pixel.

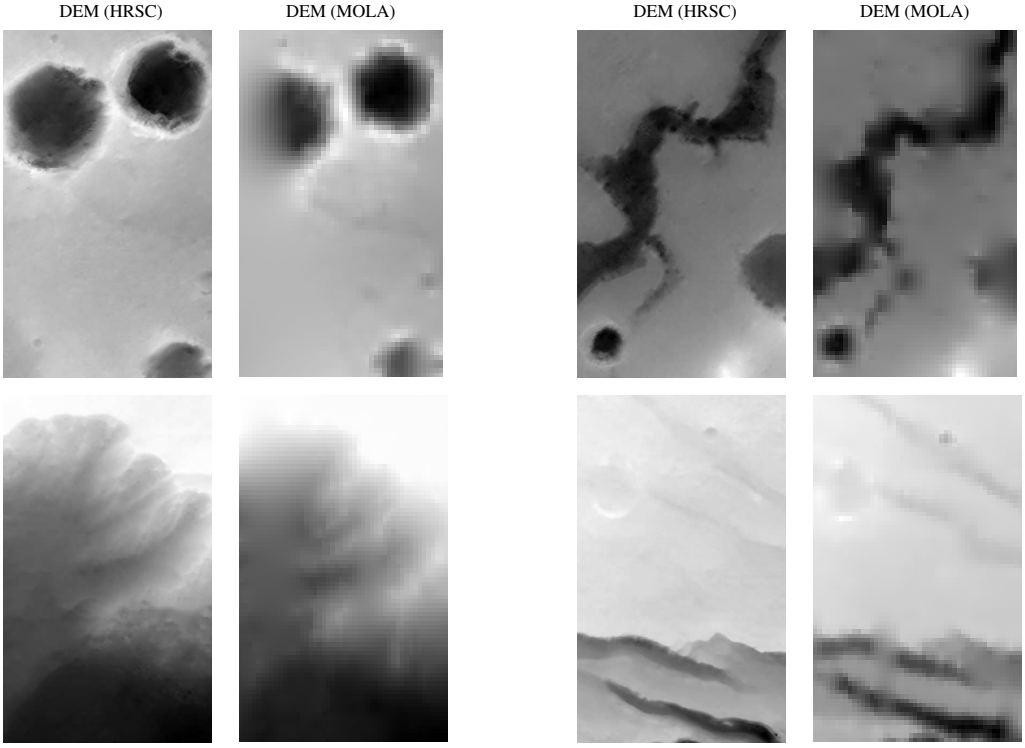

Figure 8: Visual comparison of some parts of the HRSC and MOLA DEM.

However, comparing parts of the HRSC and MOLA DEM visually reveals that several places of the MOLA DEM appear fuzzy, which is probably the result of interpolation of missing data as seen in Figure 8. This has especially a high impact for comparing orbit 1235 to MOLA, because the changes of the terrain are much larger than in the other orbits, leading to a higher standard deviation of differences. Considering this shifts the accuracy well within the expected range. Furthermore, the visual comparison of the HRSC and MOLA DEMs in Figure 8 make it clear that the HRSC DEM is much sharper, has a much higher resolution and conveys much more structural detail than the MOLA DEM.

A more extensive analysis and comparison to other methods can be found in a separate publication (Heipke et al., 2006).

4 CONCLUSIONS

SGM has originally been developed for addressing the problem of accurate matching in the presence of sharp depth changes that are typical in HRSC images of urban scenes. Martian scenes have different challenges like jagged terrain and large, almost textureless areas. The test has shown that these challenges are handled well by the presented solution. The proposed method works fully automatic and offers pixelwise matching at full image resolution. Processing is done time efficient so that large areas can be computed fast. Thus, the presented approach has not only proved to be successful for accurate HRSC processing of terrestrial data, but also for processing Martian scenes.

ACKNOWLEDGMENTS

We would like to thank Frank Scholten of the Institute of the Planetary Research of DLR for interesting discussions and for supporting the development of the terrestrial HRSC processing at the Institute of Robotics and Mechatronics of DLR through customized HRSC Level 3 data delivery as well as all other members of the Institute of Planetary Research of DLR for inspiring discussions over the last few years.

REFERENCES

Boykov, Y., Veksler, O. and Zabih, R., 2001. Efficient approximate energy minimization via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(11), pp. 1222–1239.

Gupta, R. and Hartley, R. I., 1997. Linear pushbroom cameras. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(9), pp. 963–975.

Heipke, C., Oberst, J., Attwenger, M., Dorninger, P., Ewe, M., Gehrke, S., Gwinner, K., Hirschmüller, H., Kim, J. R., Kirk, R. L., Lohse, V., Mayer, H., Muller, J.-P., Rengarajan, R., Schmidt, R., Scholten, F., Shan, J., Spiegel, M., Wählisch, M., Yoon, J.-S. and Neukum, G., 2006. The hrsc dtm test. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVI, Part 4.

Hirschmüller, H., 2005. Accurate and efficient stereo processing by semi-global matching and mutual information. In: IEEE Conference on Computer Vision and Pattern Recognition, Vol. 2, San Diego, CA, USA, pp. 807–814.

Hirschmüller, H., Scholten, F. and Hirzinger, G., 2005. Stereo vision based reconstruction of huge urban areas from an airborne pushbroom camera (hrsc). In: Proceedings of the 27th DAGM Symposium, Vol. LNCS 3663, Springer, Vienna, Austria, pp. 58–66.

Smith, D. E. and et al., 2001. Mars orbiter laser altimeter: Experiment summary after the first year of global mapping of mars. Journal of Geophysical Research 106(E10), pp. 23,689–23,722.

Spiegel, M., Schmidt, R., Stilla, U., Baumgartner, A. and Neukum, G., 2005. Improving the exterior orientation of mars express orbiter using mola data in bundle adjustment. In: AOGS 2nd Annual Meeting, Vol. 58-PS-A1013, Singapore.

Wewel, F., Scholten, F. and Gwinner, K., 2000. High resolution stereo camera (hrsc) - multispectral 3d-data acquisition and photogrammetric data processing. Canadian Journal of Remote Sensing 26(5), pp. 466–474.