FURTHER ADVENTURES IN MARS DTM QUALITY: SMOOTHING ERRORS, SHARPENING DETAILS

R.L. Kirk[a]*, D. Mayer[a], B.L. Redding[a], D.M. Galuszka[a], R.L. Fergason[a], T.M. Hare[a], K. Gwinner[b]

[a]U.S. Geological Survey, Astrogeology Science Center, 2255 N. Gemini Dr., Flagstaff AZ 86001 (rkirk@usgs.gov)
[b]German Aerospace Center (DLR) Institute of Planetary Research, Rutherfordstraße 2, D-12489 Berlin, Germany

ICWG III/II: Planetary Remote Sensing and Mapping

ABSTRACT:

We have used high-precision, high-resolution digital terrain models (DTMs) of the NASA Mars Science Laboratory and Mars 2020 rover landing sites based on mosaicked images from the Mars Reconnaissance Orbiter High Resolution Imaging Science Experiment (MRO HiRISE) camera as a reference data set to evaluate DTMs based on Mars Express High Resolution Stereo Camera (MEX HRSC) images. The Next Generation Automatic Terrain Extraction (NGATE) matcher in the SOCET SET/GXP® commercial photogrammetric system produces DTMs with relatively good (small) horizontal resolution but high error, and results are terrain dependent, with poorer resolution and smaller errors on smoother surfaces. Multiple approaches to smoothing the NGATE DTMs give very similar tradeoffs between resolution and error. Smoothing the NGATE DTMs with a 5x5 lowpass filter is near optimal in terms of both combined resolution-error performance and local slope estimation, but smoothing with a single pass of an area-based matcher, which has been the standard approach for generating planetary DTMs at the U.S. Geological Survey to date results in similar errors and only slightly worse resolution. DTMs from the HRSC team processing pipeline fall within this same trade space but are less sensitive to terrain roughness. DTMs produced with the Ames Stereo Pipeline also fall in this space at resolutions intermediate between NGATE and the team pipeline. Although DTM resolution and error each vary by a factor of 2, the product of resolution and error is much more consistent, varying by ≤20% across multiple image sets and matching algorithms. Refinement of the stereo DTM by photoclinometry can yield significant quantitative improvement in resolution and some improvement in error (improving their product by as much as a factor of 2), provided that albedo variations over distances smaller than the stereo DTM resolution are not too severe.

1. INTRODUCTION

Detailed topographic data are foundational to geoscience and engineering operations on other planets just as they are on Earth, making the assessment of digital terrain model (DTM) quality factors such as vertical precision and horizontal resolution of great interest. In the absence of ground surveys (apart from a handful of tracked landers/rovers), other references for making such assessments are needed. The High Resolution Imaging Science Experiment (HiRISE; McEwen et al., 2007), which at ~25 cm/pixel is the highest resolution Mars-orbiting camera, provides a useful reference against which to measure lower resolution imagers such as the ~6 m/pixel Context camera (C TX; Malin et al., 2007) and High Resolution Stereo Camera (HRSC, multi-line with nadir channel 12.5 m/pixel and up, stereo channels typically 2x2 averaged; Neukum et al., 2004). Downsampled to ground sample distances (GSD) appropriate to the other cameras, HiRISE DTMs have sampling-limited resolution and negligible vertical errors. The main difficulty is that most HiRISE DTMs are the width of a single stereopair, 5 km. Wider DTM mosaics have been constructed for many candidate landing sites, but such sites are by intention flat and featureless, hence poorly suited for DTM evaluation. The coverage for the Mars Science Laboratory (MSL) site in Gale crater (Golombek et al., 2012) extended onto the very rugged flank of Aeolis Mons, however, and we used it to evaluate HRSC DTMs made both at the German Aerospace Center (DLR) with the instrument processing pipeline and at the U.S. Geological Survey (USGS) with a combination of commercial and open-source software (Kirk et al., 2017). Unfortunately, the individual DTMs at Gale were registered by a non-rigorous approach (Golombek et al., 2012) with undocumented (likely not subpixel) accuracy that left significant vertical discontinuities in the mosaic. This forced us to restrict our analysis to a single HiRISE

strip and rendered our initial conclusions about DTM quality (Kirk et al., 2017; 2018) tentative. Fortunately, when similar DTMs were collected for the Mars 2020 landing site in Jezero crater (Fergason et al., 2019; 2020) more rigorous techniques were available and were used to optimize the alignment of individual DTMs and to verify their positional accuracy. The seamless HiRISE DTM mosaic also covered a much larger area (~20x18 km versus ~5x15 km at Gale) including both smooth and rugged terrains (Figure 1). Assessment of the Jezero data (Kirk et al., 2020) indicated that stereo matching errors were in the range 0.2-0.3 pixel and DTM resolution ranged from 10-20 pixels, with the USGS products having somewhat better (smaller) resolutions but slightly greater errors than those from DLR. These results were generally consistent with earlier ones for Gale (Kirk et al., 2017; 2018). In addition, a direct assessment showed that misregistration at the pixel level would not have affected the conclusions. Perhaps the most unexpected result was that the quality measures for USGS (but not DLR) DTMs varied with surface roughness: resolution was lower (better) but errors were greater on rougher terrain.

In this paper, we revisit the Jezero and Gale datasets in an attempt to answer several of the salient questions about DTM quality posed in the discussion of our previous results (Kirk et al., 2020). First, the USGS DTMs in that study were generated by using a standard process and set of parameters for automated matching (described below) that has been employed for mapping multiple landing sites (e.g., Golombek et al., 2012; Fergason et al., 2019; 2020) as well as regions of scientific interest. What tradeoffs between resolution and vertical precision are available through adjusting the matcher parameters or from postprocessing such as spatial filtering of the DTMs? Can an optimal set of parameters be identified? Does the optimal processing depend on the application, e.g., does estimating slopes over short baselines (as

---

* Corresponding author

for landing site selection) require a different approach than, say, identifying and measuring small features for geologic research? Second, how does the quality performance of other stereo matching packages and algorithms compare with the DLR pipeline and the commercial system used by the USGS? Might other software yield superior DTM quality from the same data? Third, does refining a stereo DTM by photoclinometry (shape-from-shading) yield quantitative improvements in resolution or vertical precision, or is the apparent "sharpening" of the DTM merely qualitative?

2. SOURCE DATA

2.1 Gale crater

Mapping of Gale crater included more than a dozen HiRISE stereopairs at 25 cm/pixel, covering the full landing ellipse and a substantial area of Aeolus Mons (also known informally as "Mount Sharp"; Golombek et al., 2012). The first of these pairs containing rugged terrain was designated Traverse 1 (or T1) and consists of images psp_009149_1750 and psp_009249_1750. A 15 x 6.5 km study area (latitude -4.92° to -4.67°N, longitude 137.35° to 137.46°E) within this DTM was used by Kirk et al. (2011) to assess an early multi-orbit HRSC DTM mosaic from DLR (Gwinner et al., 2010a). Subsequent comparisons (Kirk et al., 2017, 2018) used the same HiRISE data and DTMs produced by DLR and USGS from images h4235_0001_xx2 (xx = nd2, s12, s22), which has a superior signal to noise ratio (SNR). The Level 2 (radiometrically calibrated) images are available from the NASA Planetary Data System (PDS). The DLR Level 4 (single-orbit controlled; Gwinner et al., 2010b) DTM h4235_0001_dt4 that was evaluated by Kirk et al. (2020) is also in the PDS but is not used in this study. Instead, we use the USGS DTM and orthoimage as the starting point for photoclinometry.
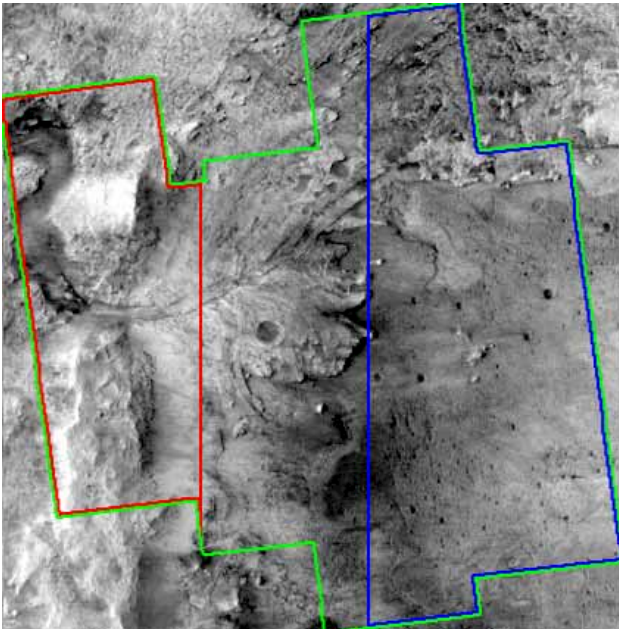

Figure 1. HRSC orthoimage of study area in Jezero crater, Mars, derived from h5270_0000 nadir image at 50 m GSD. Area covered by HiRISE reference DTM mosaic (RMS slope 3.38°) is outlined in green, western rough (11.29°) subarea in red, and eastern smooth (3.38°) subarea in blue. Quality statistics for these areas are shown in Figure 3. Equirectangular projection with north at top. Area shown is 7 x 13.75 km, centered near -4.8°S, 137.8°E.

2.2 Jezero crater

To support landing site selection, planning, and onboard navigation during landing for Mars 2020, the USGS produced DTMs from multiple HiRISE and CTX stereopairs, then coregistered them and made DTM mosaics as summarized below (Fergason et al., 2019; 2020). We used the mosaics rather than individual DTMs for this paper. The study area is defined by the HiRISE coverage, centered on the Jezero delta near latitude 18.49°N, longitude 77.41°E (Figure 1). The data cover about 290 km^2 within a 20 x 20 km region, five times the area studied at Gale. The HRSC product from DLR is an unreleased multi-orbit mosaic prepared for the Mars 2020 project, based on a subset of the available HRSC coverage for quadrangle MC-13E. An unreleased Level 5 (multi-orbit controlled; Gwinner et al., 2016) DTM covering the entire quadrangle was studied by Kirk et al. (2020) and results from it are shown in the figures below for purposes of comparison. Level 2 images h5270_0000_xx2 were used to produce the USGS DTM as described below.

3. MAPPING METHODOLOGIES

3.1 Stereo: SOCET SET

Production of stereo DTMs at the USGS with SOCET SET is described in detail by Kirk et al. (2020). Here, we summarize the process, focusing on spatial registration and stereo matching. DTM production at the USGS used the open-source ISIS system (Sides et al., 2017) for data preparation and the commercial stereo software SOCET SET ® (Miller and Walker, 1993; 1995) (more recently, SOCET GXP) from BAE Systems. Following Kirk et al. (2020), we refer to the resulting products as "USGS DTMs" although we have now generated additional DTMs at the USGS by using the Ames Stereo Pipeline (ASP; Beyer et al., 2018) and ISIS as described below. At Gale, DTMs were controlled individually by bundle adjustment based on a sparse set of ground points measured manually by reference to Mars Orbiter Laser Altimeter (MOLA; Smith et al., 2001) data. The DTMs were registered to create the mosaic used for site selection and mission planning by "rubber sheet" deformation based on manually measured tiepoints. Because Gale contains diverse albedo and slope features that are nearly ideal for testing the effects of sharpening DTMs with photoclinometry, for this study we aligned a single, undeformed HiRISE DTM (Traverse 1 or T1 in Golombek et al., 2012) to the USGS HRSC DTM by point-cloud fitting in ASP.

The ASP point-cloud fitting module *pc_align* has also been used to generate dense sets of pseudo-ground control points for more recent DTMs including Jezero, by adjusting a set of tiepoints to align with a MOLA or HRSC base. This process, described by Kirk et al. (2017), is significantly faster and more accurate than the earlier use of sparse manual ground points. At Jezero, the CTX stereopairs were controlled to HRSC in this way and all HiRISE images were then controlled simultaneously to CTX, with tiepoints between overlapping stereopairs. The HRSC images used in the USGS DTM were similarly controlled to CTX. To further improve on the spatial accuracy achieved by bundle adjustment, *pc_align* was used to fit the CTX data to the DLR HRSC base and then the HiRISE and USGS HRSC DTMs to CTX. Cross correlation of the orthoimages, which were transformed identically to the DTMs, was used to verify that the spatially resolved offsets between products were small compared to the DTM GSD (Fergason et al., 2019; 2020).

Stereo matching for the USGS DTMs was performed with the Next Generation Automatic Terrain Extraction (NGATE) module of SOCET Set/GXP (Zhang, 2006; Zhang et al., 2006). NGATE uses both area- and feature-based matching to estimate heights "at every pixel" (i.e., on a grid with GSD equal to the mean of the input images). Estimates can also be based on multiple images, but these are considered pairwise (including both orderings of each pair) rather than in a multi-ray intersection calculation. The dense height estimates from multiple pairs and algorithms are then combined by robust filtering to generate an output at the desired GSD. Four levels of additional smoothing of this output (none, low, medium, and high) may be selected. Unsmoothed NGATE DTMs from Mars images tend to have a "blocky" appearance (Kirk et al., 2008) that causes root mean squared (RMS) surface slopes to be overestimated. Rather than using the smoothing options built into NGATE, our standard procedure has been to apply a single pass of the older, area-based Adaptive Automatic Terrain Extraction (AATE) algorithm (Zhang and Miller, 1997) to the NGATE result. Our rationale for using AATE, which reduces the appearance of blockiness, is that smoothing as a side effect of area-based matching might be expected to introduce fewer errors than a filter that does not consider the image content. We have therefore used the NGATE+AATE approach for all products, including the Gale and Jezero DTMs evaluated by Kirk et al. (2020). In this paper we make the first quantitative assessment of the results of the AATE post-pass and compare it with the four levels of NGATE smoothing and with postprocessing the unsmoothed NGATE DTM with lowpass boxcar filters of various sizes.

3.2 Stereo: Ames Stereo Pipeline

Preparation of the HRSC images for processing in ASP closely followed the procedure described for CTX images by Mayer et al. (2018). The images were first controlled in ASP based on control points collected automatically and fitted to the CTX base. A sparse preliminary DTM was then collected and the images were orthorectified onto this for matching. ASP provides the capability to do multi-image matching and multi-ray intersection. This capability was used to produce the majority of the DTMs studied from the triplet of stereo and nadir channels. For some matcher parameters DTMs were also produced from the pair of stereo channels alone.

The default stereo matching process in ASP is "block matching," which consists of normalized cross correlation (NCC) followed by subpixel (SP) refinement (Beyer et al., 2020). We investigated the quality of DTMs produced without SP refinement, with refinement by parabolic interpolation of the NCC results, and by the default method of Bayes EM weighted affine adaptive window correlation. We did not test the simpler, nonadaptive affine SP algorithm that is also offered. The kernel sizes for the NCC and affine SP steps are the parameters most obviously related to DTM resolution. We therefore investigated combinations of odd NCC kernel sizes from 3–25 pixels and odd SP sizes from 7–25 pixels.

ASP also includes several matchers based on Semi-Global Matching (SGM; Hirschmüller, 2008), which attempts to find a distribution of stereo disparities that is consistent with correlation results. By optimizing a robust "cost function," SGM attempts to interpolate smoothly where the correlations are noisy (e.g., featureless areas), yet allows sharp jumps in disparity where justified by the data. The method thus offers the hope of better performance on both bland and steep terrains. SGM has been applied successfully to HRSC images (Hirschmüller et al., 2006) and was evaluated in the HRSC DTM comparison (Heipke et al., 2007). ASP offers both a generalization of the original SGM algorithm and More Global Matching (MGM; Facciolo et al., 2015), also known as smooth SGM. We tested both algorithms over the supported kernel sizes (odd, 3–9 pixels). We also compared results with and without SP refinement (which is built into SGM and MGM and does not require specifying a separate kernel size) and with two different cost functions. Cost mode 3 uses the census transform (Zabih and Woodfill, 1994) and cost mode 4 uses the ternary census transform (Hua et al., 2016).

3.3 Photoclinometry

To investigate the effects of refining a stereo DTM by photoclinometry, we used the ISIS 2 software and methods described by Kirk et al. (2003a) to process a subarea of the HRSC nadir orthoimage circumscribing the HiRISE T1 DTM (Figure 2). In particular, photometric parameters appropriate to the Martian surface at the observed phase angle were estimated and a uniform contribution to the image radiance from atmospheric haze was estimated and subtracted as described in that paper; both steps are essential if photoclinometry is to yield a quantitatively accurate result. The algorithm takes a single image as input and
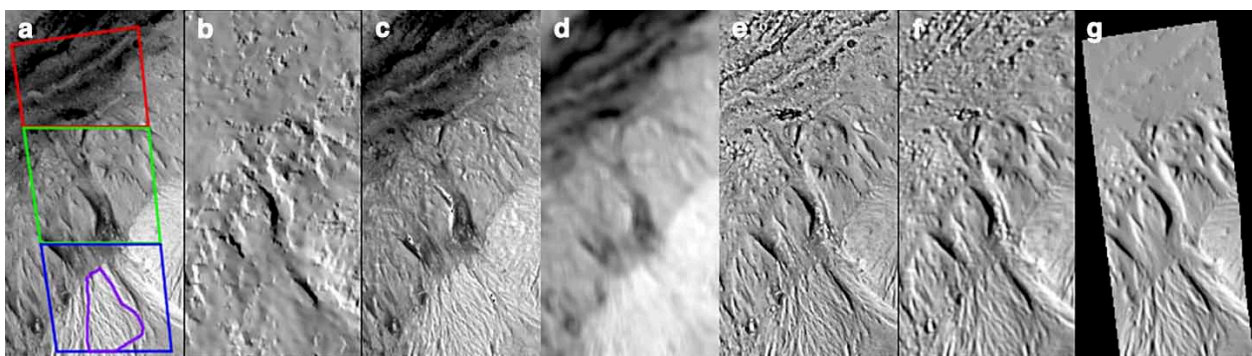

Figure 2. Processing of HRSC data in Gale crater, Mars to correct albedo variations in HRSC image and refine stereo DTM by photoclinometry. (a) HRSC orthoimage of H4234_0001 nadir image at 50 m GSD. Quality statistics for areas outlined in color are shown in Fig. 4. (b) Simulated image from USGS HRSC stereo DTM. (c) Ratio of orthoimage (after haze subtraction) to simulated image contains albedo variations and shading due to topography not resolved in stereo DTM. (d) Ratio c, smoothed at the DTM resolution of 7 posts, contains albedo variations over larger distances than this. (e) Ratio of orthoimage a (haze subtracted) to smooth albedo d contains shading plus albedo variations smaller than 7-pixel resolution. This image was used as input to photoclinometry to refine the stereo DTM. (f) Simulated image from stereo DTM refined by photoclinometry after 16 iterations. (g) Simulated image from HiRISE stereo DTM downsampled to 50 m GSD. Photoclinometry result f appears similar except where image e contained uncorrected albedo variations. All panels are in Equirectangular projection with north at top, 21.1 x 21.35 km, centered near 18.5°N 77.4°E.

necessarily assumes that the surface albedo is uniform, which is obviously not true for most of Mars. The T1 area, in particular, contains terrains ranging from a dune field in the north with low slopes and extreme albedo variations to the rugged slopes of Aeolis Mons with more subtle albedo variations in the south. We therefore utilized the HRSC stereo DTM to correct for albedo variations down to its limit of resolution as described by Kirk et al. (2006) and shown in Fig. 2. The orthoimage (with haze subtracted) was divided by a synthetic image based on the stereo DTM and smoothed at the DTM resolution to yield an albedo map. The orthoimage was then divided by the albedo map to yield a corrected image containing topographic shading with only residual albedo variations over distances smaller than the DTM resolution. The corrected image was the input for photo-clinometry, with the stereo DTM serving as the starting approximation. The photoclinometry equations were solved iteratively by under-relaxation, with relinearization after every relaxation step (Kirk et al., 2003a). The result was saved and evaluated after 1, 2, 4, … 128 steps.

3.4 Quality Assessment

Our quality assessment begins with downsampling a HiRISE reference DTM (or DTM mosaic) from its original GSD of 1 m/post to the appropriate GSD and reprojecting it to match the "target" DTM to be evaluated. We then smooth the reference DTM with boxcar lowpass filters of 3x3, 5x5, etc., posts, and measure the RMS difference between the target DTM and each of these smoothed products. Inverse quadratic interpolation of the difference at odd-integer filter sizes then yields the filter size at which HiRISE best fit the other data set (a measure of resolution) and the minimum difference (a measure of vertical precision). For brevity, we refer to these estimates as "resolution" and "error" below. Normalizing the resolution to the image GSD and converting the vertical error to an inferred matching error in pixels (which accounts for stereo convergence as well as GSD) yields consistent results for different cameras (Kirk et al., 2020).

4. RESULTS

The first test reported is a check of our process for quality evaluation. We generated synthetic target DTMs by smoothing the reference DTM with Gaussian lowpass filters of varying widths, then comparing them to the reference DTM smoothed as usual with boxcar filters. The resulting "resolution" (best-fit boxcar filter width) was consistently related to the Gaussian width, equal to 1.5 times the full width at half maximum (FWHM), or 3.6 standard deviations. Thus, although the numerical value of "resolution" may depend on the choice of filter shape and how its width is expressed, results based on different definitions can be relied on as not merely correlated, but directly proportional to one another. Conversely, using the same filter shape to compare the resolution of different DTMs should yield quantitatively reliable comparisons, although care is still needed if resolution expressed as a filter size is to be compared to other measures such as the smallest recognizable craters or other features (Kirk et al., 2020).

Figure 3a shows the matching errors for Jezero DTMs plotted as a function of resolution. The most obvious conclusions are first, that, for a given area, the different methods of smoothing the basic NGATE DTM follow nearly the same trend, and second, that areas of differing roughness follow similar but offset trends. As we previously reported (Kirk et al., 2020), the DLR DTM shows much less variation in error and almost none in resolution as a function of terrain. It is plausible that the error and resolution of the basic NGATE result vary with slope as a result of the nonlinear process by which height estimates "at every pixel" are

combined to yield the output DTM. For linear averaging of independent height samples (e.g., by a smoothing filter), one would expect the RMS error to vary inversely with filter size so that their product is constant. Figure 3b shows that the product of resolution and error for smoothed NGATE DTMs is roughly constant but displays a weak minimum whose value is almost independent of slope. This minimum occurs near the NGATE+ AATE and 5x5 boxcar results and between the low and medium smoothing options. The NGATE+AATE approach that we have used extensively is thus close to a loosely defined "sweet spot" in terms of overall DTM quality. Somewhat disappointingly, postprocessing with a simple 5x5 lowpass filter yields better resolution at a similar error level to AATE despite not taking the images into account and is undoubtedly faster to compute. Compared to these products, the DLR Level 5 DTM is significantly smoother but its product of error and resolution is similar. Smoothing the NGATE DTM with a 13x13 boxcar filter results in quality statistics very similar to the DLR product, though slightly more terrain dependent.
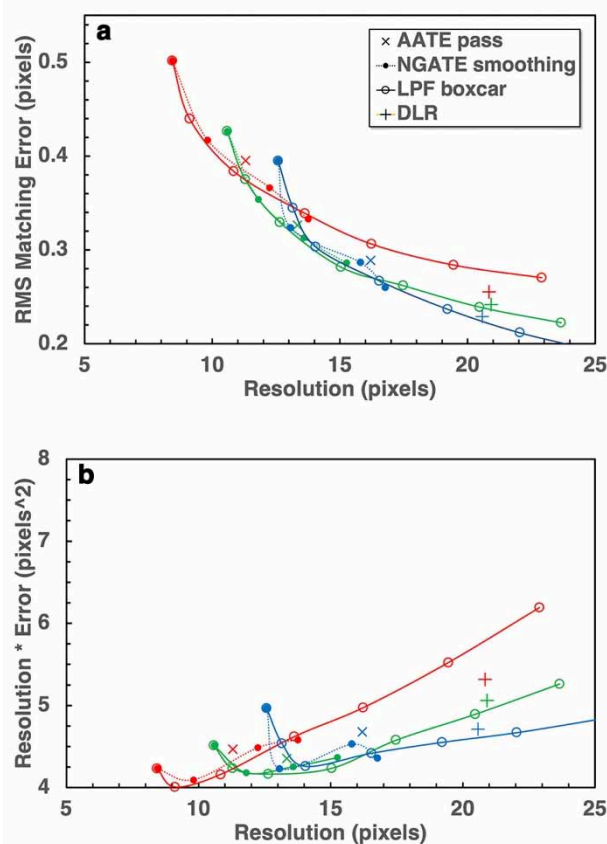

Figure 3. Quality factors for Jezero HRSC DTMs, scaled to be independent of image GSD and stereo convergence. "Resolution" is the best-fit filter width, scaled to stereo channel GSD. Matching error is a GSD-independent measure of RMS vertical error. (a) Results from commercial SOCET Set/GXP matcher NGATE. Large solid circles represent the basic NGATE output. Other symbols show the effects of various smoothing approaches (see text for full description) plus DLR Level 5 product, as listed in key. (b) As (a), showing the product of resolution and error, which would be constant for averaging of independent height estimates and varies much less than error or resolution alone. In both panels, colors correspond to the subareas of different roughness outlined in Fig. 1.

Estimating slopes over small horizontal baselines is an important application of planetary DTMs and is crucial for landing site

selection and validation. We therefore calculated the RMS adirectional slope on a one-post (50 m) baseline for the suite of USGS HRSC DTMs and compared the results to slopes computed from the reference DTM. Regardless of terrain, the "blocky" NGATE DTM substantially overestimates the RMS slope as a result of its large vertical errors. Smoothing the DTM decreases the slope estimates, bringing them into better agreement with the true slopes. No one smoothing solution yields the correct RMS slope for all three study areas; this is unsurprising given the terrain-dependent behavior built into NGATE. The 5x5 lowpass filter arguably comes close to being optimal in terms of consistency. It overestimates slopes of 7° and 11° by less than 0.8°, and the 3.4° slope of the Mars 2020 landing area by only 2°. Comparable slopes are obtained with AATE, consistent with the slope-baseline analysis of a subarea by Kirk et al. (2020), which showed that the NGATE+AATE DTM overestimated the true slopes on the Jezero rim by a consistent but small amount over a wide range of baselines, and agreed well with the slopes at Gale. Slope errors of a degree or two are small enough not to interfere with distinguishing safe from hazardous landing sites (Golombek et al., 2012). Overestimating slope hazards is clearly preferable to underestimating them in this application.

We also evaluated Jezero DTMs generated with NGATE (no smoothing) and NGATE+AATE but omitting the nadir image and using only the two stereo channels. The estimated resolution for the triplet is about 80% that for the pair, which matches the dense grid spacing used by NGATE, given that this equals the mean of the input image GSDs, and the GSD of the nadir channel is half that of the stereo channels. Introducing the nadir image reduces (improves) the error level to about 75% of that for the pair. This is less than the improvement that would be expected if the three pairs formed (nadir-stereo 1, nadir-stereo 2, and stereo 1-stereo 2) were statistically independent, but it clearly represents a simultaneous improvement in resolution and error, rather than the trade between the two that smoothing the DTM provides.

Results for DTMs produced with the ASP block matcher are shown in Figure 4a, with the NGATE and DLR results from Fig. 3b for comparison. The models differ mainly in error. Resolutions range only from 17 to 20 pixels rather than tracking the kernel size as the smoothed NGATE models do with filter size. Not surprisingly, subpixel interpolation reduces errors compared to the NCC result with no refinement, but results from the adaptive affine algorithm are better yet. Visual inspection shows that the larger errors in the models with interpolated or no SP refinement mostly take the form of local "blunders" (small patches much larger residuals than typical). The size and abundance of such blunders decrease as the kernel size is increased. Contouring artefacts, which might be expected for both whole-pixel matching and interpolation, were not observed. If the affine subpixel refinement is not used, increasing the NCC kernel size reduces the error up to a point, after which it tends to increase (worsen) resolution with little effect on the product. Varying the affine SP kernel size for fixed NCC kernel yields a similar "L-shaped" curve. The best resolution occurs at a SP kernel size of 13 pixels and has nearly the same error regardless of the NCC kernel size. SP refinement appears to produce the same results as long as elevations computed in the NCC step are sufficiently close to correct. Note that the best block matcher result falls directly on the curve for lowpass filtered NGATE models, close to the 9x9 filter result. Resolution and error vary only slightly with terrain roughness, as shown in Fig. 4a for the best model (results for the other algorithm and parameter choices show similar variation but are omitted for clarity). Slopes computed from the best block matching DTM underestimated the true values on all terrains by amounts ranging from 0.4° to 1.2°.
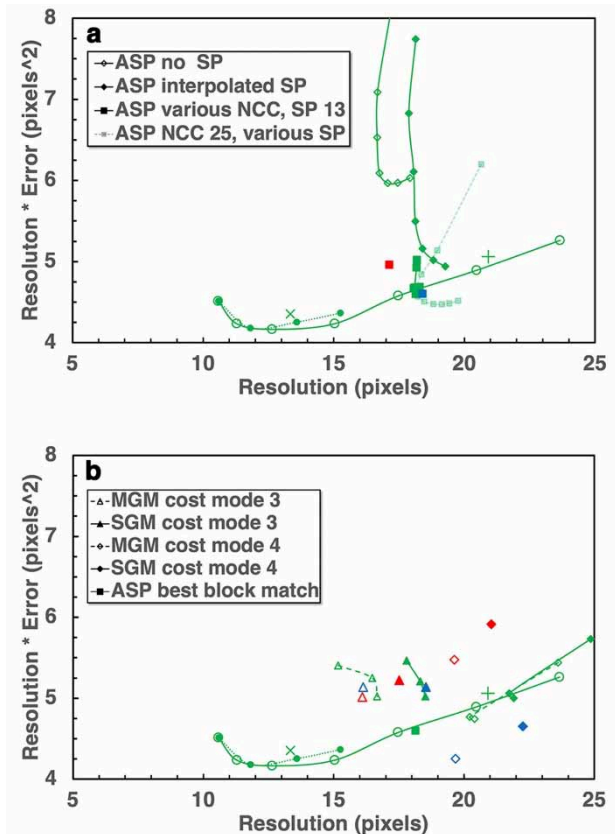

Figure 4. Quality factors for Jezero HRSC DTMs produced by using the Ames Stereo Pipeline. Product of resolution and matching error is plotted as in Fig. 3b. Colors correspond to the areas of different roughness outlined in Fig. 2. For clarity the results for rough and smooth subareas are shown only for the best DTM from each algorithm (NCC kernel 25 pixels and SP kernel 13 for block matching, kernel 9 pixels for SGM/MGM). (a) Results for block matching. For NCC with no subpixel refinement or parabolic interpolation, the product of resolution and error decreases at constant resolution as NCC kernel size is decreased, then remains almost constant resolution increases slightly. With adaptive affine subpixel refinement, varying SP kernel size yields similar behavior, but NCC kernel size has little effect. (b) Results for the ASP SGM and MGM matchers with two cost functions (see text). Results for the full area are shown for kernel sizes 5, 7, and 9 pixels, with the product of resolution and error decreasing with increasing kernel size in all cases (kernel size 3, not shown, yielded much larger errors).

Results for the ASP SGM and MGM algorithms are shown in Figure 4b. The NGATE, DLR, and best ASP block matcher results are included for comparison. DTMs produced without subpixel refinement and those made with kernel size 3 are not shown because their errors were much larger. For clarity, results for the smooth and rough terrains are shown only at kernel size 9, which produced the best product of resolution and error in each case. As expected, the ternary census model (cost mode 4) produces better results on smooth terrain but worse results on rougher terrain, whereas the terrain dependence with mode 3 is weak. The average resolution and error for mode 4 are similar to those for the DLR pipeline. MGM yields slightly better resolution but larger errors (similar resolution-error product) than SGM. The visual appearance of the SGM models is also smoother, which is opposite to what is described by Beyer et al. (2020). Contouring artefacts, which were present in parts of the SGM models studied by Heipke et al. (2007) were not observed.

"Cloth-like" textures, with artefacts along the cardinal and diagonal directions, were present in the DTMs made with smaller kernels but largely absent from the best models made with a 9-pixel kernel. The MGM algorithm with cost mode 3 over-estimated slopes on smooth terrain by 2° and yielded the correct average for the full study area, but slope estimates for all other combinations of algorithm and region were underestimated by amounts ranging from 0.6° to 1.5°.
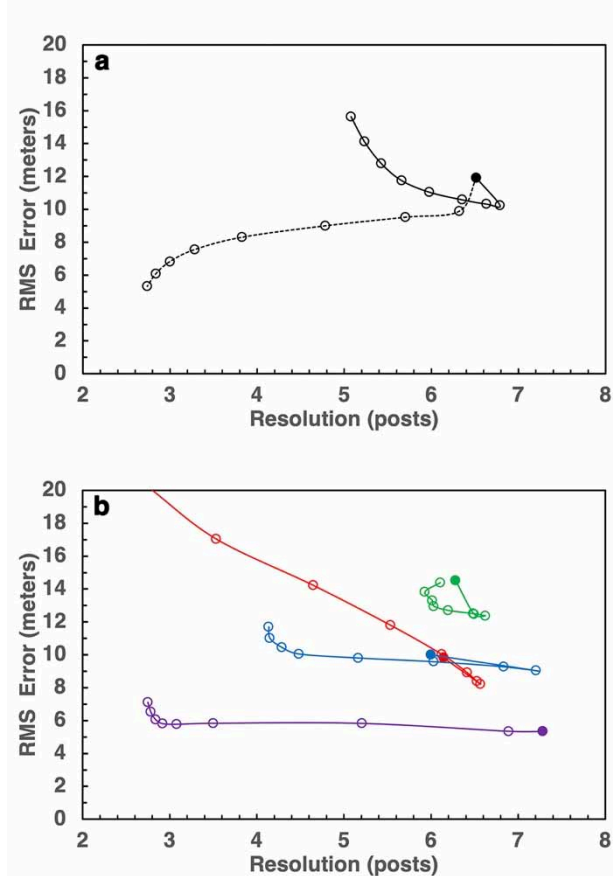

Figure 5. Quality factors for Gale HRSC DTMs refined by photoclinometry. Note that the normalization to image GSD (appropriate to stereo and used in Fig. 3a) is not used here. Best-fit filter width (resolution) is normalized to the DTM and orthoimage GSD and vertical error is in meters as measured. Solid circles show values for starting stereo DTM; open circles show results after 1, 2, 4, … 128 iterations. (a) Results for full area covered by the HiRISE T1 DTM. Solid line is for actual image data with albedo variations corrected based on stereo DTM, dashed line is for a synthetic image with uniform albedo computed from the reference DTM. (b) Results for iteration with the real image for subareas of differing slope and albedo variation outlined in Fig. 2a.

The results of refining the NGATE+AATE DTM by photoclinometry are shown in Figure 5. Note that, although the axes portray resolution and error as in Fig. 3a, the scaling appropriate to stereo is not used. Instead, resolution is scaled to the GSD of the orthoimage used to refine the DTM (same as the DTM post spacing) and vertical errors in meters are shown. In interpreting these results it is helpful to understand that the iterative relaxation process adjusts slopes across one pixel at a time, so slopes on an N-pixel baseline are only beginning to adjust after N iteration steps and generally require several times N steps to converge. Because the goal is to refine details smaller than the stereo DTM resolution (~7 posts) a few times 7 iterations are likely to be needed, and this is observed in the figure. The

resolution-error behavior can be divided into three phases. In the first few iterations, the error decreases and the resolution generally increases, similar to the effect of DTM smoothing seen in Fig. 3a. This phase represents the rapid relaxation of localized artefacts created by the stereo matcher. In the second phase, resolution decreases (improves), but after 16-32 iterations (i.e., roughly 2-4 times the initial resolution) the third phase is reached, during which the resolution remains nearly constant. Whether the error level decreases, stays the same, or increases during the second and third phases depends on the severity of uncorrected albedo variations. Figure 5a compares the behavior averaged over the whole test area for refinement based on the real image with that for a synthetic image calculated from the reference DTM. When the synthetic image, which includes no albedo variations and perfectly matches the assumed photometric behavior, is used, the error decreases in phase 2 and continues to decrease slowly in phase 3 when the resolution has reached its minimum. This minimum is ~2.5 posts rather than 1 post because computing the synthetic image requires interpolating and thus smoothing the reference DTM data. When the real image is used, errors increase slightly in phase 2 and more dramatically in phase 3. Figure 5b shows how the error is affected by albedo variations in real data. In the northern part of the study area, where relief is subtle and albedo variations are pronounced even after correction, the DTM error increases in both phases 2 and 3. In the central and southern thirds, the terrain becomes rougher and residual albedo variations less prominent. In these areas, the error remains constant as resolution improves, then increases in the final phase. When the statistics are limited to a small, rugged region with nearly uniform albedo, the behavior is similar but the resolution decreases even more, to nearly the level seen with synthetic data. Even for this subarea, however, the error increases in phase 3. The likely explanation is that even small albedo fluctuations (or even noise in the image) cause "stripes" (troughs and ridges in the direction of illumination) to form (Kirk et al., 2003b). These artefacts "grow" away from their starting points at albedo fluctuations as iteration proceeds and thus introduce errors to an increasing fraction of the DTM.

5. DISCUSSION

The results just described provide at least partial answers to many of the questions we posed previously (Kirk et al., 2020). First, quality measures for DTMs produced with four independent matching approaches (the DLR pipeline, NGATE, and the ASP block and SGM/MGM matchers) are *broadly* consistent, but resolution and error considered separately each vary by a factor of 2 or more. The product of resolution and error is much more consistent, however, varying by no more than 20% for the best results from each system. NGATE produced the highest resolution, but also the noisiest DTMs—perhaps not coincidentally because the SOCET SET/GXP systems are the only environment that allows interactive editing to selectively correct matching errors. Multiple approaches to noninteractive smoothing of the NGATE DTMs that we tested gave nearly identical results and could be adjusted to match the properties of the DLR and ASP products apart from having greater variation with terrain roughness.

Second, from a practical standpoint, postprocessing the NGATE DTM with an AATE pass as the USGS has done in most past projects, or simply by applying a 5x5 lowpass boxcar filter, is near optimal in that the product of resolution and error has a weak minimum and is independent of roughness at about this level of smoothing. Such processing also yields DTMs that overestimate surface slopes by a small and nearly consistent amount, which is appropriate for landing site studies. The appropriate amount of

smoothing for DTMs to be used in geological studies (i.e., whether to retain small features that may or may not be real versus suppressing them) may be a matter of individual taste and need. Given that post-production smoothing of the DTM gives similar results to varying the matcher parameters, a reasonable solution is to deliver DTMs with only modest smoothing and let users filter them further if desired. For ASP users, the block matcher with suitably chosen kernel sizes produces better results than the SGM/MGM algorithms. One exception is that MGM with ternary census transform (cost mode 4) produces superior results on very smooth terrains.

Third, we found that including the HRSC nadir channel image resulted in a modest but significant improvement in SOCET DTMs (about 25% reduction in the product of resolution and error) compared to using the stereo channels alone. No such improvement was seen in ASP, but it should be noted that the images were orthorectified at the stereo channel GSD, so that the nadir channel was undersampled by a factor of 2. This likely reduced or negated the channel's impact on matching errors.

Finally, our results show that refining a stereo DTM with photoclinometry can lead to significant improvement in resolution and a moderate improvement in vertical error in some cases. Using a surface photometric model that is realistic but based on the average behavior of the planet and estimating haze from shadows or by comparison to the stereo DTM suffice to obtain *quantitative* improvement of small features, not just apparent sharpening of the model. The amount of improvement depends critically on the uniformity of the surface albedo after correction based on the stereo DTM. In areas that appear completely uniform to the eye, the product of error and resolution was reduced by a factor of 2. In less uniform areas, the average improvement was less dramatic, but because DTM errors were mostly localized near uncorrected albedo features, the improvement to topography between these artefacts was likely greater. In areas with severe uncorrectable albedo variations, accuracy was worsened by photoclinometry. The gradations in albedo variation that dictate whether photoclinometry will improve or degrade the DTM are apparent to the eye, so that an informed decision on whether or not to apply the method can readily be made. A near-optimal stopping point for iteration can also be chosen in advance as 2-4 times the original DTM resolution in posts.

6. FUTURE WORK

These conclusions raise several follow-up questions. As a practical matter, can stereo matching results be improved further? Parameters controlling the DLR pipeline could be adjusted and the results compared to the standard Level 5 processing (Gwinner et al., 2016). Aspects of NGATE beyond smoothing of the final result can be controlled by choosing or even editing "strategy" files. ASP offers many features that we have not tested. For example, could its DTM filtering options improve overall quality or will they simply trade error for resolution as we have seen with NGATE? Can better results be obtained by rectifying the images before matching at the nadir rather than the coarser stereo GSD, and will the nadir channel make a more significant contribution if this is done? How must the optimal kernel sizes we found for HRSC be adjusted for images of different resolution (e.g., CTX) or of other targets such as the Moon? How does DTM quality degrade with increasing noise, lossy data compression, or (for cameras that, unlike HRSC, do not acquire both images of a pair simultaneously) with illumination differences (Kirk et al., 2016)? Use of a high-resolution reference DTM is a powerful tool with which to address all these questions.

Given that we found uncorrected albedo variations a significant obstacle to refining stereo DTMs with single-image photoclinometry, can better results be obtained with a multi-image approach that solves for both albedo and topography? Numerous formulations of the problem have been described, one of which (Alexandrov et al., 2018) is available in ASP in a form compatible with HRSC and other planetary images. Whether this capability can be tested at Jezero will depend on the availability of additional images with illumination sufficiently different from that in the HRSC observation (h5270_0000) used here.

Other questions that we posed previously (Kirk et al., 2020) remain to be studied. High-resolution reference DTMs are useful research tools for assessing matching algorithms and optimizing parameters, but because they cover limited areas they cannot be used to certify the quality of DTMs in routine production. Can indicators of quality—especially *local* quality metrics that could guide the interpretation of bumps and hollows in the DTM as either real geomorphology or mapping artefacts—be found that do *not* require a reference? The software packages used here provide information qualitatively related to errors, but not quantitative estimates of DTM precision. ASP outputs a binary "good pixel map" that simply marks pixels as successfully matched or not and a map of ray intersection error. Intersection error is also computed in the DLR pipeline. The "figure of merit" (FOM) file from NGATE combines flag values for special cases (e.g., where DTM values were interpolated for reasons such as excessive slope or were manually edited) with correlation values for most posts. We have found these products to have limited utility. In most cases only a few posts are "bad" or interpolated. Blunders by the ASP block matcher (and DLR software) are often flagged by large intersection errors, and this connection is particularly evident in extended problem areas (e.g., the shadowed side of the Jezero rim). Most posts. however, have low intersection errors (DLR and ASP) or high correlations (NGATE). Thus, most local undulations seem to be statistical inliers, variations within the vertical precision of successful image matching, rather than cases of poor or failed matching. In our experience the best way to assess the reality of individual small "features" in a DTM is to compare a terrain shaded relief to the orthoimage. A desirable goal would thus be to develop more quantitative estimators of local DTM precision from these qualitative indicators or others such as image texture. High-resolution reference DTMs of select areas are likely to be essential for calibrating any such estimators.

Finally, we noted (Kirk et al., 2020) that our assessments are based on a comparison of gridded DTMs, which are derived products interpolated from the point clouds that are the direct output of image matching and ray intersection. Direct assessment of the two-dimensional density of points, including its variations with surface texture, as well as resolution and vertical errors in such point clouds is therefore desirable and could point to ways to further improve stereo mapping accuracy. The error in ray intersection is a quality parameter that is defined for individual points (though it can be interpolated to the DTM grid) that might be useful. SOCET SET/GXP do not provide access to stereo point cloud data or intersection error, but the non-commercial DLR pipeline and ASP software packages do. With a native GSD of 1 m, HiRISE DTMs should be as useful for evaluating point clouds as they are for DTMs.

REFERENCES

Alexandrov, O., and Beyer, R.A., 2018. Multiview shape-from-shading for planetary images. *Earth Space Sci.,* 5, 652-666, doi:10.1029/2018EA000390.

Becker, K.J., et al., 2015. Criteria for automatic identification of stereo image pairs. *Lunar Planet. Sci.*, 46, 2703.

Beyer, R., et al., 2018. The Ames Stereo Pipeline: NASA's open source software for deriving and processing terrain data. *Earth Space Sci., 5*, 537-548, doi:10.1029/2018EA 000409.

Beyer, R., et al., 2020. Ames Stereo Pipeline Documentation, Release 2.7.0. (Version 2.7.0). Zenodo. http://doi.org/10.5281/zenodo.3963341.

Bland, M., et al. 2018. How well do we know Europa's topography? Assessing variability in digital terrain models. *Lunar Planet. Sci., 49,* 2193.

Facciolo, G., et al. 2015. MGM: A significantly more global matching for stereovision. In *Proc. British Machine Vision Conf.* (BMVC), BMVA Press, 90–1.

Fergason, R.L., et al., 2019. Mars 2020 terrain relative navigation support: Digital terrain model generation and mosaicking process improvement. *4th Planetary Data Workshop* (LPI Contrib. 2151), 7047.

Fergason, R.L., et al., 2020. Mars 2020 terrain relative navigation flight product generation: Digital terrain model and orthorectified image mosaics. *Lunar Planet. Sci.,* 51, 2326.

Golombek, M., et al., 2012. Selection of the Mars Science Laboratory landing site. *Space Science Reviews*, 170, 641-737, doi:10. 1007/s11214-012-9916-y.

Gwinner, K., et al., 2009. Derivation and validation of high-resolution digital terrain models from Mars Express HRSC-data. *Photogramm. Eng. Rem. Sens.,* 75(9), 1127–1142.

Gwinner, K., et al., 2010a. Regional HRSC multi-orbit digital terrain models for the Mars Science Laboratory candidate landing sites. *Lunar Planet. Sci.,* 41, 2727.

Gwinner, K., et al., 2010b. Topography of Mars from global mapping by HRSC high-resolution digital terrain models and orthoimages: Characteristics and performance. *Earth Planet. Sci. Lett.,* 294, 506–519.

Gwinner, K., et al., 2016. The High Resolution Stereo Camera (HRSC) of Mars Express and its approach to science analysis and mapping for Mars and its satellites. *Planet. Space Sci.,* 126, 93–138.

Heipke, C., et al., 2007. Evaluating planetary digital terrain models: The HRSC DTM Test. *Planet. Space Sci., 55*, 2173–2191, doi:10.1016/j.pss.2007.07.006.

Hirschmüller, H., 2008. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Machine Intelligence, 30*, 328-341.

Hirschmüller, H., et al., 2006. Stereo processing of HRSC Mars Express images by semi-global matching. *Int. Arch. Photogramm. Rem. Sens. Spatial Inf. Sci,* 36, 305–310.

Hua, H, et al., 2016. Texture-aware dense image matching using ternary census transform. *ISPRS Ann. Photogramm. Rem. Sens. Spatial Inf. Sci*. III-4, 59–66.

Kirk, R.L., et al., 2003a, Photoclinometry made simple…?, *ISPRS Working Group IV/9 Workshop "Advances in Planetary Mapping 2003",* Houston, March 2003, online at https://astropedia.astrogeology.usgs.gov/download/Research/ISPRS/Kirk_isprs_mar03.pdf.

Kirk, R.L., et al., 2003b. High-resolution topomapping of candidate MER landing sites with Mars Orbiter Camera Narrow-Angle images, *J. Geophys. Res*., 108(E12), 8088, doi:10.1029/2003JE002131.

Kirk, R.L., et al., 2006. Topomapping of Mars with HRSC images, ISIS, and a commercial stereo workstation, *Int. Arch. Photogramm. Rem. Sens. Spatial Inf. Sci*., XXXVI-4, "Geospatial Databases for Sustainable Development," Goa.

Kirk, R.L., et al., 2008. Ultrahigh resolution topographic mapping of Mars with MRO HiRISE stereo images: Meter-scale slopes of candidate Phoenix landing sites. *J. Geophys. Res*., 113, E00A24, doi:10.1029/2007JE003000.

Kirk, R.L., et al., 2011. Near-complete 1-m topographic models of the MSL candidate landing sites: Site safety and quality evaluation, *European Planet. Sci. Conf.*, 6, EPSC2011-1465.

Kirk, R.L., et al., 2016. The effect of incidence angle on stereo DTM quality: Simulations in support of Europa exploration, *ISPRS Ann. Photogramm. Rem. Sens. Spatial Inf. Sci*., III-4, 103-110, doi:10.5194/isprs-annals-III-4-103-2016.

Kirk, R.L., et al., 2017. Community tools for cartographic and photogrammetric processing of Mars Express HRSC images. *Int. Arch. Photogramm. Rem. Sens. Spatial Inf. Sci*.,XLII-3-W1, 69-76, doi:10.5194/ isprs-archives-XLII-3-W1-69-2017.

Kirk, R.L., et al., 2018. Community tools for cartographic and photogrammetric processing of Mars Express HRSC images. In *Planetary Rem. Sensing and Mapping* (B. Wu, K. Di, J. Oberst, I. Karatschevseva, eds.), Taylor & Francis, 107–124, doi:10.1201/9780429505997.

Kirk, R.L., et al., 2020. Evaluating stereo DTM quality at Jezero crater, Mars with HRSC, CTX, and HiRISE images. *Int. Arch. Photogramm. Rem. Sens. Spatial Inf. Sci*., XLIII-B3, 1129–1136, doi:10.5194/isprs-archives-XLIII-B3-2020-1129-2020.

Malin, M.C. et al., 2007. Context camera investigation on board the Mars Reconnaissance Orbiter. *J. Geophys. Res*., 112, E05S04, doi:10.1029/2006JE002808.

Mayer, D.P., 2018. An improved workflow for producing digital terrain models of Mars from CTX stereo data using the NASA Ames Stereo Pipeline. *Lunar Planet. Sci.,* 49, 1604.

McEwen, A.S., et al., 2007. Mars Reconnaissance Orbiter's High Resolution Imaging Science Experiment (HiRISE), *J. Geophys. Res.,* 112, E05S02, doi:10.1029/2005JE002605.

Miller, S.B., Walker, A.S., 1993. Further developments of Leica digital photogrammetric systems by Helava. *ACSM/ASPRS Annual Convention and Exposition Technical Papers*, 3, 256–263.

Miller, S.B., Walker, A.S., 1995. Die Entwicklung der digitalen photogrammetrischen Systeme von Leica und Helava. *Z. Photogramm. Fernerkundung,* 63(1), 4–16.

Neukum, G., et al., 2004. *HRSC: The High Resolution Stereo Camera of Mars Express*. ESA Special Publications SP-1240.

Sides, S.C., et al., 2017, THE USGS Integrated Software for Imagers and Spectrometers (ISIS 3) instrument support, new capabilities, and releases. *Lunar Planet. Sci.*, XLVIII, #2739.

Smith, D. E., and 23 others, 2001. Mars Orbiter Laser Altimeter: Experiment summary after the first year of global mapping of Mars. *J. Geophys. Res.,* 107, 23,689–23,722.

Zabih, R., and Woodfill, J., 1994. Non-parametric local transforms for computing visual correspondence. In *European Conf. Computer Vision*, 151–158. Springer.

Zhang, B., 2006. Towards a higher level of automation in softcopy photogrammetry: NGATE and LIDAR processing in SOCET SET. Paper presented at *Geocue Corporation 2nd Ann. Technical Exchange Conf.*, Nashville, Tenn., 26–27 September 2006.

Zhang, B., and Miller, S., 1997. Adaptive Automatic Terrain Extraction. In *Proc. SPIE, 3072, Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision III*, (McKeown, D.M., McGlone, J.C., Jamet, O., eds.), 27-36.

Zhang,, B., Miller, S., DeVenecia, K., and Walker, S., 2006. Automatic terrain extraction using multiple image pair and back matching. Paper presented at *ASPRS 2006 Ann. Conf.*, Reno, Nevada, 1–5 May 2006.