EVALUATING STEREO DTM QUALITY AT JEZERO CRATER, MARS WITH HRSC, CTX, AND HIRISE IMAGES

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

[a]Astrogeology Science Center, U.S. Geological Survey, 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

KEY WORDS: Mars, photogrammetry, topography, DTMs, quality control

ABSTRACT:

We have used a high-precision, high-resolution digital terrain model (DTM) of the NASA Mars 2020 rover *Perseverance* landing site in Jezero crater based on mosaicked images from the Mars Reconnaissance Orbiter High Resolution Imaging Science Experiment (MRO HiRISE) camera as a reference dataset to evaluate DTMs based on Mars Express High Resolution Stereo Camera (MEX HRSC) and MRO Context camera (CTX) images. Results are consistent with our earlier HRSC-HiRISE comparisons at the Mars Science Laboratory (MSL) *Curiosity* landing site in Gale crater, confirming that those results were not compromised by the small area compared and potential problems with spatial registration. Specifically, height errors are on the order of half a pixel and correspond to an image matching error of 0.2–0.3 pixel but estimates of horizontal resolution are 10–20 pixels. Products from the HRSC team pipeline at DLR are smoother but more precise vertically than those produced by using the commercial stereo package SOCET SET®[†]. The DLR products are also homogenous in quality, whereas the SOCET products are less smoothed and have higher errors in rougher terrain. Despite this weak variation, our results are consistent with a rule of thumb of 0.2-0.3 pixel matching precision based on many prior studies. Horizontal resolution is significantly coarser than the DTM ground sample distance (GSD), which is typically 3–5 pixels.

1. INTRODUCTION

Detailed topographic data are foundational to geoscience and engineering operations on other planetary bodies just as they are on Earth, making the assessment of digital terrain model (DTM) quality of great interest. From a user's perspective, a key question is to what extent can one expect that features seen in a DTM are present as shown, or conversely that a feature not seen is truly absent? Multiple measures of DTM quality are needed to address all aspects of these seemingly simple questions, starting with the size of features that are reliably detected (the horizontal resolution and local vertical precision of the DTM). How common are "positive" artefacts (spikes, pits, and other spurious features) and "negative" artefacts (data gaps and places where real features have been filtered out)? Can they be distinguished from real topography? These criteria all involve local properties of the DTM, which are mainly determined by the dense matching algorithm used to generate 3D points and the process by which those points are interpolated to fill the DTM grid, as well as the quality and geometric properties of the images. They form the focus of this paper. Other questions involving broader-scale reliability such as leveling and absolute accuracy depend mainly on the control process and are not considered here.

The ideal reference dataset with which to investigate all these quality measures would be independent of the "target" DTM being assessed, with high precision and absolute accuracy, high horizontal resolution and dense sampling, and broad spatial coverage. Such reference data are (barely) available for the Earth, but nowhere else. Laser altimeter data with excellent accuracy, precision, and resolution are available for the Moon, Mars, and Mercury, but tend to be sparsely sampled. Heipke et al. (2007) used Mars Orbiter Laser Altimeter (MOLA; Smith et al., 2001) altimetry to measure the absolute accuracy of DTMs produced from High Resolution Stereo Camera images (HRSC; Neukum et al., 2004) by a variety of software approaches. Because of the large footprint size (~100 m) and sparse sampling of MOLA data, Heipke et al. (2007) turned to other approaches such as DTM-image comparisons and baseline-slope statistics to evaluate DTM resolution. We apply some of their techniques in this paper.

The availability of dense topographic data registered to the target opens an additional, powerful approach to estimating resolution and precision. Such a reference DTM must have significantly better resolution and precision than that being measured, and must be coregistered to the target (which in turn means that it must not contain significant internal distortions), but absolute accuracy is not required. Simply subtracting the reference data from the target does not provide a true measure of precision, because topographic detail resolved only by the reference dataset also contributes to the difference. Instead, the reference must be smoothed to varying degrees. The amount of smoothing that yields the minimum difference provides a quantitative measure of resolution, and this minimum difference measures precision.

Suitable reference data can, for example, be obtained in the laboratory by laser ranging (Craft et al., 2017), or from using a known DTM to generate synthetic steropairs (Kirk et al., 2016). For Mars, the High Resolution Imaging Science Experiment (HiRISE), which provides the highest resolution orbital images at ~25 cm/pixel (McEwen et al. 2007), is an ideal reference for lower resolution cameras such as Context camera (CTX; ~6 m/pixel; Malin et al., 2007) and HRSC (multi-line with nadir channel 12.5 m/pixel and larger, stereo channels typically 2 x 2 averaged). Downsampled by averaging by a factor of 20 or more to resolutions appropriate to the other cameras, HiRISE DTMs have sampling-limited resolution and negligible vertical errors. The main difficulty is that most HiRISE DTMs are derived from a single stereopair, and thus are ~5 km (only 100 posts of a typical HRSC DTM) wide. Larger 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. Fortunately, recent rovers have the capability to drive out of their landing ellipses, and mapping in support of these missions has

* Corresponding author
† Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

included more rugged (and scientifically interesting) adjacent terrain as well. DTMs for the *Curiosity* landing site in Gale crater (Golombek et al., 2012) extended onto the very rugged flank of Aeolus Mons. We previously used these data to investigate the resolution and precision of HRSC DTMs made at both the German Aerospace Center (DLR) and the U.S. Geological Survey (USGS) (Kirk et al., 2011; 2017; 2018). Unfortunately, as summarized below, the process by which the Gale data were collected led to concerns about both the size of the usable study area and the accuracy with which the reference data could be registered to the target DTMs. In this paper, we apply the same method of analysis to the *Perseverance* landing site in Jezero crater (Fergason et al., 2019; 2020), where the coverage of suitable terrain is larger and all data have been registered to known, high accuracy. Our results based on HiRISE reference data are corroborated by independent approaches to estimating resolution similar to those of Heipke et al. (2007).

2. SOURCE DATA

2.1 Gale

Mapping of Gale crater eventually 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 superior signal to noise ratio (SNR). The Level 2 (radiometrically calibrated) images are available from the NASA Planetary Data System (PDS). The DLR DTM is the standard Level 4 single-strip controlled DTM h4235_0001_dt4, also in the PDS.

2.2 Jezero

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. 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. A Level 5 (multi-orbit) product covering the entire quadrangle according to the standards described in Gwinner et al. (2016) is being prepared for PDS release. Level 2 images h5270_0000_xx2 were used to produce the USGS DTM as described below.

3. MAPPING METHODOLOGIES

3.1 DLR: HRSC Team Pipeline

The HRSC processing pipeline used at DLR is entirely automated and designed to take full advantage of the multi-line scanner geometry of the images. Products are controlled to MOLA by a sequential photogrammetric adjustment (Gwinner et al., 2010b). The Gale DTM is a Level 4 product, derived from a single-orbit image set as described by Gwinner et al. (2009). Production of multi-orbit Level 5 products such as that used at Jezero is described by Gwinner et al. (2016). Dense image matching is described in detail by Gwinner et al. (2009). The images are filtered to reduce noise and compression artefacts, and orthorectified to reduce scale errors and parallax distortions. Area-based matching is applied, consisting of normalized cross-correlation followed by sub-pixel refinement by adaptive least squares. Matching is performed at a "pyramid" of resolution levels, because image quality usually varies within a single image strip (hujndreds of kilometers long) for HRSC, depending on atmospheric and illumination conditions. Points from different resolution levels are filtered separately, then combined by weighted interpolation. This procedure improves matching performance in areas of poorer image quality and thus reduces the occurrence of matching gaps, at the price of a small reduction of point precision in areas with higher image quality. Ground coordinates are computed by multi-ray intersection based on all available images, which provides information to eliminate bad matching results.

3.2 USGS: SOCET SET

Production of DTMs at the USGS uses the same software for all image types. The open-source ISIS3 system developed by the USGS (Sides et al., 2017) is used for data preparation and commercial stereo software (SOCET SET ® from BAE Systems; Miller and Walker, 1993; 1995) is employed for control and DTM generation. BAE has since introduced SOCET GXP as the successor to SOCET SET, but it uses the same adjustment and matching software so our results will apply to it also. ISIS3 was used to format output products for delivery, and for much of the analysis described below. Kirk et al. (2008) describe the mapping procedure for HiRISE in detail; that for CTX is similar. Kirk et al (2017; 2018) describe the procedures for HRSC in detail. They also describe a more efficient approach to generating ground control that was applied to the Jezero HiRISE and CTX data. The differences, and the subsequent processing done to refine and assess the geometric registration of the Jezero products (Fergason et al., 2019; 2020) are relevant to the reliability of our DTM-to-DTM comparisons, so we summarize them here.

At Gale, HiRISE (and CTX) DTMs were produced individually over a period of several years. They were controlled by using the older procedure (Kirk et al., 2008), in which ground control points were measured interactively. A small number of points in level areas were constrained in elevation, as interpolated from the MOLA data. An even smaller number of points on features common to the MOLA and image data were constrained in all three dimensions. Accuracy of the control was therefore limited by the sample and grid spacing of MOLA. Kirk et al. (2011) investigated the consistency of the overlapping HiRISE DTMs and found offsets on the order of 100 m horizontally and 10 m vertically. In addition, it was discovered late in the process that geometric distortions for CTX were not corrected properly, resulting in horizontal and vertical distortions of tens of meters. The CTX products were therefore adjusted horizontally to match the DLR HRSC data by "rubber sheet" transformation based on interactively measured ties, then the HiRISE products were warped to match CTX. This process left the horizontal accuracy of registration uncertain, but likely at the few-pixel level given the use of interactive measurements. Elevation differences were minimized by least-squares adjustment of vertical offsets to the DTMs, but mismatches of up to 10 m vertically remained as a result of uncorrected tilts (Kirk et al., 2011). These seams in the DTM mosaic led to the decision to restrict analysis to the single T1 HiRISE DTM.

The USGS HRSC DTM was produced later (Kirk et al., 2017) and controlled separately by as outlined below. Kirk et al. (2017) found this DTM to be slightly misaligned with the earlier products and shifted it to align it (by eye) with the HiRISE data. This process yielded single-pixel registration accuracy at best. The sparsity of small topographic features made it difficult to assess distortions that the warping process could have introduced into the reference DTM.

At Jezero, the new control procedures described by Kirk et al. (2017; 2018) were applied to HiRISE and CTX as well as HRSC images. In this process, a sparse set of tiepoints (but still substantially denser than the interactive measurements used in the past) is created and fitted to MOLA by using the point-cloud fitting application *pc_align* of the Ames Stereo Pipeline (ASP; Moratto et al., 2010). These points are then converted to ground control points and constrained to their fitted locations. To further improve the consistency of registration, the individual CTX DTMs were adjusted with *pc_align* to match the HRSC base (Fergason et al., 2019; 2020). Pseudo-ground-control points for the HiRISE images were then generated by fitting clouds of tiepoints to the mosaicked CTX DTM. All HiRISE images were then adjusted in a bundle based on this ground control plus image-to-image and pair-to-pair tiepoints. Finished HiRISE DTMs were further adjusted by fitting to the CTX base with *pc_align*. The orthomosaics were transformed along with the DTMs, and the horizontal registration within and between datasets was assessed by matching the images with IMCORR (https://nsidc.org/data/velmap/imcorr.html). The median of the spatially resolved offsets horizontally were 30 m for CTX to DLR Level 5 and 2 m for HiRISE to CTX (Fergason et al., 2019; 2020). The USGS HRSC DTM was controlled to the CTX base by using *pc_align*. It was fitted to the CTX data once complete but no orthoimage was produced. Its alignment with HiRISE DTM mosaic was therefore checked as part of the DTM comparison process described below and found to be accurate to 18 m. Thus, all datasets are aligned horizontally with fractional-post precision. Vertical discrepancies are also small but are unimportant since we focus on the dispersion in elevation differences rather than the mean.

DTMs were produced by using the Next Generation Automatic Terrain Extraction (NGATE) module (Zhang, 2006; Zhang et al., 2006). Regardless of the ground sample distance (GSD) selected for the output, this software performs both area-based and feature-based matching "at every pixel" (actually on a grid with spacing equal to the mean image GSD). Each possible pair of images is matched separately but the results are not combined in a multi-ray intersection calculation as for the HRSC pipeline. Instead, the multiple matching results (area- and feature-based, different image combinations, at multiple closely spaced points) are combined by robust filtering to estimate the elevation of a post in the output DTM. In our experience, this algorithm does a good job of finding the ground surface, but tends to produce a DTM that appears "blocky" on close examination (Kirk et al., 2008, their Fig. 18); both feature-based matching and nonlinear filtering would be expected to yield rather sharp jumps between small regions of the DTM. We therefore refine the NGATE DTMs by performing one pass of the older, area-based matcher Adaptive Automatic Terrain Extraction (AATE; Zhang and Miller, 1997, which smooths the DTM slightly but is more likely to maintain consistency with the images than a simple lowpass filter. SOCET SET provides tools for interactive editing, but no editing was done on the DTMs used in this study.

4. QUALITY ASSESSMENT

4.1 Comparison to HiRISE DTMs

At both Gale and Jezero, our primary approach to quality assessment was to compare DTMs post-by-post with smoothed HiRISE data. This process began with downsampling the HiRISE DTM mosaic to the appropriate GSD and reprojecting it to match the target DTM. We then smoothed the reprojected HiRISE DTM with boxcar lowpass filters of 3 x 3, 5 x 5, etc., posts, and measured the root mean square (RMS) difference between the target and each of these smoothed products. Interpolation of the difference at odd-integer filter sizes then yielded the filter size at which HiRISE best fit the target (a measure of resolution) and the minimum RMS difference (a measure of vertical precision).

Our results appear in Table 1. Filter width is expressed in DTM posts, meters, and image pixels; vertical precision is given in meters and also converted to RMS matching error ρ in pixels according to the equation $EP = \rho\ GSD\ /\ (p/h)$ where EP is the expected vertical precision (equated with the measured error) and p/h is the parallax-height ratio (Becker et al., 2015). Normalizing the results by GSD and p/h in this way allows us to compare matcher performance for the different cameras, as shown in Figure 1. We find that the results are consistent when the HRSC stereo channel GSD (rather than the nadir GSD, which is a factor of 2 smaller) is used. This is reasonable given that the stereo channels contribute the most parallax. Results (both precision and resolution) for the same camera and processing are consistent at roughly the 15% level between sites. The USGS products have significantly larger errors and better (smaller) apparent resolution than the DLR DTMs. The CTX results resemble those for the DLR DTMs (better precision and poorer resolution) despite being generated with the same software as the HRSC USGS DTMs.

Given the concerns described above about internal distortions and pixel-level misregistrations of the Gale DTMs, we investigated the sensitivity of the results to registration errors. We repeated the analysis with the HRSC DTM offset by one and two pixels from its nominal position in each cardinal direction. The RMS error at optimal smoothing showed a smooth minimum with

Site	Data	Image GSD (m/pixel)	Processing	DTM GSD (m/post)	Best-fit filter width (DTM posts)	Best-fit filter width (m)	RMS vertical error (m)	Best-fit filter width (pixels)	RMS matching error (pixels)
Gale (Aeolus Mons)	HRSC	32.4	DLR Lev 4	50	14.0	699	11.3	21.6	0.239
			USGS	50	6.88	344	13.2	10.6	0.281
Jezero	HRSC	26.9	DLR Lev 5	50	11.3	563	9.48	20.9	0.242
			USGS	50	7.18	359	12.8	13.5	0.326
	CTX	5.72	USGS	20	5.26	105	3.55	18.4	0.266

Table 1. DTM precision and resolution based on comparison to smoothed HiRISE data.
GSD = ground sample distance. RMS = root mean squared. Stereo channel GSD used for HRSC.

respect to these offsets. Interpolating, we found that the best-fit position was 6.1 m west and 17.5 m north of the nominal location, for an overall error of 18.5 m (0.37 post). Fits to the results also yielded estimates of sensitivity of the quality factors to small misalignments. A shift of one pixel increases both the optimal smoothing and the RMS error by ≤5%; two pixels of misregistration increases the values by ≤20%. Thus, registration errors of one or a few pixels between our Gale DTMs may have led us to overestimate smoothing and precision by no more than 10-20%, consistent with the ~15% agreement between the Gale and Jezero results. Given that the "rules of thumb" for these quantities (discussed below) are usually expressed with 50% uncertainty, the variation is not significant.

Dataset	Parameter	Region East	Region All	Region West	d log param d log slope
HiRISE	RMS slope	3.38	7.09	11.29	
HRSC DLR	Filter width	20.58	20.92	20.84	0.010
	Sensitivity	7.94	5.00	3.21	-0.750
	Match error	0.229	0.242	0.255	0.090
HRSC USGS	Filter width	16.20	11.34	13.30	-0.299
	Sensitivity	8.92	5.00	4.03	-0.661
	Match error	0.289	0.326	0.395	0.261
CTX	Filter width	18.34	18.40	18.46	0.006
	Sensitivity	6.81	7.11	7.90	0.123
	Match error	0.201	0.266	0.345	0.450

Table 2. DTM precision and resolution as a function of surface slope. Best-fit filter width and matching precision in pixels as in Table 1. "Sensitivity" is a measure of the relative width of the minimum in error vs. filter size.

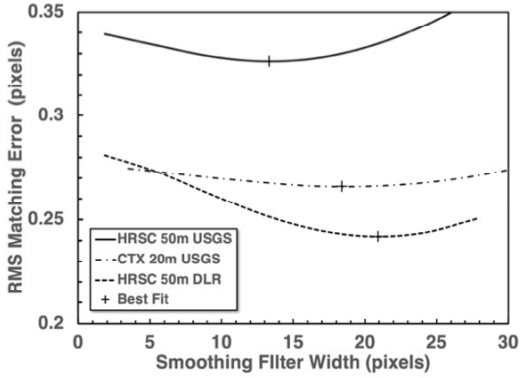

Figure 1. Estimated matching error of target DTMs as a function of smoothing of reference DTM. Values for odd-integer filter widths are connected by a smooth curve; cross indicates the interpolated minimum error at best-fit width.

An important question is whether these results are predictive for the quality of other DTMs made with images from the same camera and analyzed with the same software, or whether they are somewhat specific to the terrains studied. The consistency between the Gale and Jezero results supports the former expectation, but the diversity of terrains within the Jezero study area provides an opportunity to test it. We therefore analyzed subareas smoother and rougher than the average. The smooth region (east of 77.45° longitude) lies entirely on the Jezero crater floor east of the delta. The rough region (west of 77.35°) consists primarily of the crater rim. Table 2 gives the RMS adirectional slope for these areas and the full HiRISE DTM, measured over 50 m baselines, along with the optimal smoothing filter width and matching error (both in pixels). Also shown is a measure of the sensitivity of the error to the filter width (the change in filter width needed to increase the error by 1% from the minimum). Results for the DLR Level 5 DTM are nearly independent of terrain. Matching errors increase with roughness for the DTMs produced in SOCET SET, but much less than proportionately; the logarithmic derivatives (Table 2) are less than 0.5. The smoothness increases with roughness (again, much less than proportionately) for HRSC but, oddly, not for CTX. The sensitivity of the smoothing estimate is also nearly constant for CTX but decreases with roughness for both HRSC products.

4.2 Slopes as a Function of Baseline

A fundamentally different approach to quantifying DTM resolution is to plot the RMS slope against the horizontal baseline over which it is evaluated. Slopes at every baseline value can be calculated efficiently by using Fast Fourier Transform methods, and such baseline-slope curves have been an important tool in landing site selection for numerous missions (e.g., Kirk et al. 2008; Golombek et al. 2012). As shown in Figure 2a, the slope curves can contain information about the data collection process as well as the surface topography. This approach addresses a wide range of spatial scales and does not require a reference DTM; resolution can often be inferred from a break in the derivative of the baseline-slope curve. A reference is nonetheless useful as an indication of the true surface slope behavior, so that only deviations from this are interpreted as data effects. The main drawback of looking at baseline-slope curves is that the results can be ambiguous. For example, an upturn of the curve at short baselines due to localized artefacts (noise) in the DTM can potentially mask the leveling out that is due to limited resolution. Figure 2b (modified from Kirk et al., 2018) shows the curves for slopes along north-south baselines on Aeolis Mons, averaged over a 3.75 x 12.8 km area. The DLR DTM is increasingly deficient (relative to HiRISE) in slopes at baselines shorter than about 1500 m. The "knee" in the curve can be estimated, somewhat subjectively, by drawing tangents to the curve and locating their intersection, at about 560 m. The curve from the USGS DTM agrees closely with that from HiRISE (though the gap between curves widens at ~700 baseline) making it difficult to draw conclusion about the resolution of the USGS HRSC DTM.

Figure 2c shows north-south slopes averaged over a 3.35 x 12.8 km area on the rim of Jezero crater. Results for CTX are not shown because the area analyzed is slightly different as a consequence of the different GSDs and the requirement that the Fourier transform length be a power of 2 posts. As at Gale, the curve for the DLR DTM starts to depart from the HiRISE curve around 1500 m baseline and is flat for short baselines, with a "knee" around 490 m. The ratio of this transitional baseline to the optimal filter width is thus similar for both study areas. What is remarkable is how closely the slope curve for the optimally smoothed HiRISE DTM follows that for the HRSC data.

Slopes for the USGS HRSC DTM at Jezero significantly exceed those for HiRISE, indicating that the USGS DTM is noisy at baselines ≤1000 m. Though the HRSC DTM agrees best with HiRISE in elevation when the latter model is smoothed (Table 1), such smoothing will clearly *increase* the discrepancy in slopes. We are thus led to consider smoothing the USGS HRSC DTM, trading reduction of noise for (further) reduced resolution. Using a 9 x 9 lowpass filter yields a DTM very similar in appearance and quality to the HRSC product. It matches smoothed HiRISE data with an optimal filter width of 19.5 pixels (525 m), yielding an RMS error of 9.54 m (corresponding to a matching precision

of 0.243 pixels). The slope spectrum for this smoothed version of the USGS DTM agrees closely with the DLR (and appropriately smoothed HiRISE) data. On the other hand, applying a 5 x 5 filter to the HRSC DTM yields baseline-slope behavior similar to that of the full-resolution HiRISE data. Unfortunately, this agreement is partly fortuitous, since the slopes in the HiRISE model come entirely from real topography, whereas those in the HRSC DTM combine the effects of (smoothed) topography and errors.

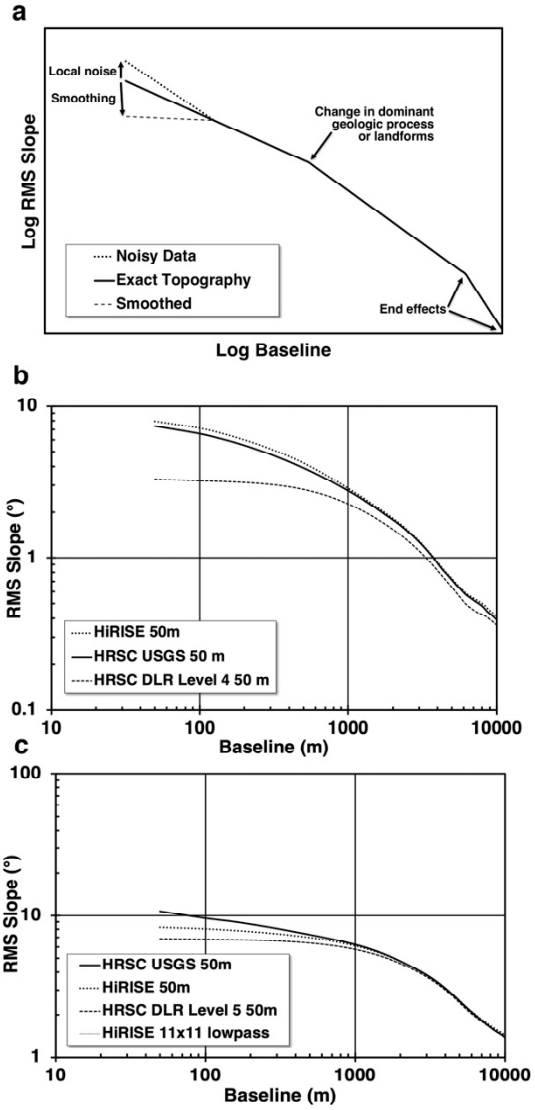

Figure 2. Slope as a function of baseline. (a) Schematic effects. (b) Gale data from Kirk et al. (2018). (c) Jezero data.

4.3 Qualitative and Semiquantitative Assessments

Visual examination of the original and smoothed DTMs sheds additional light on the quantitative comparisons described in the previous section. Presenting the DTMs in the form of synthetic shaded relief images (Figure 3) emphasizes the local texture and details in which we are primarily interested. Examining the DTMs directly (as grayscale or color-coded images) leads to essentially the same conclusions. The inescapable first impression is that the target DTMs are substantially less detailed than the HiRISE data downsampled to the same GSD. The CTX and DLR Level 5 HRSC DTMs each resemble their respective optimally smoothed HiRISE product quite closely. In particular, both the target and smoothed HiRISE DTMs appear homogeneous, with similar scales of detail in both the rugged western and smooth eastern areas. Close comparison shows that surfaces in the target DTMs appear slightly rougher than in the smoothed HiRISE data. This difference is particularly noticeable for CTX.

The appearance of the HRSC USGS DTM is more complex. It contains features of similar size to those in the smoothed HiRISE reference, but also numerous smaller bumps and hollows. Regardless of their dimensions, many of these variations appear to be spurious (topographic "noise") but some correspond to real surface features. These include some small features which are seen in the unsmoothed but not the smoothed HiRISE data. Some larger real features appear "broken up" into clusters of small humps in the HRSC DTM. A few relatively steep slopes (e.g., in the channel walls) appear to be resolved relatively accurately at widths as small as 150 m. On the other hand, the prominent 900-m diameter impact crater in the center of the delta, which is subdued but visible in the DLR DTM, is almost entirely absent. All of these effects were also observed in the Gale crater T1 area (Kirk et al., 2017; 2018). The greater effective smoothness and reduced amplitude of errors in smoother terrain that we inferred from quantitative comparisons is not readily apparent to the eye.

The errors in the target DTMs take the form almost entirely of compact fluctuations (bumps and hollows) either at about the size of the best fit smoothing filter or. for the HRSC USGS DTM, at this size and smaller. Spikes or pits of amplitude exceeding local relief (indicating matching blunders) were not observed. Neither were gaps in the data caused by a lack of matched points, but surprisingly large features (e.g., the 900-m crater) were almost entirely missed. The prominent pattern of correlated errors over small rectangular areas seen with the SOCET SET AATE matcher (Kirk et al., 2003b) was not observed in the DTMs produced with NGATE and smoothed with one AATE pass.

Heipke et al. (2007) measured a quantity related to DTM resolution by counting impact craters visible in the shaded relief and in the orthoimage. The crater diameter at which half the craters identified in the image could be seen in the shaded DTM varied from 2 to 5 km for different processing approaches. The stereo channel GSD was 28 m, similar to our image set. We were unable to make useful crater counts because the small area of the Jezero study area set by the HiRISE coverage contains very few craters visible in the HRSC DTMs. This may be a consequence not only of the horizontal resolution but also of the vertical precision of the DTMs; the images show that the majority of craters present are degraded, with floors filled to nearly the surrounding level. Nevertheless, the few craters visible, along with some small knobs, provide bounds on the equivalent measure of resolution. The CTX DTM contains more craters, but rather than counting them we made a subjective assessment of the smallest craters and knobs that are reliably present.

The HRSC USGS DTM contains only one clear impact crater, with diameter 1800 m (just outside Jezero, hence not seen in Figure 3). Its rim is well resolved. As noted, the 900-m crater is not visible. Knobs 500 m across seen in the orthoimage are generally identifiable in the shaded relief (though many similar appearing bumps do not correspond to real features). Knobs 300 m across are occasionally present, those 200 m and smaller are generally not. In the HRSC DLR DTM, the 1800-m crater is visible, as is a 1600-m crater beyond the edge of the USGS coverage. The 900-m crater is visible but indistinct. Knobs as small as 600 m across are visible; a few real knobs smaller than this can also be identified but appear to be ~600 m wide. In the CTX data, a great many small craters are seen in the shaded relief, but even more in the images. Some craters 150 m in diameter are distinguishable from noise in the elevation data, but many craters

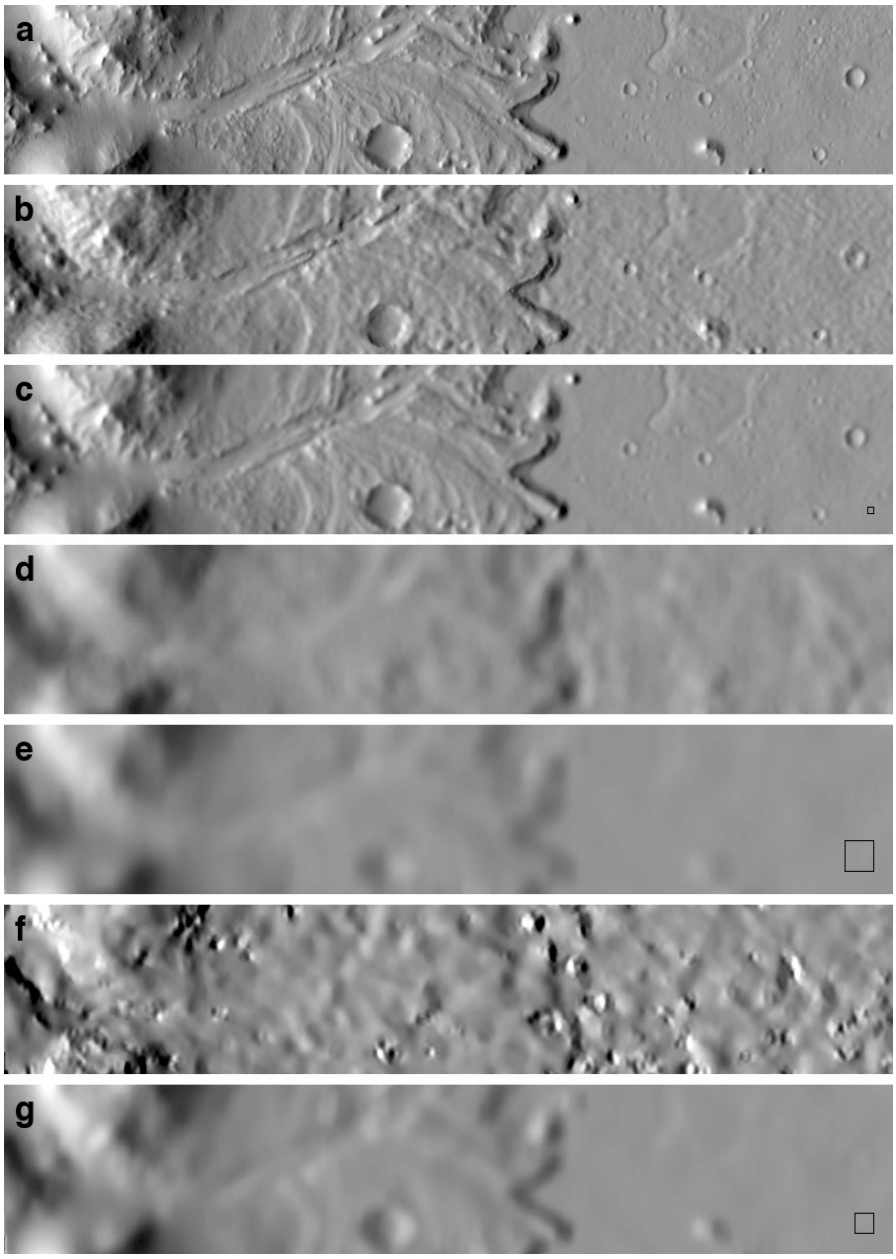

Figure 3. Shaded relief portrayal of Jezero DTMs. (a) HiRISE downsampled to 20 m/post. (b) CTX at 20 m/post. (c) HiRISE at 20 m, smoothed 5 x 5 to match CTX (see Table 1). (d) HRSC DLR Level 5 at 50 m/post. (e) HiRISE at 50 m/post smoothed 11 x 11. (f) HRSC USGS at 50 m/post. (g) HiRISE at 50 m/post, smoothed 7 x 7. All images 9 x 4 km, centered at 77.4°E, 18.5°N, Simple Cylindrical projection, north at top, illuminated from left with identical contrast stretch. HRSC products have been enlarged to aid comparison. Squares in panels (c), (e), (g) indicate the size of smoothing filter applied to HiRISE to match target DTM resolution.

this size are not seen. At 300 m diameter most craters are visible. A few knobs as small as 50-70 m across are visible but appear broadened to ~100 m. Thus, small features are blurred to about the scale of the optimal smoothing filter estimated above, and craters are sometimes visible at diameters 1.5x this smoothing width, but reliably visible at about 3x this size. Resolution as measured by crater visibility thus seems to lie at the low end of the range found by Heipke et al. (2007). This is unsurprising given that significant improvements have been made to both the DLR and USGS processing techniques in the interim (Gwinner et al., 2016; Kirk et al., 2016).

Finally, we examined elevation statistics for flat areas of the Jezero crater floor to make crude estimates of the vertical precision of the DTMs, independent of a reference model. The images and HiRISE DTM show areas that are topographically smooth yet have abundant image texture as a result of albedo variations. The largest such rectangular area in the HRSC DTM was 7.5 x 3 km. The standard deviation of elevations in this box was 11.6 m for the DLR dataset and 12.2 m for USGS. In the CTX DTM, real topographic variations are apparent and the largest featureless box we could identify was 1 x 1 km. The elevation standard deviation was 2.64 m. If these elevation dispersions are attributed to matching error, the corresponding matching precisions are 0.296, 0.309, and 0.199 pixel, in good agreement with the results for the smooth area (Table 2).

5. DISCUSSION

The results presented here for vertical and image-matching precision are not unexpected. Matching precision on the order of

0.2 pixel has long been a rule of thumb for predicting vertical precision (e.g., Cook et al., 1996). We have studied matching precision in SOCET SET with a variety of approaches applied to images from very different cameras (Kirk et al., 1999; 2003b; 2008; 2017) as well as simulated images (Kirk et al., 2016). Finding values in the range 0.2–0.3 pixel (as here) we have recommended this range as a rule of thumb for predicting vertical precision for particular images, and for designing stereo cameras and observations to meet some required precision.

Our results for horizontal resolution are somewhat surprising, though consistent with Kirk et al. (2016; 2017; 2018). Because we have used such disparate approaches to try to quantify the horizontal resolution of the target DTMs, it is appropriate to discuss what is meant by the term (we exclude at the outset the erroneous usage of resolution as a synonym for pixel size or GSD) and how the different measures are expected to compare. In contemporary usage (e.g., Wolf et al., 1980), resolution is usually defined as the size of a minimally discernable gap between features, or a *line*. (Earlier usage referred instead to the separation between centers of features, equal to a line *pair*.)

Smoothing with a lowpass boxcar filter will attenuate signals with a wavelength equal to the filter width while passing longer-wavelength signals. Although the peak-to-peak wavelength is equivalent to the separation rather than the gap between features (i.e., a line pair), it must be bigger than the filter width to be visible. Thus, the width of the smallest resolvable line or other feature is likely to be close to the filter width, and this is what we observe in comparing the broadened appearance of small knobs to the amount of smoothing inferred from comparing target and reference DTMs. Area-based image matching is a far more complex process than filtering, but to the extent that a finite matching patch is used, the patch width places a limit on resolution that can be equated to a resolvable line. Matchers are also known to display more complex algorithm- and scene-dependent behavior that can introduce artefacts at other spatial frequencies, but these effects are beyond the scope of our paper.

Slopes are computed for a discrete baseline, and should be just detectable when that baseline extends from (say) the top of a local high to the floor of a barely resolved adjacent gap; this is equivalent to the resolvable line width. “Resolution” defined by the diameter of an identifiable crater requires more thought. At a minimum, the diameter might be equivalent to the width of a line pair, for a simple craterform defined by a high-low-high pattern of elevations. *Recognition* of a landform such as a crater is usually taken to require at least 3 to 5 resolution elements. Thus, we would expect detected craters to be at least as large as the optimal smoothing, and possibly two or three times larger. Considered in this light, the various resolution estimates for each Gale and Jezero dataset are mutually consistent. Though our work shows the value of a reference dataset for evaluating DTM quality, this consistency indicates that useful quality estimates can, in some cases, be obtained without such a reference.

Many DTMs are generated with a GSD between 3 and 5 image pixels. (For HRSC team products, GSD is set at about twice the mean image GSD which is often, as here, 4x the nadir GSD.) We have offered this as a rule of thumb for DTM design in past work (e.g., Kirk et al., 2003; 2008; 2016) and justified it on the grounds that smaller GSD would not increase resolution. The rationale is that, for some matching tools, three pixels is the smallest (odd) patch size that could be used in area-based matching, so posts spaced more closely than this would not provide independent elevation information. Our results indicate that this lower limit is not violated, but also is not approached closely, at least when the average behavior over large regions is considered. The discrepancy is understandable given that matching software may choose patch sizes larger than the minimum possible. Nevertheless, choosing a GSD of 3–5 pixels is a reasonable rule for DTM design, because it is desirable to oversample resolved features, and because the density of good matches (therefore resolution) can be locally greater in areas of optimal image texture. What we do not recommend is to take either 3–5 pixels or the GSD of a DTM as its resolution; the true resolution is in the range of 10–20 pixels in the examples presented here. A more conservative rule of thumb based on this range is thus more appropriate for designing stereo cameras and observations.

Perhaps our most surprising result is the spatially variable quality of the USGS DTMs. Not only do the precision and best-fit smoothing vary with roughness, individual areas contain a mixture of features (and noise) ranging from the size of the best-fit filter to much smaller. These variations make sense in light of the description (Zhang, 2006; Zhang et al., 2006) that the NGATE matcher produces a dense set of candidate matches and then filters them in a robust (nonlinear) way to populate the output DTM. The details are proprietary, but it is plausible that the algorithm applies more smoothing (yielding smaller precisions) when it detects smoother terrain. It seems also to apply a mix of stronger and weaker smoothing locally, resulting in the mix of relief at multiple scales. Why the precision was terrain sensitive for CTX data but the smoothing was not is unclear. The obvious differences between CTX and HRSC images are higher signal-to-noise ratio and smaller GSD. Perhaps higher resolution allows CTX to detect and match localized, high contrast albedo variations, though this might be expected to yield consistently good rather than coarse resolution. The HRSC DLR DTM has more homogeneous noise and resolution, indicating a different and less terrain-sensitive processing approach. The slightly greater smoothness overall also indicates a different tradeoff between precision and resolution. A very similar trade can be made by filtering the USGS DTM after matching, but this does not entirely hide the roughness-dependent behavior.

We conclude with words of warning to the community of planetary DTM users. The true resolving power of stereo DTMs may not be as great as the GSD would suggest. Furthermore, it can vary within a DTM, and small “features” may represent a mixture of actual surface detail and artefacts, so that they should be interpreted (if at all) with caution and always in light of the images. Therefore, it is not good practice to simply optimize horizontal resolution without consideration of the effect on vertical error. The achievable trades between horizontal resolution and vertical precision is likely to depend somewhat on image and scene details such as SNR, illumination, and terrain roughness. The optimal trade in any case may be different depending on the application, with slope accuracy prized for landing site selection but detection of fine details perhaps more valuable for some geologic studies or minimization of artefacts for others.

6. FUTURE WORK

Several directions for future investigation are evident in light of this work. First, the high-quality reference data provided by HiRISE can be used to test and optimize matcher performance. For example, can parameters of the NGATE matcher such as internal smoothing be selected to yield reasonably reliable short-baseline slope estimates for a variety of terrains and illumination conditions? Second, assessing the properties of DTMs produced with other software, such as the Ames Stereo Pipeline (Moratto et al., 2010) is of interest. Third, the intrinsic precision and density of points obtained by image matching should be studied

separately from the quality of DTMs interoplated from those points. With access to the algorithms and software (necessarily for noncommercial software), both steps might be improved based on such evaluations. Fifth, our quantitative analysis in this paper addresses average properties of a DTM, so derivation of local quality parameters supporting the interpretation of specific features is desirable. Finally, photoclinometry (Kirk et al., 2003a) has been suggested for improving stereo-derived DTMs (e.g., Kirk et al., 2006). Comparison with a reference DTM would quantify its effects on resolution and precision.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support of the National Aeronautics and Space Agency (NASA) Mars Express Project, the Planetary Geology and Geophysics Cartography program (2005-2015) and the NASA-USGS Interagency Agreement for planetary mapping (2016 on) for the work described here.

REFERENCES

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

Cook, A.C., et al., 1996. Clementine imagery: Selenographic coverage for cartographic and scientific use. *Planet. Space Sci.*, 44, 1136–1148.

Craft, K.L., et al., 2017. A Stereophotoclinometry model of a physical wall representing asteroid Bennu. *Lunar Planet. Sci.*, 48, 1964.

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.

Kirk, R.L., et al., 1999. Digital photogrammetric analysis of the IMP camera images: Mapping the Mars Pathfinder landing site in three dimensions. *J. Geophys. Res*., 104 (E4), 8868–8888.

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. Remote 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 Planetary Science Conference*, 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. Remote 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. 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 Remote Sensing and Mapping* (B. Wu, K. Di, J. Oberst, I. Karatschevseva, eds.), Taylor & Francis, 107–124, doi:10.1201/97804295 05997.

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

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.

Moratto, Z.M., et al, 2010. Ames Stereo Pipeline, NASA's open source automated stereogrammetry software. *Lunar Planet. Sci.*, 41, 2364.

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.

Wolf, P.R., Johnson, S.D., Keating, T.J., Kerr, W.E., Mezera, D.F., Pimentel, L.T., and Sieker, F.A., 1980. Definitions of terms and symbols used in photogrammetry. In *Manual of Photogrammetry*, 4th Edition (Slama, C.C., Theurer, C., and Henriksen, S.W., Eds.), American Society of Photogram–metry, Falls Church, Virginia, 1056 pp.

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

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 Annual Technical Exchange Conference*, Nashville, Tenn., 26–27 September 2006.

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 Annual Conference*, Reno, Nevada, 1–5 May 2006.