Analysis of Flash Lidar Field Test Data for Safe Lunar Landing

Andrew E. Johnson, Jason A. Keim and Tonislav Ivanov
aej@jpl.nasa.gov, Jason.A.Keim@jpl.nasa.gov, Tonislav.I.Ivanov@jpl.nasa,gov
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109

Abstract— In May 2008, the Autonomous Landing and Hazard Avoidance Technology (ALHAT) Project conducted a helicopter field test of a commercial flash lidar to assess its applicability to safe lunar landing. The helicopter flew several flights, which covered a variety of slant ranges and viewing angles, over man-made and natural lunar-like terrains. The collected data were analyzed to assess the performance of the sensor and the performance of two algorithms: Hazard Detection (HD) and Hazard Relative Navigation (HRN). The collected flash lidar data were also used to validate a high fidelity flash lidar software model used in ALHAT Monte Carlo simulations. The field test results, combined with prior simulation results, advanced the technology readiness level of the HD algorithm to TRL 5 and the HRN algorithm to TRL 4.[1][2]

TABLE OF CONTENTS

1. INTRODUCTION

The Autonomous Landing and Hazard Avoidance Technology (ALHAT) Project is developing sensors and algorithms to increase safety during lunar landing [3]. Two new algorithms, executing during the final approach phase, are being developed: on-board hazard detection (HD) to determine a safe landing site [4], and hazard relative navigation (HRN) to enable accurate landing at the selected safe site. Lidar was selected as the primary sensor for these new algorithms to accomplish ALHAT's task of providing technology that works autonomously under any lighting conditions. In particular, flash lidar [8] is a promising new technology – providing high data rate and instantaneous capture of an entire range image. Currently, ALHAT is developing a flash lidar sensor with a 5cm range precision that will be able to detect 30cm-high hazards when viewed from a 1000m slant range at a 15-degree path angle. Slant range is the distance to the target along the line of sight, and path angle is the angle between sensor boresight and the horizontal [2].

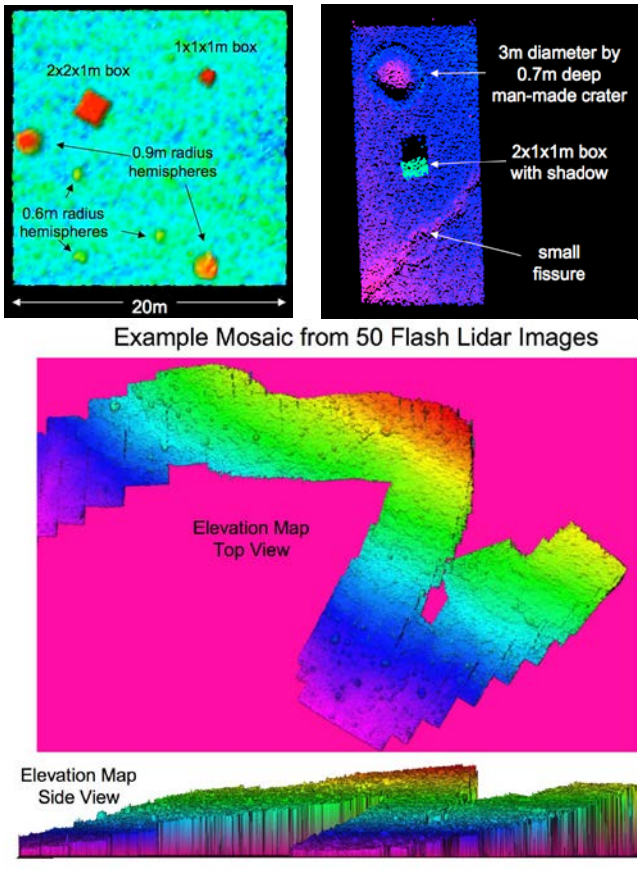


Figure 1: Examples of flash lidar data collected.

The goal of the first Field Test (FT1) was to assess the commercial state of the art by testing a custom-built commercial flash lidar. The expected performance for this field test sensor was the detection of 30cm high hazards from a 500m slant range at a nadir viewing direction. As described below, the sensor was able to generate range

[1] 978-1-4244-3888-4/10/$25.00 ©2010 IEEE.
[2] IEEEAC paper #1659, Version 1.2, Uploaded October 30th, 2009

images at 500m, but did not have the range precision required to detect 30cm hazards. As a result of this finding, numerous design improvements are being investigated to obtain the 5cm range precision required of the ALHAT flash lidar. Such improvements include accurate detector calibration, uniform laser pulse shaping, and implementation of a faster time-of-flight clock. Single and multiple image-processing techniques are also being investigated to decrease noise and increase resolution. The ALHAT sensor will also obtain a greater maximum slant range by using a more powerful laser and a more sensitive detector.

This paper briefly describes the sensors, the data storage system, test terrain and the trajectories flown. The bulk of the paper describes the performance of the sensor as a function of viewing angle and slant range, the HD performance relative to hazard size, and the HRN accuracy and error rejection performance. A comparison of real to simulated HD results is made and used to extrapolate the HD performance of the ALHAT sensor being developed, from the performance of the field test sensor. The paper concludes with a discussion of TRL advancement.

2. TEST OVERVIEW

The FT1 objectives were to:

- Test a commercial flash lidar in a relevant environment and use this information to guide the development of the ALHAT flash lidar sensor
- Test HD and HRN algorithms using data collected with a real sensor in a relevant environment and use this information to improve algorithms
- Collect data for validation of the flash lidar sensor model used in the POST II Monte Carlo simulation
- Identify areas to increase the fidelity of the sensor model
- Advance sensor and algorithm TRL

Sandbox analysis prior to FT1 showed that the slant range and path angle had a strong influence on the flash lidar and the HD algorithm performance. Studies also showed that the size of hazards influenced detection performance, while the distribution of hazards (number of hazards of each size per unit area) influenced the probability of detecting a safe landing site [2]. Consequently, FT1 was designed so that hazards of a variety of sizes, placed in fields with different hazard distribution, could be imaged from different angles and distances.

To obtain a variety of slant ranges and path angles, as well as descents toward the targets, a helicopter was used as the test platform. An inertially stabilized gimbal was mounted to the front of the helicopter. The gimbal contained the flash lidar, two Inertial Measurement Units (IMU), an orientation sensor, two digital cameras and an analog camera. A GPS attached on the fixed structure above the gimbal provided position estimates. One IMU was used for stabilization of the gimbal while the other was used for continuous attitude propagation. The digital cameras were used for occasional absolute attitude determination. The orientation sensor was not actually used while the analog camera was for real-time feedback on gimbal pointing. A comprehensive description of the FT1 hardware, ground infrastructure and trajectory reconstruction is given in [7].

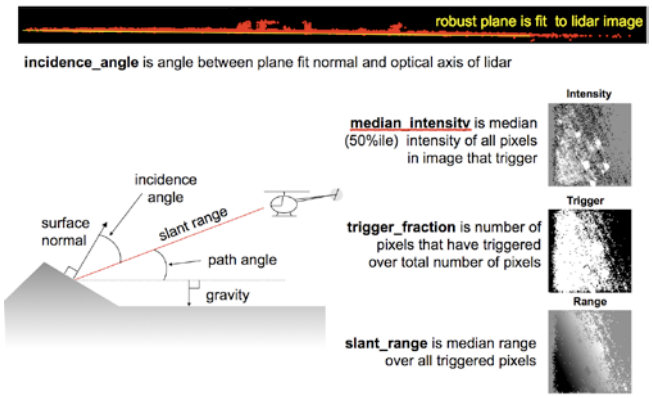


Figure 2: Metrics to assess sensor performance.

The testing was conducted at NASA Dryden Flight Research Center (DFRC) and Death Valley National Park. At DFRC, there were two different target sites: one on the Lakebed and one in the Borrow Pit. The Lakebed site was very flat and consisted of 11 hazards in close proximity: 9 hemispheres of various sizes and reflectivity, a large box and a small box. The Lakebed site was designed for lidar characterization. The Borrow Pit site had numerous hazards made out of 1x1x1m boxes, fields of hemispheres with 5% and 10% distribution, and two 3m-wide craters. The Borrow Pit site was designed for assessing hazard detection and safe landing probability. The final site was at Mars Hill in Death Valley National Park. Mars Hill has numerous rock fields of varying distribution, as well as steep and shallow slopes. The purpose of the Mars Hill site was to obtain lidar data from natural, as opposed to man-made, hazards and to test slope hazard detection.

Figure 1 (top right) shows a single flash lidar image taken of the Lakebed site from a 430m slant range at a nearly nadir view. Due to the nadir view, the projection of the field of view is nearly square. The hazards near 1m in height are clearly visible; smaller hazards show up but are close to the noise floor of the sensor. Figure 1 (top left) shows 3D points from a single flash lidar image of the Borrow Pit target taken from a 184m slant range at an oblique view angle, 27˚ from horizontal. The oblique view stretches the 3D points along the viewing direction. The image clearly shows the detection of a 3m-wide crater and a 2x1x1m box. Due to the oblique view, there are 3D points on the side and top of the box, and there is a shadow behind the box and behind the front edge of the crater. Figure 1 (bottom) shows a mosaic of 50 flash lidar images taken at the Mars Hill site. The 3D points from all images are projected into a single elevation map to construct the mosaic. The numerous rocks and slopes of the hill are clearly seen in the map.

To assess the performance of the sensor and the hazard detection algorithm, reference digital elevation maps (DEM) of the sites were built. For the Lakebed and Borrow Pit sites, these "truth DEMs" were constructed by generating hemisphere, box, and crater objects at the surveyed positions and orientations of the site hazards . The terrain between the hazards was filled in by a surface fit through all the survey points. A plane was fit for the Lakebed site and a quadric surface was fit for the Borrow Pit site. For the Mars Hill site, a previously constructed DEM was used. In the analysis presented in this paper, we only used the Lakebed DEM shown in Figure 3.

To compare the lidar data to the truth DEM, the position and attitude of the lidar in the truth DEM coordinate frame is required. Because the flash lidar pixel size is small (0.20m at 500m), this position and attitude knowledge needed to be very accurate (0.4 mrad for attitude and 0.2m for position). To achieve this accuracy, the following two-step procedure was used. First, the entire lidar trajectory for the flight was reconstructed using IMU, GPS and camera measurements. This process [7] placed the lidar hazards within 1m horizontally of the truth hazards. Furthermore, the lidar had a time-varying range bias which, combined with the above mentioned horizontal trajectory errors, caused the lidar data to be misaligned horizontally and vertically relative to the truth DEM. To clean up this misalignment, the flash lidar data and truth DEM were correlated using a procedure based on the HRN algorithm [5]. The end result was precise alignment to 0.1m or less.

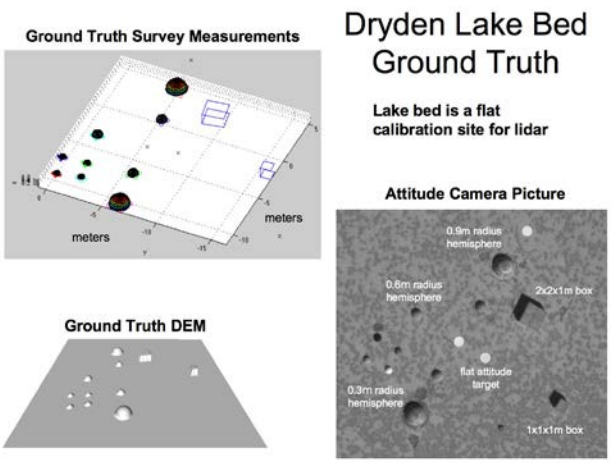


Figure 3: Ground truth DEM for the Lakebed test site at DFRC.

3. SENSOR ANALYSIS

Lidar hardware

The flash lidar used in FT1 was provided by NASA Langley Research Center. This lidar was procured from Advanced Scientific Concepts and then retrofitted with a real-time data collection and timing system. The lidar is comprised of a 1.57 micron, diode-pumped Nd:YAG laser source and receiver optics. The laser source is diffused to 3^o to actively illuminate the target site observed by the receiver optics. This receiver has a field of view slightly less than 3^o and is composed of an InGaAs array of 128×128 pixels and optics with a 250mm focal length and a 120mm diameter aperture. After calibrating the sensor projection function, the receiver provides range and bearing information for each pixel. For FT1, the sensor operated at a frame rate of 8 Hz.

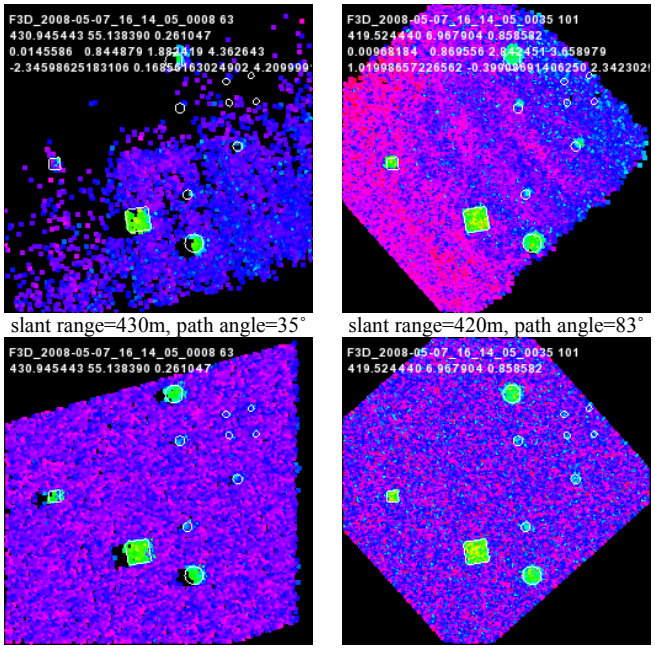
In the flash lidar hardware, range is measured for each pixel. The range is determined by finding the peak time-of-flight of the returned laser pulse for each pixel [1]. The digitization of the light energy hitting a pixel happens continuously but only the most recent 20 digitization results are stored in a circular buffer. When the light energy passes a threshold, the pixel "triggers" and the 20 digitization results are written to memory. If the return laser power does not exceed the threshold, the pixel does not trigger. The top left image in Figure 4 shows a flash lidar image that has a low trigger fraction (0.26). As can be seen, many pixels do not have range measurements. This image is taken at the operational limits of the sensor's hardware.

The flash lidar range measurements are assembled in a flash image and, when interpreted, form a cloud of 3D points. Using standard robust estimation techniques [9], a plane can be fit to the 3D points. For example, a plane fit algorithm that ignores outliers was used at the Lakebed site, where many hazards stick up from a relatively flat ground. From the plane parameters, the slant range and the incidence angle relative to the target plane can be computed (see Figure 2). In terrain with no slope, such as the Lakebed site, the trajectory path angle is the complement of the incidence angle.

Lidar Trigger Fraction and Maximum Slant Range

Based on the trajectory reconstruction, lidar images from one of the flights over the Lakebed site that had more than 10 points falling within the horizontal bounds of the truth DEM were tabulated. For each of these 2423 images, the trigger fraction, slant range and path angle were computed. The trigger fractions were binned by path angle in 15˚ increments and slant range in 50 m increments. The median trigger fraction for each bin was computed and the results are plotted in Figure 6. The inset table shows the number of images in each bin.

Based on Figure 6, we observed the following trends. For all path angles, the trigger fraction decreases as the slant range increases. This expected trend is due to the decrease in laser energy reaching the detector from greater slant ranges. We also observed that as the path angles decreases the trigger fraction also decreases. This trend is expected because the diffuse reflectance from the surface decreases as the path angle decreases.



slant range=430m, path angle=35˚ slant range=420m, path angle=83˚

Figure 4: Real (top) and simulated (bottom) elevation maps for a pair of slant ranges and path angles. Color scale is purple (-0.5m) to red (1.5m).

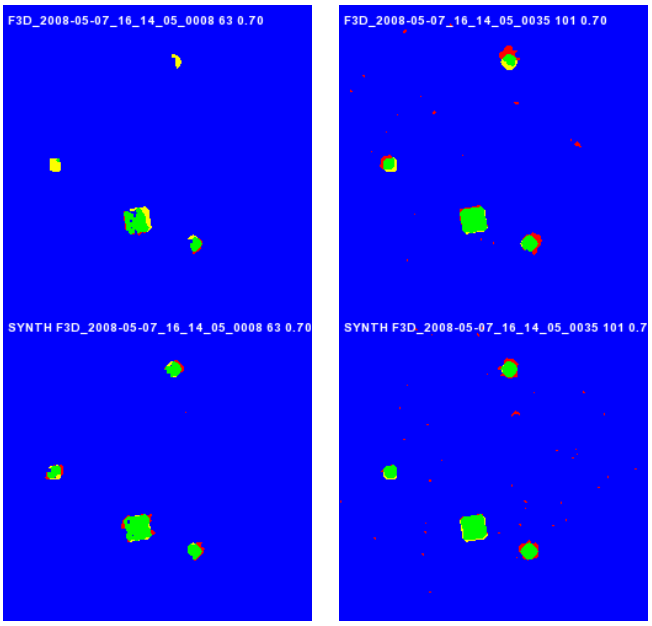


Figure 5: Real (top) and simulated (bottom) hazard detection results with a 70cm roughness detection threshold for the four elevation maps shown in Figure 4.
GREEN = TRUE POSITIVE **RED = FALSE POSITIVE**
YELLOW = FALSE NEGATIVE **BLUE = TRUE NEGATIVE**

To use the sensor for HD or HRN, high trigger fractions are required. Trigger fractions as small as 90% can be tolerated since the DEM generation software is capable of filling in small holes due to the missing pixels. Trigger fractions below 90% result in holes that are too large to fill in and become treated as hazards, which is undesirable. For near nadir viewing (82.5˚ path angle), the lidar produced trigger fractions above 90% out to 400m slant range and then the trigger fraction quickly rolled off. For the opposite extreme of a 0˚ to 15˚ path angle[3], the trigger fraction stayed above 90% only until around 250m. Thus, given the 90% threshold, the maximum operational range of the sensor varied from 400m for nadir viewing to 250m for the extreme oblique viewing angles.

The FT1 flash lidar was designed to have a maximum range of 500m at nadir viewing, but did not meet this expectation. There are several causes that explain this shortfall. During the test, a window was placed in front of the lidar to reduce drag and minimize contamination. It is predicted that the window and its coating reduced the amount of light received by 28%. Also, the data used in this analysis are from a dry lakebed surface, which has a different reflectivity from the lunar surface. These all have an effect on the absolute maximum range values, but do not change the sensitivity of the sensor to range and path angle.

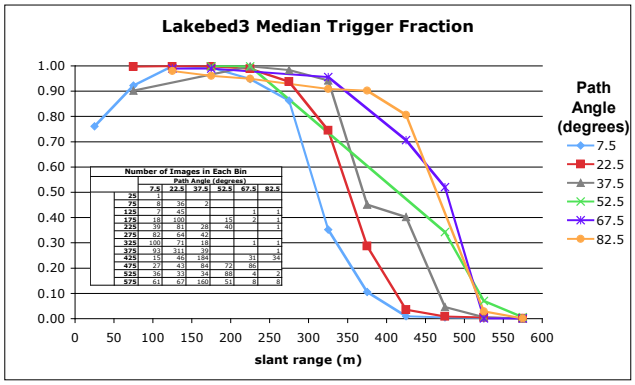


Figure 6: Median trigger fraction for 2423 images taken of the lakebed binned by slant range and path angle

DEM Elevation Error and Lidar Range Error

The 3D point sets, derived from the flash lidar images, were converted into DEMs after the data-driven alignment procedure was applied. With this procedure, 451 3D point sets from the image data over the Lakebed site were successfully aligned to the truth DEM. The elevation difference between the truth DEM and each lidar DEM was computed at each pixel, and the mean and the standard deviation of the elevation differences across the image were computed. Finally, as was done for the trigger fraction, the standard deviations of the elevation errors were binned by slant range and path angle, and the median standard deviation was computed for each bin. Figure 7 shows the results from these steps.

For viewing close to nadir (82.5˚ path angle), the median standard deviation in elevation error was about 0.20m. Since the lidar is pointed straight down at nadir viewing, 0.20m is taken as an estimate of the FT1 sensor's random range error. However, this is an overestimate of the true capability of the sensor because of several field test limitations. The data

[3]For Apollo 11,12 and 14 path angle was ~16˚; for Apollo 15,16, and 17 path angle was ~25˚

used to compute the estimate were from a slant range around 400m. At this distance, the return laser energy is low, so the signal-to-noise ratio (SNR) is small, meaning more errors. Imaging from a closer slant range should improve the range error. Also, the pixels in the array have intensity-dependent range biases, which had not been corrected prior to FT1. Performing calibration for each pixel will decrease the range error. Finally, the alignment between the lidar and truth DEMs was not perfect, so this could introduce elevation errors near hazards.

As the path angle decreases, the errors in the range measurements decrease in the vertical direction and increase in the horizontal direction. This trend of decrease in elevation error with decreasing path angle was also observed in ALHAT simulations.

Elevation errors also influence hazard detection performance: as the random errors increase, the detection rate decreases while the false alarm rate increases. This trend is discussed in the next section.

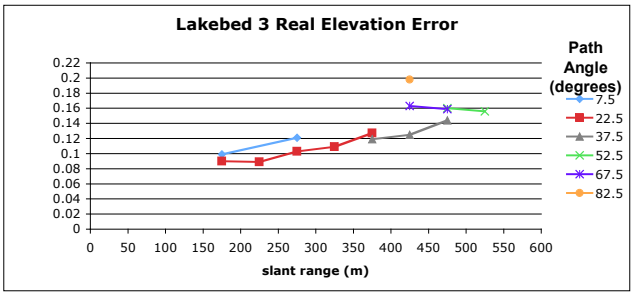


Figure 7: Median standard deviation of elevation error for different path angles and slant ranges.

4. HAZARD DETECTION ANALYSIS

Hazard detection was assessed using the 451 aligned 3D lidar point sets. Performance was analyzed in two ways: comparative hazard analysis and individual hazard analysis. Comparative hazard analysis utilizes the same analysis tools used in the ALDAC1 POST II simulations [6]. Individual hazard analysis is a new way of analyzing HD performance. This analysis aims to assess the impact of the false alarms. The goal is to eliminate enough small false alarms so that a safe site can be found for landing. However, this constraint comes at the expense of not detecting small hazards.

To validate the high fidelity sensor model used in ALHAT, simulated flash lidar images of the truth DEM were generated using the position and attitude of each real flash lidar image (Figure 4, bottom). The range precision parameter in the model was set to 20cm to match the experimental value discussed in the previous section. The HD analysis tools were also run on the simulated imagery to generate statistics that could be compared to the ones from the real data in order to validate the sensor model.

Comparative Hazard Analysis

Comparative hazard analysis compared hazard detections from the real lidar data to those from the reference simulated data. The comparative analysis tool counts correctly detected hazards (true positives = TP), missed hazards (false negatives = FN), incorrectly detected hazards (false positives = FP) and correctly detected no-hazard areas (true negatives = TN). This counting is done at a pixel and a component level. Pixel-level analysis compares the real and simulated detections on a pixel-by-pixel basis. Component-level analysis first groups the adjacent pixel detections into connected components, and then compares the components in both real and simulated data. Components do not have to overlap to be matched; they just need to be within a small number of pixels from each other. This gives component-level analysis some robustness to misalignment.

Figure 5 shows real and simulated detection results with a 70cm detection threshold for the elevation maps shown in Figure 4. At this threshold, both the detection and the false positive rate are comparable between the real and the simulated data. However, in the top left image, one hazard is completely missed and one has very few detected pixels on it.

A quantitative assessment at a 70cm detection threshold is shown in Figure 8. This figure shows detection metrics for a sequence of 51 flash lidar images with very good alignment to the truth DEM and with high trigger fractions. The metrics for the real data are compared to the same metrics for the simulated data. The horizontal axis of the plots is the image number in the sequence.

The bottom right plot compares DEM coverage percentage, which describes the fraction of the truth DEM area that has valid lidar elevation data. The high agreement (correlation coefficient r = 1.00) confirms that the real data have high trigger fractions. The top right plot compares the elevation error metric, and the good correlation (r=0.76) indicates that the trends follow each other in both types of data.

The top left plot compares the true positive detections on a per-pixel basis and the agreement is very good (r=0.95). The simulated data has slightly more correctly detected pixels because there is no misalignment error. The true positive component detections (top middle) agree (r= 0.74), but not as well as the pixel level detections. This may be due to discretization during the component grouping, which will cause a small difference in the number of detections to have a greater impact on the correlation coefficient.

The bottom left plot in Figure 8 shows that the number of false-positive pixels in the real data is significantly higher than the number in the simulated data. The component-level false positives, shown in the bottom middle plot, are in much better agreement, and there are very few false-positive components. This indicates that the false-positive pixels are grouped around detected hazards (caused by a small

misalignment between the truth and lidar DEMs), so when the components are generated, these pixels get grouped with a detected hazard component and are not identified as a separate false positive component.

Based on this comparative hazard analysis, the simulated and real data agree well under 3 circumstances: the simulation sensor model range precision is adjusted to 20cm, the real data is accurately aligned to the truth DEM, and the trigger percentage of the real data is high. The simulation sensor model has been validated for these conditions.

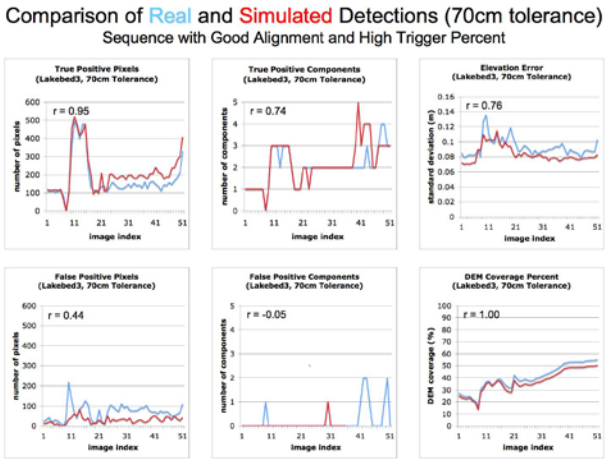


Figure 8: Comparison of real and simulated detections for an image sequence with good alignment and high trigger percent.

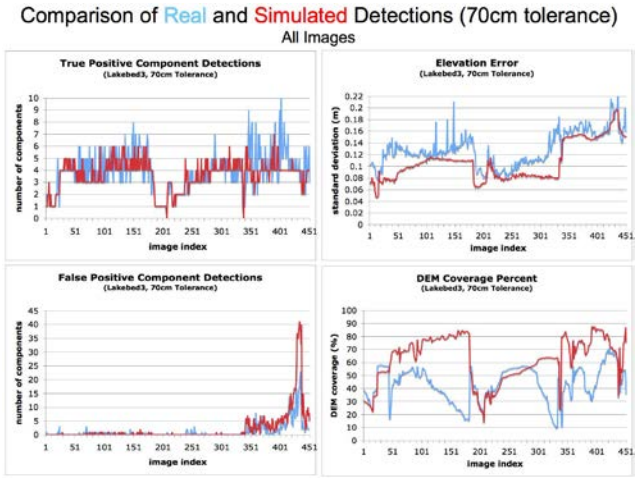


Figure 9: Comparison of real and simulated detections for all images.

As it is expected, the agreement is not quite as good when the data are misaligned or include a significant percentage of pixels that have not triggered. Figure 9 shows metrics for all 451 images aligned to the truth DEM. The segment shown in Figure 8 is between image indices 200 and 250 in this figure. The real and simulated component detections are correlated (true positive component correlation = 0.58, false positive component correlation = 0.88) but there are some noticeable disagreements. One cause of these disagreements is the lower trigger fraction in the real data compared to the simulated data, as indicated by the lower DEM coverage percentage (bottom right plot). Slight misalignments may also come into play as indicated by the higher elevation error in the real data relative to the simulated data (top right plot).

Individual Hazard Analysis

In the comparative hazard analysis, there were a significant number of false positives when the roughness detection threshold was set to a low value. These false positives are treated as real hazards in the safe landing site selection software. Thus, the presence of too many false positives may leave no hazard-free location to land. Nevertheless, despite the false positives, it is desirable to set the detection threshold as low as possible in order to detect the small hazards. Usually, the size of the smallest hazard desired for detection is set in advance. Thus, the HD analysis done up to date has focused on maximizing the detection probability of hazards greater than a predetermined height. A study was conducted to determine the sensor's range precision and field of view needed to detect such hazards [2]. Attempts were made to minimize the false-positive rates by appropriately choosing the detection threshold for such hazards, but there was no clear understanding of the tolerable amount of false positives introduced in the process. Such understanding is needed to make comprehensive statements about the hazard detection ability of the FT1 flash lidar. This ability is the smallest height at which as many as possible real hazards are detected, while as few as possible false positives are introduced.

To quantify these false positives, previous analysis used the false alarm rate (FAR = # FP / (# FP + # TP)). This metric depends on the number of actual hazards (TP) in the area being imaged. As this number decreases, the FAR increases, which can be misleading since the number of FPs has not changed. Instead, a more intuitive metric for quantifying false positives is the number of false positives per area. In ALHAT, the relevant area is the Vehicle Footprint Dispersion Ellipse (VFDE), which is defined as the area under the lander at touchdown grown by the GN&C position error change since the hazards were detected. Any safe landing site must be larger than the VFDE. The new technique for FP quantification is illustrated in Figure 10. Each square is the size of a VFDE and represents one of five disjoint landing areas. At least one of these areas needs to be free of false positives so that there is a place to land. For the FT1 analysis, we used the criterion that there be no more than 1 false positive component in an area the size of 5 VFDEs. Notice that this constraint is independent of the number of actual hazards in the landing area.

FT1 comparative hazard analysis, as well as ALDAC1 analysis, also showed that the numerous false positives were usually small. In the ALDAC1 analysis, these small false

positives contributed to the failure of a few trials, where the HD software picked a landing site that contained a hazard for landing. In these trials, the configuration of the false positives was such that the landing location, which is chosen to be farthest from all detected hazards (true and false positives), was actually near a small undetected hazard. The remedy is to eliminate the small hazard detections (both true and false) from consideration in the safe site selection process. This is accomplished by placing a component detection threshold on the hazard size in pixels; hazards below the threshold are eliminated, while hazards on or above the threshold are kept. The consequence of this procedure is that the vehicle must be tolerant to all hazards below the threshold. Figure 10 illustrates how these two constraints are applied. An elevation map (top) is generated and from that a roughness map is computed. The roughness detection threshold is applied and detection components are identified (middle). Next, the component detections with number of pixels smaller than the component detection threshold are eliminated (bottom). The large detections remain but there are more places to land than if the small detections were kept.

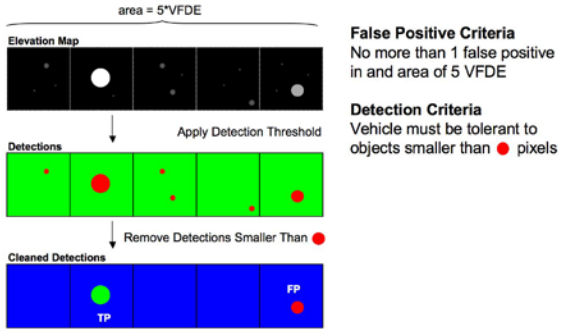


Figure 10: Individual hazard discrimination criteria.

Individual hazard detection analysis was applied on the data from the Lakebed site. The first step was to determine the number of pixels detected on each hazard in the site as a function of the roughness detection threshold. To do this, a roughness map was computed from each lidar DEM. The roughness threshold was then varied from 20cm to 150cm, in 10cm intervals, and applied to the roughness map. A binary detection image was generated for each threshold. The detection pixels in this image were grouped into connected components. These components were compared to the ones generated from the truth DEM. If any two corresponding components intersected, a TP hazard occurred and the number of pixels inside the intersection was considered as the size of that hazard. If the component did not intersect, a FP hazard occurred and the number of pixels inside the component from the lidar DEM was considered as the size of that hazard. This process was done for all 451 lidar DEMs aligned to the truth DEM.

The results from this step are shown in Figure 11. The solid curves represent the median number of TP pixels on each hazard object (over all 451 images) as a function of the detection threshold. Rightmost is the curve for the 2x2x1m box, which is the largest hazard and thus has the most pixels on it. Next is the 1x1x1m box, which has more pixels on it than the 0.9m radius hemispheres at the higher detection thresholds. However, for lower detection thresholds, because hemispheres have a larger area closer to the ground, the 0.9m hemispheres had greater number of pixels than the 1x1x1m box. After that are the curves for the three 0.6m radius hemispheres. These curves exhibit some differences amongst themselves. All hemispheres were made from the same plastic material and were covered with a thin layer of dirt from the lakebed; thus, reflectivity differences are not the source of the variability. Instead, the variability is likely due to the fact that some hemispheres only appear in certain images, and thus the statistics for the three hemispheres were based on a collection of images with different path angles and slant ranges. The final set of curves is for the four 0.3m radius hemispheres. Each of these hemispheres has a different reflectivity, which causes the number of pixels on each hazard to vary accordingly. The white target is the most reflective and thus has the most pixels.

The second step was to assess the number of false positives as a function of detection threshold. Since any false positive greater in size than the component detection threshold is considered a hazard, this analysis looks at the cumulative number of false positives greater than a certain size.

For each image at each detection threshold, the number of false positives in the image was tabulated as a function of false positive size in pixels (e.g., 4 of 1 pixel size, 2 of 2 pixels, … 0 of 50 pixels). These counts were then converted into the cumulative number of FP of size *n* pixels or greater. Each image had a different area, so, to normalize out this effect, the cumulative counts of FP per image area were converted into cumulative counts of FP per VFDE area by multiplying the counts by the image area divided by the VFDE area (380m^2 for a 22m diameter VFDE).

This data was then used to find the component size threshold for each detection threshold where 80% of the images had no false positives. The result is plotted by the dashed curve in Figure 11. This curve divides the plane into an upper region that has less than a 20% chance of a false positive per VFDE and a lower region that has more than a 20% chance. The rightmost point on the curve is at (110cm,1pixel). This point indicates that, with a 110cm detection threshold, 80% of the images do not have a false positive of size 1 pixel or greater within an area the size of 1 VFDE.

Assuming that false positives in each image are independent (which is true if the FP are generated solely based on random uncorrelated range measurements), then each image is a random sample from the distribution of false positives per VFDE. If 80% of the images do not contain a false positive, then the 1 false positive in 5 VFDE criterion is satisfied.

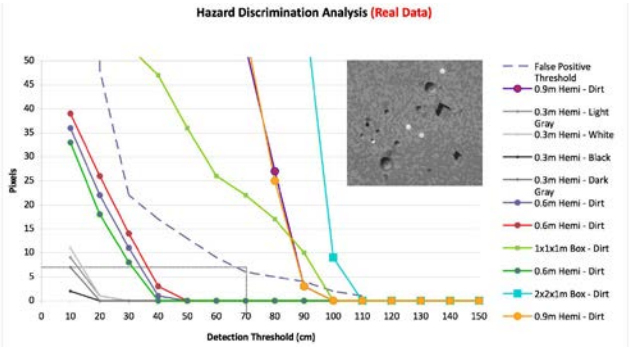


Figure 11: Individual hazard analysis for real data. A 70cm detection threshold and a 10 pixel component size threshold allows for detection of the larger hazards with less than 20% chance of a false positive per VFDE.

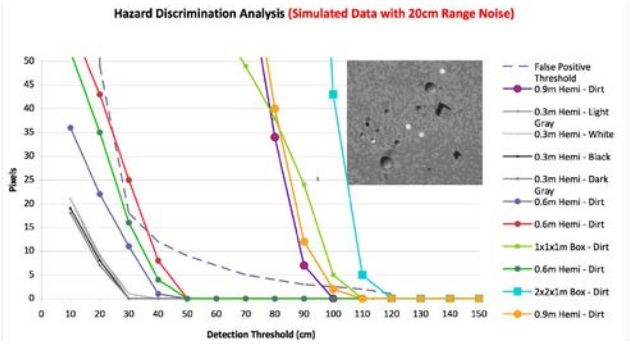


Figure 12: Individual hazard analysis for a simulated flash lidar with 20cm range precision. The similarity between this plot and the same plot for the real flash lidar data demonstrates that the sensor model used in ALHAT simulations is a reasonable representation of the truth.

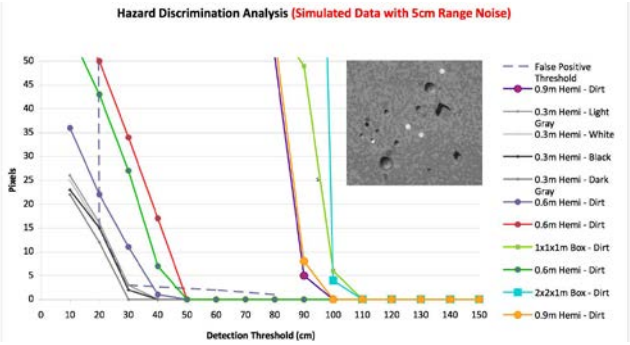


Figure 13: Individual hazard analysis for a simulated flash lidar with 5cm range precision. This represents the expect performance of the flash lidar under development by ALHAT.

To meet the one-false-positive-per-5-VFDE criterion, the detection threshold and component-size threshold must be selected above the dashed false-positive curve. Because all of the curves for the 0.60m and 0.30m diameter hemispheres are below this dashed curve, the hemispheres cannot be detected with any threshold setting.[4] The larger objects can be detected with a range of thresholds. One possible setting is a detection threshold of 70cm and a component size threshold of 10 pixels. This setting allows the detection of the 2x2x1m box, the 1x1x1m box and the 0.9m radius hemispheres, while it limits the number of false positives to less than 1 per 5 VFDE areas.

The same individual hazard detection analysis was applied to the simulated lidar images generated for Section 3 and the results are shown in Figure 12. In general, these curves are similar to the ones form the real data. One discrepancy however, is that the false positive curve in the simulated data actually goes below one of the 0.6m diameter hemisphere curves. This indicates that for some viewing directions and path angles the 0.60m hemispheres may be detectable while also assuring sufficiently low false positives. Another discrepancy is that the four 0.3m diameter hemispheres all have very similar curves in the simulated data. This is because the different reflectivity of the spheres is not modeled in the truth DEM from which the simulated data is rendered. The final discrepancy is that the 1x1x1m box has significantly more pixels for any given detection threshold in the simulated data. This is possibly due to pixels not triggering on the edges of the box in the real data and such edge effect is not modeled in the sensor model. In order to predict the performance of the flash lidar under development by ALHAT, a set of simulated flash lidar images was generated with a 5cm range precision and the same analysis was applied. The resulting plot is shown in Figure 13. In this plot, the false positive line intersects the hemispheres that are 30cm high. This indicates that the sensor under development could detect 30cm high hazards while still limiting false positives to one per 5 VFDE (just barely). The detection rate could be increased for these hazards as the expense of slightly more false positives.

These results should be interpreted as the nominal performance over the part of the test space (slant range, path angle, hazard size) covered by the 451 images from the Lakebed3 flight. Given a different set of images, the performance may improve or degrade. At the very least, this analysis is an example of how HD analysis will be conducted in the future.

5. HAZARD RELATIVE NAVIGATION ANALYSIS

The assessment of hazard relative navigation was straightforward relative to the assessment of HD. The HRN algorithm [5] was applied to consecutive pairs of lidar images that viewed the lakebed target and the alignment result was compared to the reconstructed trajectory.

[4] It should be reiterated that these results are for the commercial flash lidar tested in FT1. Later it will be shown that the flash lidar being developed by ALHAT will be able to detect these hazards

HRN Analysis Approach

First, pairs of consecutive images from all 2423 images that viewed the lakebed targets were determined. Due to data gaps this resulted in 2322 image pairs for HRN assessment. To test the robustness of the HRN algorithm, no culling of the pairs due to low trigger fraction, path angle or slant range was performed. The 3D points from each pair were transformed to the local level frame using the reconstructed trajectory. The misalignment corrections from data-driven alignment were not used for two reasons: first, the number of image pairs would have been significantly reduced and second, the misalignments were fairly static so they had little effect on consecutive image correlations.

Next, a random 3D position shift was applied to the second image in the pair and elevation maps were generated for the first and second (shifted) image[5]. No attitude shift was applied; the attitude for each image from the trajectory reconstruction was used. The 3D shift was uniformly distributed in each axis between -1 m and 1 m (±1m cube). A track point was selected in the first image and a patch of elevation around this point was extracted for correlation with the second image. The correlation search window was set to ±3m in both horizontal directions around the selected track point position. The elevation patch in the first image was correlated with the search window in the second image. The 3D shift computed from the correlation peak was then compared to the true 3D shift to compute an error residual.

Along with this residual, four correlation metrics (peak height, peak width, peak ratio and track point contrast) were recorded. Peak height is the height of the highest peak in the correlation search window and it ranges between -1 and +1, with +1 indicating excellent correlation. The peak width is a measure of how wide the peak is in pixels. The peak ratio is the ratio of the highest peak to the second highest correlation peak. Track point contrast is the variation of the local terrain relief over the track point patch measured in meters per pixel. The HRN algorithm used these four correlation metrics internally to determine if a correlation was valid. To do that, thresholds are placed on the correlation metrics and if any are violated, the HRN estimate is considered invalid. Through the use of these checks, the HRN algorithm is able to throw out most incorrect estimates. Another important metric for assessing performance is the Circular Error Probability (CEP). The 95th percentile CEP is the radius of a circle that contains 95% of the HRN residuals within it.

Some examples of HRN tracking are shown in Figure 14. The middle and right column show valid correlations; the left column shows an invalid correlation. The top row of each column shows the first image and the selected track point. The second row shows the second image after it has been shifted synthetically to simulated motion errors. The third row shows the patch from the first image overlaid on its best match location in the second image. Notice that the large box, which is the feature with the sharpest corners and greatest elevation change, was selected for tracking by HRN in the two valid examples. Although the box was on the edge of the image in the middle example, the correlation was still able to find the correct matching position. In the left example, neither the box nor the hemisphere was selected for tracking because the track point selector had been constrained to the center of the elevation map. This was done to prevent selection of points that might not be in the overlap between images. As a result, a low contrast patch was selected and the correlation metrics mark this pair as invalid.

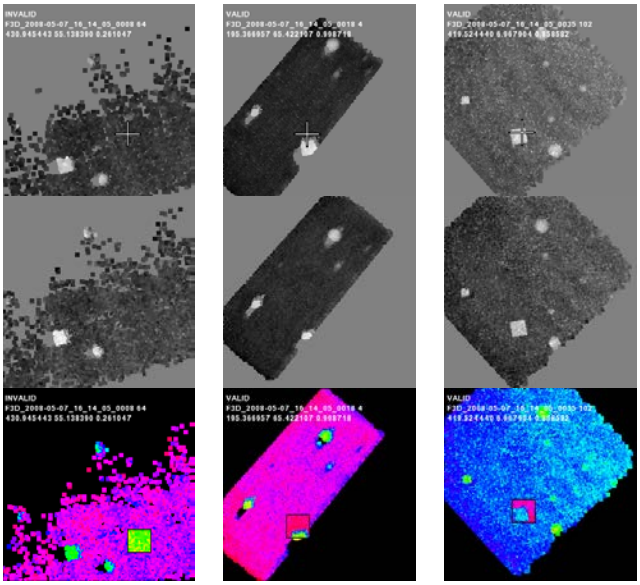


Figure 14: HRN examples.

HRN Performance

The HRN algorithm was applied to all image pairs. Correlation thresholds were set to peak height >0.5 (unit less), peak width < 15 pixels, peak ratio > 1.1, and contrast > 0 (no threshold). With these thresholds, 1086 out of 2322 pairs were marked as valid. For all valid pairs, the HRN error residuals on the x, y and z coordinates are shown in Figure 15. The elevation error (z) is typically less than the 0.20m range precision of the sensor due to averaging over the correlation patch.

A scatter plot of the horizontal errors is shown in Figure 16. The ±1m random shift bounds are shown with the dashed green box and the ±3m search window is shown with the red box. The errors are clustered around the origin with only a few gross outliers. The Gaussian nature of the errors near the origin is apparent in the zoomed view. A case-by-case investigation revealed that the 6 pairs with errors outside the ±1m box are due to a track feature being selected right on the edge of the flash lidar image. This is a challenging correlation situation and it is avoided by forcing HRN to pick track points away of the DEM boundaries. For all valid results, the 95th percentile CEP was 0.32m. Although this

[5] To facilitate automatic processing a fixed resolution (0.1m) was used.

error is greater than one DEM pixel, HRN tracking is pretty good given the large noise in the sensor. Also, this statistic is for all valid path angles and slant ranges in the data set, which includes many images with low trigger fractions where correlation could be challenging.

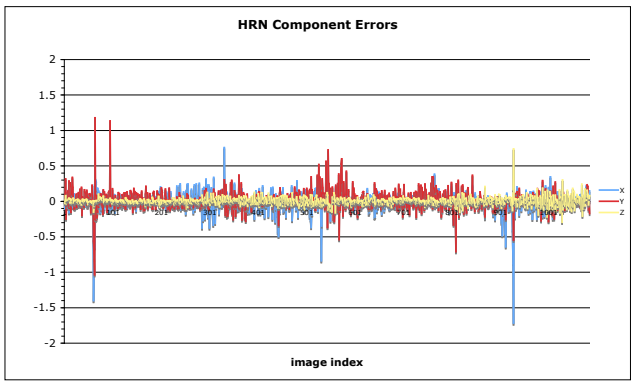


Figure 15: HRN estimate errors for 3 coordinate axes.

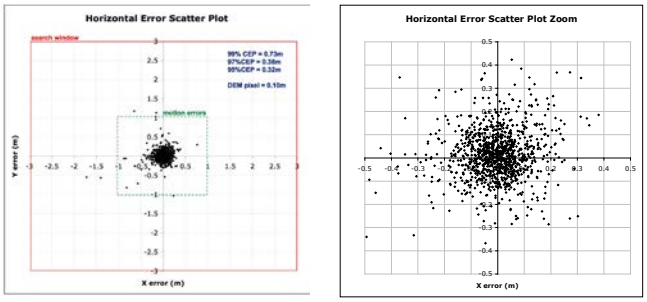


Figure 16: Errors in the horizontal HRN estimate marked as valid: all (left) and zoom to central peak (right).

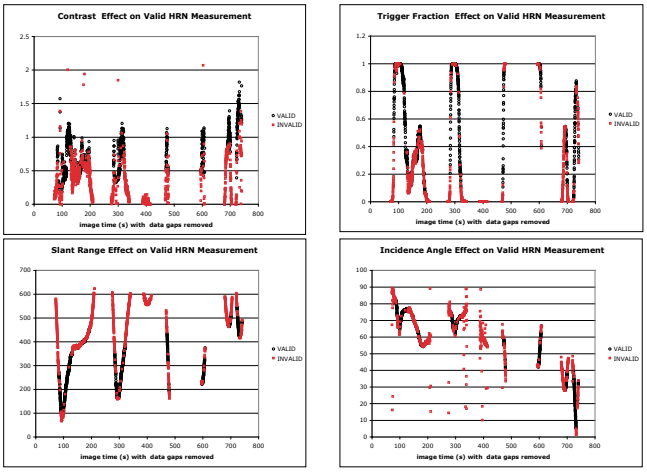


Figure 17: Sensitivity of the HRN algorithm in obtaining a valid estimate relative to the image contrast, trigger fraction, slant range and incidence angle.

The effects of environmental variables on the correlation being marked as invalid were investigated and the results are shown in Figure 17. The top left plot shows the effect of contrast on each image pair. Basically all selected track points with contrast below 0.20m/pixel resulted in invalid correlations. Since 0.20m is the range error of the sensor, this result makes intuitive sense; the terrain is going to have at least as much relief as the noise placed on it by the sensor. Although it was not done in this analysis, it would be a good idea to set the contrast threshold at the range error of the sensor. The top right plot shows the effect of trigger fraction on measurement validity. HRN does not work well at very low trigger fractions, and it is surprising that it actually does work in some cases, for images that have trigger fractions less than 10%. On the other hand, HRN occasionally fails on images that have nearly 100% trigger fraction; however, this failure could be due to many other factors such as poor manual gimbal pointing or track point area not appearing in both images due to high velocity near surface. The bottom right plot shows the effect of slant range. HRN works at slant range ranging from 100m to 550m, but in general performs worse at higher altitudes. The bottom right plot shows that there is no discernable effect of path angle; HRN works at all path angles.

6. CONCLUSION

The FT1 analysis first assessed the flash lidar in terms of its sensitivity (pixel trigger fraction) and range measurement precision as a function of path angle and slant range. The results show that the lidar has a worst case range error (random noise) of 0.20m one sigma. When imaging though the gimbal window (28% loss), the lidar has a maximum range between 400m for nadir viewing and 250m for oblique viewing (15˚ from horizontal). These results are for dry lakebed material that does not necessarily have the same reflectivity as the lunar surface.

Hazard detection results show that the lidar and algorithm have a 100% detection rate for 70cm high hazards while keeping the probability of a false positive detection less than 20% per VFDE. Smaller hazards can be detected but at the expense of more false positives. The hazard detection performance is not adequate to detect small hazards required for safe lunar landing and the main source of the difference is the relatively high range noise of the commercial sensor flown in FT1.

The hazard detection results were also compared to results obtained from simulated flash lidar imagery with 20cm range precision. The real and simulated hazard detection results agreed well when the flash lidar is in its nominal operational regime. This validated the implementation of the ALHAT flash lidar simulator, which was used in the Monte Carlo simulation in POST II [2]. Simulated results for a sensor having a 5cm range precision were also generated, and these showed that the detection of 30cm high hazards with a low false positive rate was possible. This field test analysis, along with the recent comprehensive analysis of the HD tests space through Monte Carlo simulation has advanced the HD algorithm from TRL4 to TRL5.

The critical components of the HRN algorithm were also tested using consecutive flash lidar images. After processing more than 2000 image pairs, the results show that the HRN algorithm provides motion estimates with an accuracy of 0.38m (95^{th} percentile circular error probability), while it is able to reject most incorrect estimates using internal algorithm checks. The processing of a significant amount of real data, combined with a recent stand-alone simulation of the HRN algorithm with lunar terrain, has advanced the HRN algorithm from TRL3 to TRL4.

ACKNOWLEDGMENT

The work described in this publication was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract from the National Aeronautics and Space Administration. The work was funded by the NASA Exploration Technology Development Program and would not have been possible without the flash lidar sensor provide by NASA Langley and the field test system provided by JPL.

REFERENCES

[1] A. Bulyshev, D; Pierrottet, F. Amzajerdian, G. Busch, M. Vanek and R. Reisse, "Processing of three-dimensional flash lidar terrain images generating from an airborne platform" *Proc. SPIE, Vol. 7329,* April 2009.

[2] Jody Davis et al., "Advances in POST2 End-to-End Descent and Landing Simulation for the ALHAT Project," *Proc. AIAA/AAS Astrodynamics Specialists Conference*, Honolulu, HI, August 2008.

[3] Chirold Epp and Tom Smith, "Autonomous Precision Landing and Hazard Detection and Avoidance Technology (ALHAT)," *Proc. IEEE Aerospace Conf.*, Big Sky, MT, March 2007.

[4] A. Johnson, A. Huertas, R. Werner and J. Montgomery, "Analysis of On-Board Hazard Detection and Avoidance for Safe Lunar Landing," *Proc. IEEE Aerospace Conf.*, Big Sky, MT, March 2008.

[5] Andrew E. Johnson and Miguel SanMartin, "Motion Estimation from Laser Ranging for Autonomous Comet Landing," *Proc. Int'l Conf. Robotics and Automation*, pp. 132-138, April 2000.

[6] A. Huertas et al., "Performance Evaluation of Hazard Detection and Avoidance Algorithms for Safe Lunar Landing," *Proc. IEEE Aerospace Conf.* Big Sky, MT, March 2010.

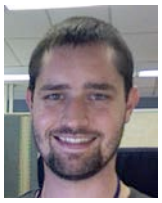
[7] J. Keim. et al., "Field Test Implementation to Evaluate a Flash Lidar as a Primary Sensor for Safe Lunar Landing," *Proc. IEEE Aerospace Conf.* Big Sky, MT, March 2010.

[8] R. Stettner and H. Bailey, "Eye-Safe Laser Radar 3D Imaging," *Proc. SPIE*, vol. 5412, SPIE, Apr. 2004, pp. 111–116.

[9] C. Stewart, "Robust Parameter Estimation in Computer Vision," *SIAM Review* Vol. 41, No. 3, 1999, pp. 513-537.

BIOGRAPHIES

Dr. Andrew E. Johnson *is a Principal Member of Technical Staff in the Optical Navigation Group at JPL. He is the JPL Project Manager and terrain sensing algorithm lead for the Autonomous Landing and Hazard Avoidance Project which is developing technology for safe and precise landing for the next generation manned lunar lander. At JPL he works on development, validation and flight implementation of computer vision systems for planetary landers and Mars rovers.*



Jason A. Keim *received his BSBME from the University of Southern California in 2002 and MSME from California State University, Los Angeles, with emphasis in rigid body dynamics and control. In 2002, Jason joined the Guidance and Control Analysis Group at the NASA Jet Propulsion Laboratory. In that time, Jason's primary focus has been the development and validation of formation flight algorithms and technologies for missions such as NASA Starlight, TPF and DARPA F6. Additionally, he has contributed to the autonomous surface operations of the Mars Science Laboratory, data processing and trajectory reconstruction for the Autonomous Landing and Hazard Avoidance Technology program, and other research and technology development programs.*

Tonislav Ivanov *received his MEng in ECE from Cornell University in 2007. After graduation, he joined the Computer Vision Group at JPL. He works on lidar data processing, field test setup, and is helping develop the hazard detection algorithm for the Autonomous Landing and Hazard Avoidance Project. He also works on recognition using stereo data for robot human awareness and lunar terrain characterization for future missions to the Moon.*