AUTOMATED EXTRACTION OF VERTICAL OBSTRUCTIONS FROM LIDAR DATA

David A. Kohlbrenner, Geospatial Specialist
Todd A. Jamison, Chief Scientist
Lindsay M. Kedzierski, Geospatial Analyst
Observera Inc.
3856 Dulles South Court, Suite I
Chantilly, Virginia 20151
dkohlbrenner@observera.com
tjamison@observera.com
lkedzierski@observera.com

ABSTRACT

High quality aeronautical information is a necessity for safety of flight and navigation both in the military and commercial communities. Automated collection of Vertical Obstruction (VO) data in support of safety of flight is a critical technical capability desired by VO collection and maintenance programs. Current collection and maintenance regimes rely on various reporting and manual extraction methods, with varying degrees of quality. The US military services would like a global VO database containing 90% of all VOs over 150 ft. The ability to meet this standard is very high on the requirements list levied on the National Geospatial-Intelligence Agency (NGA) by the military services.

Lockheed Martin Denver Aerospace (LMC) developed an algorithm for extracting VO data from point cloud information. This algorithm was applied to data collected by the LIDAR sensor on the US Army Buckeye platform. The Buckeye system collects both LIDAR data and coincident 4” GSD digital imagery. Observera evaluated the quality of the VO data that was generated by the LMC algorithm to determine its utility for automatically populating and updating VO databases.

The evaluation compared the algorithm generated VO detections to evaluator derived “truth.” The evaluation data set was collected over Iraq during 2005-2007. The evaluation area included approximately 1800 SqKm over 28 sites, containing approximately 1400 VOs over 100 ft varying in terrain type and VO type. The analyst “truth” data was developed using ortho-images generated from the Buckeye digital imagery and both QuickBird and WorldView 1 satellite imagery. The height and type for each VO over 70 feet in the study area was measured and recorded, then compared it to the algorithm VO detections.

Analysis showed that the LIDAR-algorithm system had a detection probability (Pd) of 0.75 for VOs greater than 100ft at a False Alarm Rate (FAR) of 0.036 FA/SqKm. The performance came very close to meeting objective requirements for automated VO detection established during previous studies. Discussion includes definitions of detection categories (i.e., what is a VO, what is a false alarm, what detections fall in neither category). Additional results presented include Pd and FAR as a function of other variables, false alarm attribution, outlier analysis, height measurement analysis, confidence analysis, and analysis of “Non-Traditional” VO detections.

The algorithm development by LMC and the evaluation by Observera were performed under separate US Government contracts.

Key Words: LIDAR, automated detection, vertical obstruction, Buckeye, NGA, performance evaluation.

INTRODUCTION

High-quality, global Vertical Obstruction (VO) data is a goal of the worldwide aeronautical community in order to improve safety of flight and navigation. The US National Geospatial-Intelligence Agency (NGA) is the functional manager for VO data for the US military, and maintains a database of VOs called the Digital Vertical Obstruction File (DVOF), which is available through several sources. While global coverage is a goal of the system, current methods to collect, populate and update the databases are resource intensive. Furthermore, the quality and reliability of current VO information from existing manual collection programs varies significantly depending up on its source.

Development of automated methods for extracting VO data from remote sensing instruments may be a technological solution for populating these databases in an efficient manner. In particular, there has been interest in extracting VO data from both synthetic aperture radar (SAR) imagery and Light Detection and Ranging (LIDAR) data because of their ability to generate detailed surface elevation information.

The US military services have established a requirement for the global VO database to include 90% of all VOs over 150 ft. The scope of populating such a worldwide VO database is potentially a very costly endeavor. Complexities include the sheer volume of existing VOs, the need to periodically update the database, and the need to characterize the location, height, and type of each VO.

Robert Harrison, David Yip and their group at Lockheed Martin Denver Aerospace (LMC) developed an algorithm for extracting VO data from LIDAR point cloud data. They used LIDAR data collected by the US Army Buckeye program and provided by the US Army Topographic Engineering Center, now called the Army Geospatial Center (AGC).

Observera, Inc. was tasked by NGA to evaluate the quality of the VO data extracted by the algorithm from the Buckeye LIDAR data. The results serve several purposes: (1) they support understanding the ability of the Buckeye LIDAR along with the LMC algorithm to provide useful VO data; (2) they provide a baseline for understanding the causes of missed detections and false alarms, and identification of potential mitigation strategies; and (3) they provide a basis for establishing predictions, requirements and expectations for future automated LIDAR VO detection concepts.

The two primary metrics of the evaluation were the Probability of Detection (Pd) and the False Alarm Rate (FAR), although we also performed false alarm attribution, outlier analysis, height measurement analysis, confidence analysis, and analysis of “non-traditional VO” detections.

RELATED WORK AND REQUIREMENTS

Observera has performed several evaluations of automated VO algorithms for US Government agencies since 2002. As part of our earlier work, we developed a set of derived requirements in conjunction with NGA Office of Global Navigation (PV) and the US Air Force against which automated VO algorithms could be judged. These requirements were needed because the general nature of the services requirements were not specific enough for acceptance or rejection of an automated solution. For example, “all VOs above 150 ft” could be met by placing 400 ft VOs in the database every 50 ft or so on the Earth’s surface. Of course, this would not meet the “real” requirement, because the value of the VO data for low-level flight would be nil.

In establishing requirements, we focused on the combination of Pd and FAR that would be acceptable for safety of navigation and would support confidence in the data by pilots, navigators and mission planners. It is important to understand that for any type of automated detection system, Pd and FAR are intimately tied together. The sensitivity of an automated detection system can generally be adjusted up or down, having the effect of increasing or decreasing both the Pd and FAR. Thus, as one increases the sensitivity of a detection system, more correct object detections will occur (increased Pd), but more false detections will also occur (increased FAR). The sensitivity is usually adjusted to a point that reflects the best relative values of detection rates and false alarms. Because of this dependency, the Pd and FAR must be specified as a pair of requirements that must be met simultaneously.

We established a set of threshold requirements – the minimum needed in order to be considered acceptable – and a set of objective requirements – the goal for an operational capability. We established requirements for detection of VOs in order to meet the Air Force requirement of 150 ft and above, and another set of requirements for detection of VOs between 100 ft and 150 ft, as there were additional needs expressed for detecting shorter VOs. We note that there is interest in VOs that are less than 100 ft as well, however, the feeling at the time the requirements were established was that the technology to automatically detect and separate VOs from other objects at such short heights was infeasible for the foreseeable future. Table 1 gives the derived requirements for Pd and FAR for automated or semi-automated VO detection systems.

Table 1. Derived requirements for automated / semi-automated VO detection

Threshold Requirements	Pd	FAR
VOs ≥ 150 ft	>40%	<0.05 FA/SqKm
VOs ≥ 100 ft and < 150 ft	>25%	<0.05 FA/SqKm
Objective Requirements		
VOs ≥ 150 ft	>90%	<0.01 FA/SqKm
VOs ≥ 100 ft and < 150 ft	>75%	<0.01 FA/SqKm

It may seem to some that the FARs are extremely stringent. When establishing the requirements for VO detection, one might be inclined to believe that the Pd is the most important issue, and that any FAR is acceptable, since one does not want a flyer to fly into a VO because it was not detected. However, this thought process can be very misleading. A user's confidence in the VO database is related to the database's accuracy. The problem with higher false alarm rates is that VOs are relatively scarce objects, so false alarms contaminate the database very quickly. We use a measure called *contamination rate* to describe this and it represents the ratio of false alarms to the total number of database entries. As an example, we estimate that the average number of VOs per land area for major areas of interest could vary from about 0.1 VO/km^2 to 1.0 VO/km^2. Using a value of 0.25 VO/km^2 – a middle ground value – the threshold requirements (Pd=0.4 and a FAR=0.05) yield a contamination rate of 33% - meaning that 1/3 of the objects detected by a system having those characteristics would be false detections. This is a significant factor when it comes to confidence in the data. It turns out that the "contamination rate" is the requirements driver to keeping the FAR at very low levels.

We also established a requirement for height accuracy associated with the detections. This requirement stated that the measured height should be within 20% of the actual height, at least 90% of the time. This is a relatively loose requirement, since most aviators give themselves wide horizontal and / or vertical margins around any VO. However, as we shall see, this requirement may significantly affect detection performance.

We believe that these derived requirements can serve as a baseline against which most automated / semi-automated VO detection capabilities being considered for use by the NGA can be judged. For the purposes of this evaluation, we were not specifically testing against these requirements, but did use them as a reference for our analysis.

STUDY AREA AND DATA

The data were acquired over Iraq between 2005 and 2007 by the US Army Buckeye system during operational demonstrations within that country. The full collections covered an area of over 21,000 square kilometers from 145 mostly urban sites that were subdivided into 595 approximately 6 km x 6 km tiles. The Buckeye LIDAR payload was an Optech Airborne Laser Terrain Mapper (ALTM) 3100 LIDAR system that was used to generate digital elevation models of the collected area. The ALTM operates as a time-of-flight measurement system. A laser pulse is emitted by the sensor and the time it takes for the pulse to strike a surface and return to the receiver is measured. This roundtrip time, when combined with the position and attitude of the platform and sensor look angle, is used to generate two 3D coordinates at each surface location (Fischer *et al*., 2008) – one for the first return and one for the last return of each pulse. The nominal horizontal accuracy is estimated to be around 1 m and the vertical accuracy is less than 1 m, although these values were not verified by the authors. The Buckeye system also includes an electro-optical (EO) sensor – a color camera – that delivers high-resolution digital imagery with approximately a 4" Ground Sample Distance (GSD) that was collected coincident with the LIDAR. Most of the data was collected at a nominal collection altitude above 9,000 ft. The LIDAR from multiple passes was combined into tiles for use and dissemination. The imagery was orthorectified and mosaicked into large tiles for each site using the LIDAR derived DEMs. Due to the mosaicking processes and the potential for multiple flights over the same area, it is highly likely that some portions of the imagery data and LIDAR data were not instantaneously coincident.

The extraction of VOs was not part of the intended use of the LIDAR data; however, once it was suggested both the AGC and NGA saw the potential value for VO applications. The processing executed by the LMC algorithm extracted over 18,500 VOs from the 595 tiles of "first return" XYZ point cloud data. LMC did not use the "last return" point cloud data that was also available for each tile. Of the 595 tiles processed, 133 tiles did not have any algorithm-detected VOs. For the evaluation, we obtained the LIDAR XYZ point clouds as text files and the algorithm results from LMC, and we acquired the 4" coincident imagery (MrSID files) of the same tiles from the Buckeye team at the AGC. We also obtained QuickBird and WorldView 1 imagery from NGA to extract the height values for the "truth" data.

Resources were not available to evaluate all of the tiles, so a statistical sampling approach was used. In particular, one of our evaluations objectives was understanding the relationship between performance and environmental factors, such as terrain type and terrain cover. We established a statistical sampling objective of 1800 VOs, which translated into approximately 55 tiles or about 5% of the total collected area. Our tile selection method was manual, and was based on a somewhat random selection of tiles designed to give a fair distribution of terrain types and terrain cover. One of the biggest issues in attempting this "fair" distribution was that there were very few tiles with rugged terrain or heavy forests – it is Iraq, after all. We were concerned about biasing our results with too many tiles that had no VOs, so we limited such tiles to 2. Otherwise, all tiles were selected without regard for the quality of LIDAR data or VO results. Ultimately, the evaluation included 58 tiles from 28 sites covering 1,789.65 Sq Km. Figure 1 shows the distribution of terrain type/land cover for the analyzed tiles.

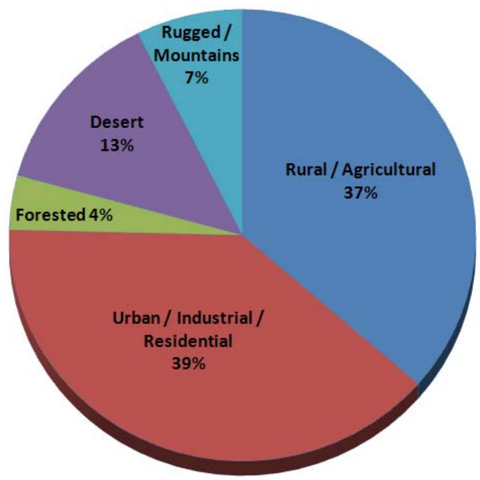

Figure 1. Terrain type/land cover of analyzed tiles.

THE VO DETECTION ALGORITHM

The VO detection algorithm uses the first return information from the LIDAR to determine the likely presence of a VO. While the details are proprietary, the algorithm compiles evidence to support a decision that a VO is present at each ground location in the LIDAR tile. Various parameters control the operation of the algorithm, such as:

- The threshold height above ground level (AGL), above which objects are candidate VOs,
- A maximum cutoff height, above which objects are considered too high to be VOs,
- The minimum spacing between VOs,
- The assumed height of ground clutter, and
- The size of the support region and the amount of support required in order to declare a VO.

Prior to the start of the evaluation, the LMC engineers had performed an initial run of their algorithm on the LIDAR data, the results of which were discussed with both NGA and Observera personnel. As a result of those discussions, the algorithm was modified slightly and run with a different set of parameters to provide results for the evaluation. One particular modification was the addition of a single "confidence" metric that was intended to quantitatively reflect the evidence supporting the detection. In the initial algorithm, there were four metrics that measured different aspects of the quality of the detection. While these were still reported, a fifth metric that summarized these was synthesized by the LMC engineers to represent this confidence factor. This confidence factor was used in key parts of our evaluation. The list of the parameters used in the runs-for-record for the evaluation are given in Table 2.

Table 2. Key algorithm parameter settings

Minimum VO Height Threshold	80 ft (24.4 m)
Maximum cutoff height	2300 ft (701 m)
Ground clutter height	50 ft (15.24 m)
Radius of support area	30 m (~100 ft)
Minimum VO spacing	20 m (~65 ft)
Minimum support required	3 points

Although the objective of our evaluation was to assess VO detections above 100 ft, the height threshold was set to 80 ft because the height measurement process had errors associated with it, which were unknown at the start of the evaluation. In order to assure that underestimation of the height of a VO did not cause it to be eliminated from the results, we chose a 20 ft margin of error for the algorithm.

LMC generated results for each LIDAR point cloud tile, which were provided in both a comma separated value (.csv) text file format and an ESRI® Shape file format. Each VO record had eighty-seven attribute fields. These included such parameters as the VOs position, height, units, date, datum, quality metrics, the confidence metric, etc. Another attribute associated with the Confidence Metric is a tag to identify flying objects or outliers. Since the LIDAR point cloud is of such high quality, it has the ability to capture planes and other objects in flight. To accommodate for these types of detections and identify them appropriately, the developers devised a field that the algorithm populated with either a “Yes”, “No”, or “Unknown”. If the algorithm populated this field with a “Yes” it would also set the Confidence Metric to 0.0%. To meet this condition, the VO identified must not have any point cloud data for 344 feet below its highest point. This indicates a scenario where the VO is most likely associated with a feature that is not ‘connected’ to the ground.

EVALUATION METHODOLOGY

The evaluation was based on a comparison of algorithm generated VO detections to evaluator derived “truth” data. A statistical sample of the collected area was selected for the evaluation. Analysts derived “truth” data for the selected areas, and then assessed each algorithm detection against the truth. Scoring relied on an agreed upon set of definitions for different types of detections. The results were aggregated and analyzed by the authors. The evaluation was conducted on-site at the headquarters of Observera, Inc. in Chantilly, Virginia.

Definitions

Prior to analyzing the data, we worked with NGA and military service representatives to define the classes that would be used in characterizing vertical obstructions for this evaluation. This was necessary (as experience showed from prior studies) because there had been and continues to be some confusion and discussion about how exactly a VO should be defined and how automated systems should be graded. Specifically, the NGA DVOF program has a definition of a vertical obstruction that was generated without the foresight of automated technologies. It is important to recognize that there are classes of vertical objects that do not fit the definition, but that, if detected by an algorithm, do not typically have a negative impact on the database or its users. Thus, it was critical to agree upon the definitions used in scoring the detection system prior to the assessment phase of the evaluation. As a result we have three primary categories of detection assessments.

We defined the term “Traditional VO” to refer to fixed, man-made objects standing above the surrounding terrain. Traditional VO detections are those that are most important to the VO user community, and are used as the basis for the probability of detection calculations within our evaluation. The four primary classes of Traditional VOs are:

- Towers
- Cranes
- Buildings and Bridges
- Tethered Balloons

We defined the term “Incorrect VO Detection” to refer to detections for which no associated vertical object could be identified. These detections are the “False Alarms” in our evaluation and provide the basis for the False Alarm Rate calculations. The specific classes of False Alarms are categorized by their underlying causes, which vary from detection system to detection system. In the case of this evaluation, we only positively identified one underlying cause – flying birds. All other incorrect detections are classified as "Unknowns".

We defined the term "Non-Traditional VO Detection" to refer to vertical objects that are not used in the calculation of Pd or FAR. Most of these do not fit the "Traditional VO" definition, but that are, in fact, vertical objects that, if inserted into the VO database, would not negatively affect database users. We note that some argue that several of these non-traditional VOs – for example trees above 100 ft – should be classified as traditional VOs, however, this significantly complicates the "truth" data development. We further classify non-traditional VOs as follows:

- Short Traditional VOs – Objects that meet the definition of a Traditional VO, but are shorter than the height threshold of our evaluation (<100 ft). This category is necessary because the algorithm height threshold is set to 80 ft and so it is likely to detect many VO objects less than 100 ft.
- Tall Trees – These are trees that exceed the height threshold used for the evaluation (≥ 100 ft).
- Natural Features – These are naturally occurring objects taller than the algorithm's threshold (≤ 80 ft), which includes short trees or forests (between 80 and 100 ft) and abrupt terrain variations, such as cliffs.
- Extra Hits on Same Feature – This class is associated with large features that generated multiple detections, such as buildings. Properly, only one detection on such a feature is used in the Pd calculation and the other detections are put into this category. This is necessary because the algorithm does not currently aggregate neighboring detections into a larger "area" VO.
- Aircraft/Non-Tethered Balloons – Any free-flying, man-made aircraft is included in this category.
- Extra Hits on Same Bird Flock – This category is used when a larger flock of birds results in more than one false detection. Since the detections are locally related, it was decided that only one of the detections should be counted in the false alarm category, and the others are classified in this category. (This is False Alarm equivalent to the "Extra Hits on Same Feature" category.)
- Power Lines – Power lines are actually a type of traditional VO classified as a "linear VO feature" in DVOF, however for this evaluation only the power line pylons are included in the "Traditional VO" category because the direct detection of the power lines themselves from LIDAR was expected to be very unlikely and the LMC algorithm was not designed to detect linear features of this type. In fact, the LIDAR did occasionally detect the power lines directly and the algorithm did occasionally put VO detections along the power lines. Moreover, power line corridors can easily be deduced from the detected pylon locations.

Sample Selection and Truth Data Preparation

We executed a quick review of each of the imagery datasets that was delivered to determine terrain type / cover. The five terrain categories we settled on were Rural/Agricultural, Urban/Industrial/Residential, Forested, Desert, and Rugged/Mountains. This information was then cross-referenced with a list of sites where commercial satellite reference imagery was available for roughly the same time period. We eliminated the Baghdad area, due to its high density. We selected 124 candidate tiles for the evaluation based on getting a reasonable cross-section of terrain types and limited the selection to two tiles with no detections. Worldview® and Quickbird® datasets with a 0.5-1.0m GSD (basic 1B format) were obtained over these tiles. Some tiles were removed from the pool because the reference imagery had issues, such as clouds, low light, high sun angle, etc., that limited the ability to measure VOs.

We prepared the "truth" data using brute-force examination of the 4" Buckeye orthoimages using ERDAS Imagine®. Each site/tile was systematically reviewed without first knowing where the algorithm had detected VOs. The process involved choosing a site, overlaying the tile footprints and generating an analyst grid. The analyst grid was a series of 200 m x 200 m (.04 Sq km) review boxes that covered the entire site. They were generated so the analyst could efficiently review the entire scene in a box-by-box fashion while not overlooking any areas. To aid in this, the analyst marked each row of the grid as they were completed. Whenever a potential VO was found, the analyst used the RemoteView Pro® software to locate and measure the height of the object in the commercial satellite imagery. RemoteView Pro® was used for this part of the process since it was the only tool available that had a built-in mensuration toolkit. Most often, the height was measured using the top of the feature to the top of shadow (the most accurate method), although sometimes other methods were used. If a VO was found to be greater than 70 ft AGL, then the analyst would enter the feature as a point attribute in a Shape file using the 4" Buckeye imagery in ERDAS Imagine® and attribute the feature with a unique ID, height in feet, class, type (sub-categories of four of the classes), and terrain surrounding the feature. Table 3 gives a summary of the Class and Type (where applicable) definitions that were used for the evaluation.

It should be clear from this description that the "truth" data should not be considered perfect data, hence our use of the quotes around the term. However, given the resources available, this was a reasonable compromise for this evaluation. We have very high confidence that the Pd of the truth data is high (~99%) and that its FAR is near zero. We anticipated the potential for missed VOs within the truth data, and compensated for that in the assessment by reviewing each algorithm detection on the imagery. Most of the uncertainty in the truth data is in the height measurements, which have an unknown associated accuracy. We point this out in our height comparison section.

Table 3. Summary of class definitions, including type sub-class, if applicable (*denotes Non-Traditional VO types)

Class Definitions
1- Towers ***(Tower Types: telephone pole / A type communication / A type power / T type power / light pole / minaret / smokestack / water / tower / communication antenna / A type light)***
2- Cranes ***(Crane Types: mobile / tower / gantry / hammerhead)***
3- Buildings and Bridges
4- Tethered Balloons
5- False Alarm ***(False Alarm Types: birds / unknown)***
6- Short Traditional VO (< 100 ft feature)*
7- Tall Trees (≥ 100 ft)*
8- Natural Feature* ***(Natural Feature Types: ≤ 100 ft short trees, forest / terrain)***
9- Extra Hits on Same Feature*
10- Aircraft / Non-Tethered Balloon*
11- Extra Hits on Same Bird Flock*
12- Power Lines*

Assessment

The detected VOs were directly compared to the "truth" VOs by displaying them simultaneously in ERDAS Imagine. We assessed the primary detection type – Traditional, Non-Traditional, or Incorrect – for each VO detected by the algorithm, as well as any subcategories associated with the detection. We also identified missed detections. For metric purposes, we used the "truth" height for determining the correct "Traditional" detections, so that only those VOs in the truth set with a measured height greater than or equal to 100 ft AGL were compared to the algorithm detections for scoring correct or missed detections.

Where the algorithm detected a VO that did not correspond to a truth VO above the height threshold, our procedure was more involved. First, we checked to see if it corresponded to a truth VO that was between 70 and 100 ft, in which case it was classified as a non-traditional detection of Class 6. If it did not correspond to a truth VO, we analyzed the Buckeye imagery to determine if a VO had been missed in the truth generation. In the case that it was, the truth file was amended and the VO was correctly classified as a traditional VO detection. As a quality check, only about 1% of the truth VOs were found this way. If it was not, then we attempted to determine if some other non-traditional VO had caused the detection, in which case it was classified as such. In the event that no traditional or non-traditional VO caused the detection, it was categorized as an Incorrect Detection (i.e., False Alarm) and a detailed analysis of the imagery was performed to determine if strong evidence of birds existed in the imagery so it could be classified appropriately.

The assessments were entered into a Microsoft® Excel database that automatically tabulated all of the results and generated some of the analysis metrics. In the spreadsheet, the data for each site was contained on one worksheet. Each truth VO and each detection for a site was contained on that site's worksheet. In addition, the statistics for all VOs and detections for the site were automatically summarized on the sheet. A summary worksheet then aggregated the data from all of the site sheets to provide overall assessments.

ANALYSIS AND RESULTS

We analyzed the data from a number of different perspectives, in order to best answer the questions posed by our customer. We would like to note up-front that the results obtained during this evaluation are applicable to the VO detection "system" consisting of the Buckeye LIDAR and the LMC Denver VO detection algorithm. It is difficult to separate issues related to either this LIDAR or this algorithm alone. In addition, the Iraq LIDAR data was collected over areas with limited environmental conditions – dry atmosphere, limited vegetation types, minimal forests, low-density canopies, and minimal rugged areas – and the evaluation did not include any VOs directly in bodies of water. Thus, caution should be exercised in translating these results to performance of other LIDAR-algorithm systems or other types of terrain cover.

Detection Summary and Key Metrics

The algorithm detected 4,731 objects in the 1,789.65 SqKm study area. Of these detections 1,054 were correct detections of Traditional VOs ≥ 100 ft and an additional 352 detections where the analyst identified the object to be < 100 ft. The algorithm detected 65 objects that were identified to be incorrect detections (false alarms). In the same study area, Observera analysts identified 2,168 objects as Traditional VOs where 1,410 were measured to be ≥ 100 ft and 758 were measured to be < 100 ft. Thus, the overall Pd (VOs ≥ 100 ft) was 75% and the overall FAR was 0.036 FA/SqKm.

In the Non-Traditional VO detection category the algorithm detected 3,612 objects which include the aforementioned 352 Traditional VO features that were less than 100 ft. These detections are summarized in Figure 2 below. It is interesting to note that trees (both short trees < 100 ft and large trees ≥ 100 ft) make up 82% of the Non-Traditional VO detections.

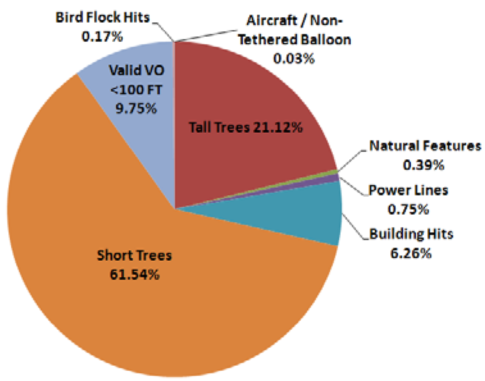

Figure 2. Non-Traditional VO detections.

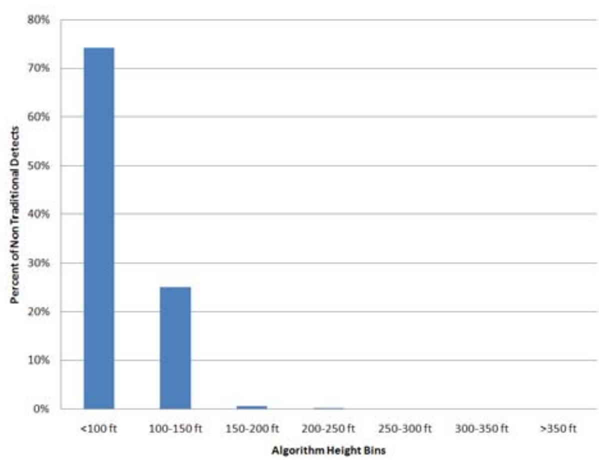

Figure 3. Non-Traditional detections by estimated algorithm height.

False Alarm Analysis

Of the 65 Invalid VO detections (i.e. "false alarms"), about half were identified as Birds and half were Unknown. It should be noted that most of the unknown false alarms exhibited similar characteristics in the LIDAR point clouds as the birds in flight. For this reason, we suspect that birds were the cause of many of these "Unknowns" and either may not have been visible in the imagery due to poor contrast, or the imagery may not have been collected at the same time as the LIDAR.

Further examination of the false alarms by placing them into height bins produces the histogram in Figure 4. It is important to note that the majority (79%) of false alarms are < 150 ft and should be the focus of any false alarm mitigation efforts.

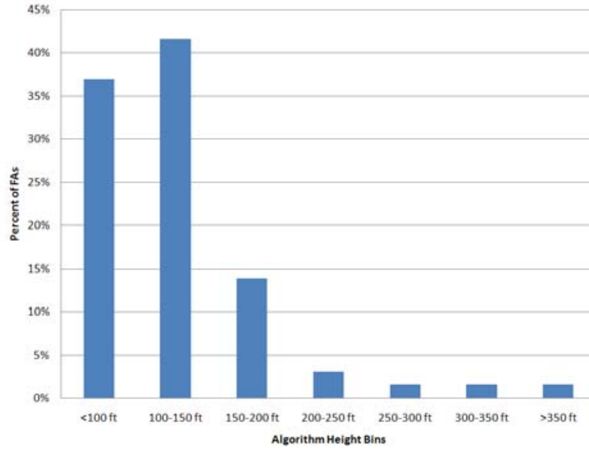

Figure 4. False alarm detections by algorithm height bin.

Analysis by Site

During the assessment process we compiled the results on a site by site basis before they were summarized into the overall key metrics. This allowed for the Pd and FAR to be calculated for each site, and the identification of sites that exhibit outlier behavior. Figure 5 shows a histogram of the Pd for the 28 sites that were analyzed (less the two zero-VO sites), sorted in descending order. Note how three of the sites have a relatively low Pd (< 0.5) compared to the rest. Figure 6 shows the FAR of all 28 sites, sorted in ascending order. Note how three of these sites have a relatively high FAR (> 0.1) while the remainder have a very low or even zero FAR. It is interesting to note that the low Pd and high FAR sites do not overlap, indicating that the underlying mechanisms of the outlier behavior are probably different.

An important caution when dealing with outliers: it is tempting to assume that outliers are not representative of normal system behavior and can be ignored. Certainly, analyzing data with outliers removed is useful, as it provides insight into potential best-achievable performance and it can help us to identify issues requiring further study; however it must not be assumed that outliers can be ignored. In this evaluation, the sources of outlier behavior are not known and there could be multiple underlying causes. Outliers must be attributed to fixable anomalies before they can safely be ignored when projecting operational capabilities of this or any similar system. Consider, for example, that most false alarms may be caused by birds. If that is the case, then the higher FARs are not due to a physical process, but rather an environmental factor for which a better detection strategy would be needed.

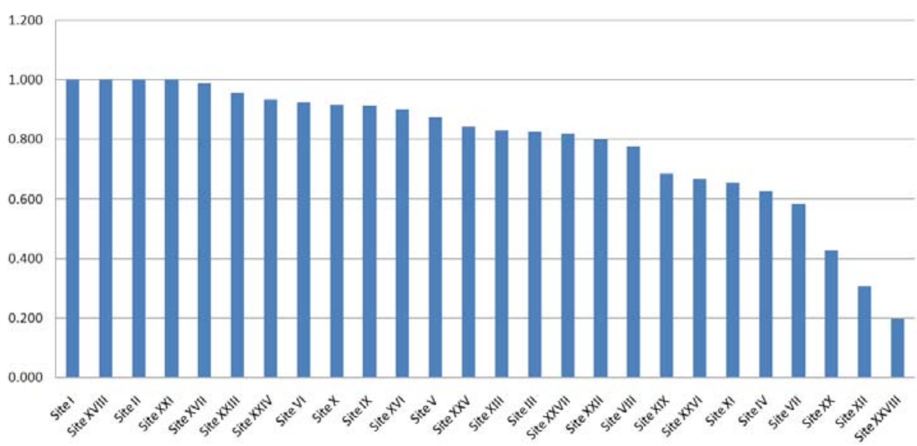

Figure 5. Probability of detection (Pd) by site.

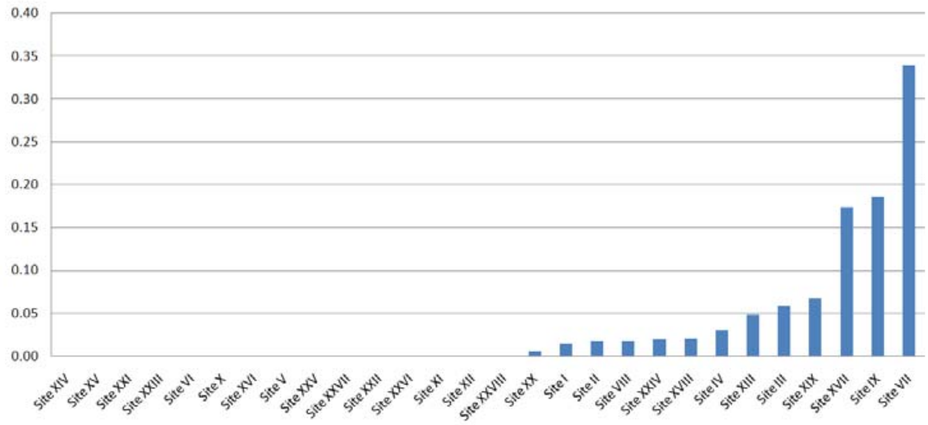

Figure 6. False alarm rate (FAR) by site.

Detection Analysis

Figure 7 shows a histogram of the detection percentage of all of the Traditional VO detections, sorted into 15 ft height bins. This shows a trend where the five lower height bins have a lower Pd (≤ 82%) as compared to the ones in the higher height bins which have a significantly higher Pd. While this information is useful in analyzing the Pd of VOs of varying heights, it should also be pointed out that the aforementioned five bins represent 94% of the detections while the top six bins represent the remaining 6% of the detections.

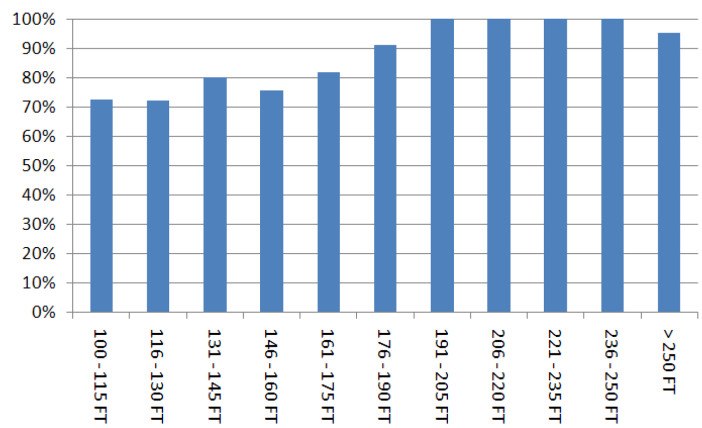

Figure 7. VO detection percentage by height bin.

By sorting the data into the two height ranges we used in our requirements analysis, we can see in Figure 8 that the system exceeds the Pd threshold requirements for both bins and approaches the objective requirements. While this histogram does not take into consideration the FAR it does show us that both height bins are within a two percentage point margin of meeting the objective Pd requirements.

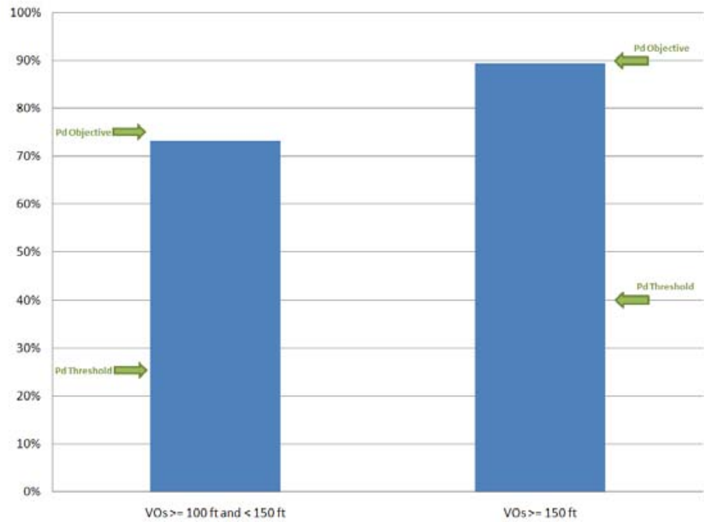

Figure 8. Detection percentage of VOs >= 100 ft and <150 ft & VOs >= 150 ft.

Confidence Analysis

We used the confidence metric to develop a series of Receiver Operating Characteristic (ROC) curves that provided insight into the overall system performance capabilities. The ROC curve is a standard method for finding an optimal balance between Pd and FAR for automated detection systems (e.g., automated target recognition, signal detectors) whose name comes for its use in communication receivers. A ROC curve is formed by sorting data in descending order by its confidence metric, then for each confidence value, plot the Pd and FAR using just the data above that confidence threshold. As Pd increases so does the FAR, but as the confidence threshold decreases, the FAR increases faster than the Pd. ROC theory allows for finding the "optimal" operating point for a given application by selecting a Pd and FAR along the curve at which to operate and thresholding the results at the corresponding confidence level.

Figure 9 shows ROC curves from five different filtering scenarios. Plot 1 (red) represents all of the detections made by the algorithm (≥ 100 ft). Note its relationship to the threshold and goal requirements (blue dots). The data meets the threshold requirements for Pd and FAR, but falls short of the objective requirements. Plot 2 (light blue) shows the ROC curve for the detections in the 100-150 ft range and Plot 3 (dark blue) shows the curve for the detections greater than or equal to 150 ft. While Plot 2 is still well below the objective requirement for its height range, Plot 3 approaches very close to its corresponding objective requirement.

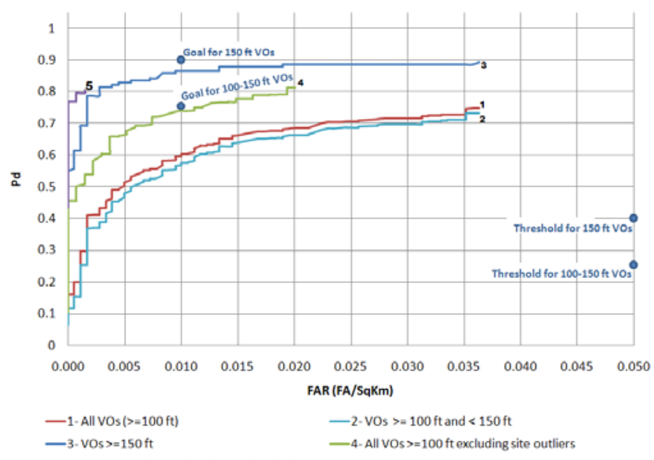

Figure 9. Performance of different operating scenarios.

Plot 4 (green) shows the overall results when the outlier sites are excluded, which is encouraging if the causes of outlier behavior can be identified and mitigated. Plot 5 (purple) shows VOs ≥ 150 ft using a more complex filter.

In this plot, the six outlier sites are removed and the algorithm VOs are thresholded at an algorithm estimated height of 120 ft and a confidence of 20. While this particular filtering approach has a Pd of 80% and does not meet the objective Pd, it does exhibit an extremely low FAR of 0.0015, which is equivalent to 1 false alarm for every 700 km^2!

Height Analysis

Our height analysis was not a true "error" analysis, since both the "truth" and algorithm measurements are imperfect. However, we did compare the height values and found a root-mean-square-deviation (RMSD) between the data of 22.4 ft and a negative bias of the algorithm relative to the analyst height of 11 ft. When analyzed as a function of height and assuming that the "truth" heights are accurate, the deviations meet, but do not exceed, the height accuracy requirement of 20%. That said, our analysis of the height information shows that the height estimates are probably worse than this, but the 80 ft threshold limited our ability to analyze larger errors. We are confident that poor height estimates contribute to a significant number of missed detections for the lower height bins and we believe additional work to improve height estimation accuracy is warranted to improve Pd.

CONCLUSIONS AND THOUGHTS ON IMPROVEMENTS

An evaluation of an automated LIDAR VO extraction system was conducted. The Buckeye LIDAR combined with the LMC VO algorithm provides very good to excellent detection rates at very low false alarm rates. Further analysis of the missed detections in the context of the algorithm may be of value for improving the algorithm, as we suspect some may be due to the "narrowness" of features and some may be due to LIDAR quality issues. Many of the missed detections that were analyzed often had sufficient amount of data points within the point cloud that could have supported detection. Although we noticed qualitative differences in the LIDAR datasets, insufficient information was available regarding flight, sensor, and processing differences to draw any conclusions regarding the effects of LIDAR parameters on VO data extraction quality. Future LIDAR collects should include documentation to support a more thorough analysis of these effects.

Analysis of the confidence metric validates its utility for optimizing system performance. Our assessment of likely contamination rates with current performance is around 5% for both height bins, which is acceptable. Improvements to the confidence metric are likely possible with additional effort, which should improve false alarm rejection and reduce contamination rates even further. Post processing of the data may also be of value to database users. Examples include the grouping of neighboring detections into area VO objects or the extraction of power lines from pylon detections into linear VO objects.

While the system performed well at detecting VOs, it does not provide a capability to identify / classify the features – a capability that is desired for support to the DVOF database. Incorporating capabilities to discern a feature was not an intended capability of the algorithm, but may be an area for research that would allow for the reduction of certain types of detections that are not of interest to the users of the data, such as trees.

The results of this evaluation are highly encouraging, and we conclude that automated VO detection from LIDAR is a viable and valuable capability that can meet key VO requirements over non-denied areas.

ACKNOWLEDGEMENTS

The authors would like to thank Lockheed Martin Denver Aerospace and the U.S. Army Geospatial Center for providing data and support, which was crucial in completing this evaluation. The authors would also like to thank the following organizations for providing software and support: Applied Imagery for QT Modeler, Overwatch Geospatial Systems for Remote View Pro, and ERDAS, Inc. for ERDAS Imagine.

REFERENCES

Fischer, R.L., B.G. Kennedy, M. Jones, J. Walker, D. Muresan, G. Baxter, M. Flood, B. Follmer, X. Sun, W. Chen, and J.G. Ruby, 2008. Development, integration, testing, and evaluation of the U.S. Army Buckeye System to the NAVAIR Arrow UAV, *Proc. of SPIE*, Vol. 6963:696318.