Analysis and Removal of artifacts in 3-D LADAR Data

J. Tuley
jtuley@cs.cmu.edu

N. Vandapel
vandapel@cs.cmu.edu

M. Hebert
hebert@ri.cmu.edu

CMU-RI-TR-04-44

August 2004
Updated January 2005

Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

© Carnegie Mellon University

Abstract

Mixed pixels are points in LADAR data which occur when the measurement beam intersects two objects at within a certain range of distances apart. The resulting return signal represents data from both objects; the laser electronics read this as a single point, but not at the real location of either object. As such, mixed pixels present an obstacle to accurate analysis of a scene. We present a survey of mixed pixel phenomena and a method of removing mixed pixels from scenes stored as 3-D points. Our method does not depend on the sensor used to originally collect the data, and is efficient and effective.

Prepared through collaborative participation in the Robotics Consortium sponsored by the U.S Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-209912.

The authors would like to thank Arne Suppe and Christoph Mertz for help in understanding the SICK laser beam shape and imaging the SICK laser beam.

Contents

1 Statement of Problem **5**

2 State of the Art **6**
2.1 Sensing 6
2.2 Mixed pixel detection and filtering 6

3 Mixed Pixel Phenomena **7**
3.1 Sensor Description 7
3.1.1 Physical configuration 7
3.1.2 Beam characteristics 7
3.1.3 Data interpretation 8
3.1.4 Mathematical model 8
3.2 Experimental Setup 9
3.2.1 Test series 1 9
3.2.2 Test series 2 9
3.2.3 Test series 3 10
3.2.4 Test series 4 11
3.2.5 Test series 5 11
3.2.6 Test series 6 11
3.3 Phenomena Observed 11
3.3.1 Edge-occurring mixed pixels 12
3.3.2 Thin-structure formed mixed pixels 13
3.3.3 Mixed pixels unique to SICK-type lasers 14
3.3.4 Mixed pixels forming from interaction with ground 15
3.4 Configurations Producing Mixed Pixels 15

4 Approach to Mixed Pixel Filtering **16**
4.1 Philosophy 16
4.2 Algorithm Variations 18
4.2.1 Surfaceness triggered 18
4.2.2 Surface normal triggered 18
4.2.3 Surface-ness and surface normal triggered 18
4.3 Marking variations 19

5 Performance Calculation **19**
5.1 Description of Metrics 20
5.2 Effectiveness 21
5.3 Efficiency 21
5.4 Performance against Point Clouds 22
5.5 Performance Overview 22

6 Conclusions **23**

References **25**

A Additional Tables and Figures **26**

1 Statement of Problem

Our long term goal is to devise methods for fast and accurate three-dimensional (3-D) representation of natural scenes containing vegetation for autonomous robot operations, both ground and aerial. Tasks envisioned are obstacle detection, path planning, scene understanding, target detection and tracking in cluttered 3-D environments. We have already reported related work in several publications; see [Vandapel et al., 2003, Vandapel et al., 2004, Hebert and Vandapel, 2003]

Perception in terrain with vegetation is particularly challenging because of the porosity of the vegetation, the presence of thin structures and scene occlusions. The high level of clutter introduces occlusions that will make data very dependent on the point of view, limiting the field of view of the sensors. In addition, data integration in time is very difficult. The tree canopy prevents reliable GPS information acquisition, rough terrains are harsh on pose estimation based on an Inertial Measurement Unit and terrain based methods for localization are sensitive to the quality of ground surface recovery process after vegetation filtering.

When using a LADAR[1] system which projects beams of any larger than infinitesimal size, it is possible for a single beam to overlap two objects in the scene with different locations. When this occurs, it is possible for the distance reported by the laser to be the true position of neither object, but instead have its value influenced by both. The point may be reported in front of the foreground object, behind the background, or more likely, somewhere in between. (The sensor used in the experiments described here is capable only of producing returns between the two objects.) Points of this nature are known in literature as *mixed pixels* and can present a serious problem to the process of identifying and categorizing objects in a scene. In this paper we refer to pixels which are not mixed as *correct pixels* to emphasize the fact that these are the "good" returns from the LADAR and truly represent the scene.

Because mixed pixels occur at geometric or photometric discontinuities in the LADAR's view, scenes with high numbers of edges or discontinuities (in color or distance) are more prone to the formation of mixed pixels. This can occur especially with vegetation (e.g., grass, branches) or scenes with thin objects (e.g., wires). In these cases, it is sometimes possible for the proportion of mixed pixels in the scene to exceed that of good pixels, and when objects are spaced closely in the scene it is also possible that the objects become less distinct, sometimes even appearing joined.

Thus it is desirable to remove mixed pixels for several reasons. First, they reduce the overall accuracy of the representation of the scene. Second, they make it difficult or impossible to correctly classify objects in the scene. Toward the goal of improving these applications, we have developed a system for the identification and removal of mixed pixels.

In the next section we will present the state of the art. Following that, we introduce the mixed pixel phenomena we have observed and how they were observed. Then we describe our approach to filtering mixed pixels, and finally discuss the performance of our methods.

[1] Light Amplification for Detection and Ranging

2 State of the Art

2.1 Sensing

Depending on the application, several sensing modalities can be used in terrain containing vegetation: 1) color and multispectal camera for 2-D scene classification to identify vegetative cover; 2) millimeter wave (MMW) radar, synthetic aperture radar, to see through the vegetation to detect target or obstacles even at long range but with low resolution. Global properties can be extracted from the pulse returns about the area illuminated but cannot be used for high fidelity reconstruction [Matthies et al., 2003]; 3) echolocation to see through grass at short range and low resolution [Murray et al., 2003]; 4) stereo vision with specialized pre-processing image filters can used to reconstruct forest scenes in 3-D at short range [Ansar et al., 2004, Ansar and Matthies, 2003]; 5) but LADAR sensors are the most commonly used.

Laser based sensors are widely used in scene modeling, ground robot navigation and aerial terrain mapping. Different technologies can be found: pulsed time of flight (TOF) laser versus continuous wave laser (CW). Narrow beam TOF lasers can provide a single return (first or last pulse) or multiple returns, even complete the waveform of the returning signal [Schilling et al., 2002]. Flash LADARs provide sampled waveforms but the scene is illuminated thoroughly with a set of wider beam laser diodes [Ludwig et al., 2003].

Single return LADAR are sensitive to mixed pixels but multi-return LADAR are as well. Such LADAR will produce two or more returns only if the distance between the two targets illuminated is above some threshold; for example, 1.8 m in the aerial LADAR used in one of our former projects [Vandapel et al., 2003].

2.2 Mixed pixel detection and filtering

The existence of mixed pixels is a long standing problem but very few papers explicitly address the problem.

Continuous wave (CW) LADAR are particularly prone to such phenomena. In [Kweon et al., 1991] Kweon presents a characterization of the Perceptron laser rangefinder. Mixed pixels are observed at range discontinuities and illustrated with a single example. In [Hebert and Krotkov, 1992, Hebert and Krotkov, 1991] Hebert formalizes the characterization and in addition to the Perceptron laser, uses the Erim laser, another amplitude-modulated continuous-wave laser. Mixed pixels are geometrically interpreted. The author concludes that strong edges are unreliable. In [Ye and Borenstein, 2002] reports mixed pixel observations from a SICK laser similar to the one we use. In [Matthies et al., 2003] Matthies also reports observing mixed pixel for a SICK laser.

In [Adams, 1993, Adams and al., 1996] Adams explicitly models the mixed phenomenon for a scan line CW LADAR and proposes a method to detect discontinuities. In [Adams, 2001], he proposes a feature detection scheme based on a smoothness constraint for data from outdoor scenes. An Extended Kalman Filter is used to predict range measurements.

A large body of work exist in range image segmentation using edge detection. There are differences with our current work. We do not assume that the data will come

in a rasterized fashion, so we work in 3-D, not in the range image space. The scenes in which we are interested do not contain large, smooth areas, but instead a large number of small and thin areas. We are not trying to segment range data at first but instead filter the data for later processing.

3 Mixed Pixel Phenomena

Our experiments have revealed several classes and types of mixed pixel phenomena. Some of these are the result of using LADAR in general, while some are sensor dependent. As such, we will describe first the sensor which recorded these phenomena (3.1), the conditions which generated them (3.2), and finally describe our observations (3.3).

3.1 Sensor Description

The experiments described in this paper used a SICK LMS-291-SO5 industrial Time of Flight (TOF) laser mounted on a pushcart (A.1). This laser was chosen due to its availability and common usage in robotics sensing research.

3.1.1 Physical configuration

The SICK laser is a one-dimensional scanning LADAR (thus a single scan returns information about a two-dimensional plane). In order to obtain a full three-dimensional scene we have mounted two SICKs on a pushcart and independently actuated each with an Animatics SM3420D Smart Motor, stepping at 0.25^o. The SICK itself has several scanning modes; we have chosen to scan 100 points at 1^o separation. A typical scan consists of 800 steps of the motor. One unit is mounted in an upright position at a height of approximately 127 cm and actuated on its x axis, so that as the laser scans horizontally the mounting sweeps vertically. The second is mounted on its side at approximately 138 cm, sweeping around the z axis, so that the scan line moves horizontally. The laser height was chosen to produce a large Field of View (FOV). All of the data presented here was collected with the side-mounted laser. This configuration has a field of view which is 115^o wide an 100^o high, producing an 800×100 range image.

3.1.2 Beam characteristics

The SICK instrument projects a beam which is rectangular in shape, with two regions of high intensity at either end of the beam, as in Figure 1. Such beam shape is consistent with generation by a double stacked laser diode. The image was captured using a WAT-902HS CCD camera[2], equipped with a 12 mm lens, a 916 nm filter and a shutter speed of 1/250. The beam is swept by a rotating mirror, set at 45^o to the original beam. This has the effect of rotating the projection of the beam, so that at the edges the high-intensity regions are perpendicular to the scan line. At the center, these regions are parallel to the scan line. Figure 1 illustrates the projection of the beam at extreme left

[2]This camera was generously provided by the NAVLAB group

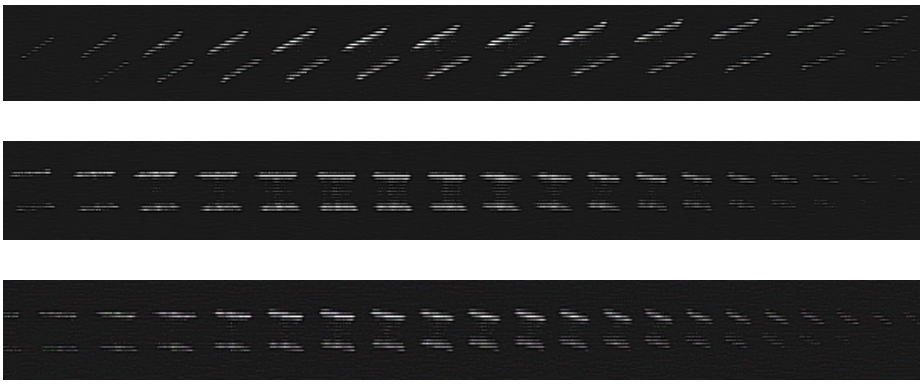


Figure 1: **Portion of SICK scan line.** Captured with an infrared camera, this sequence show the pair of high intensity area produced by the beam and its rotation as the laser sweeps.

(top image) and center (middle image) and right (bottom image). In the configuration we use, the projected beam is 0.68^o wide and scans through a 100^o field of view with 1^o angular resolution.

3.1.3 Data interpretation

The data returned by the laser is converted to rectangular coordinates with (x, y, z) representing the right, forward and upward offsets (respectively), i.e. the sensor is at the origin. All of our analysis and filtering is performed on this data, which only contains the points scanned by the laser in 3-D space. We do not keep the connectivity (row and column identification) of the points. If the data is collected from a moving sensor platform and stored as absolute coordinates, we also keep the location of the sensor with each scan. Thus, our approach may be used with data from other 3-D TOF laser scanners by considering either a single scan or by accumulating data over time as the scanners move.

These points are read into a *proto-volume*, which is a space made up of *proto-points* (i.e. a sparse voxel representation). Points are accessed using a hash key based on the relative position of the points. This structure combines the essential features of several points into one, saving space and reducing the computational complexity of subsequent use of the points. (For a full description of the proto-volume and proto-point system, see [Vandapel et al., 2004]).

3.1.4 Mathematical model

In order to better understand the observed mixed pixel phenomena, we employ a model of the return signal of the laser beam. By considering the portion of the beam reflected back by the target and the portion returned by the background (ground plane or vertical wall), it is possible to explain qualitatively the phenomena observed.

To model the energy waveform we choose a simple reflectance model:

$$E(t) = \sum_{i \in targets} \delta(t - t_i) A_i \frac{\rho_i \cos(\theta_i)}{r_i^2} \quad (1)$$

with: ρ, the reflectance; r, the range; θ, the laser incidence angle; A, the area illuminated; δ, the Dirac function. We normalize the area illuminated to conserver the flux of energy.

The non-zero length of the pulse and the dynamic of the receiver will smooth the energy waveform. We represent both effect as a Gaussian $G(t_i, \sigma_i)$.

$$E(t) = \sum_{i=targets} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(t-t_i)^2}{2\sigma_i}} A_i \frac{\rho_i \cos(\theta_i)}{r_i^2} \quad (2)$$

Several techniques can be used to detect pulses [**?**]: threshold, center of gravity, maximum, zero crossing of second derivative and constant fraction discriminator (CFD). CFD methods are not sensitive to signal amplitude change and are robust to pulse width change; these methods look for the zero crossing of the difference between the attenuated signal waveform and a version of the original waveform delayed by a fraction of the pulse width. We use such an approach.

This model improves our understanding of several phenomena, but it is not a perfect solution. One problem is that the beam intensity is not truly homogeneously distributed. Another is the lack of complete manufacturer's technical data on essential characteristics of the laser such as beam duration and trigger condition, which increases the error of the current model. However, the model is still suitable for explaining most phenomena, which will be further discussed in Section 3.3.

3.2 Experimental Setup

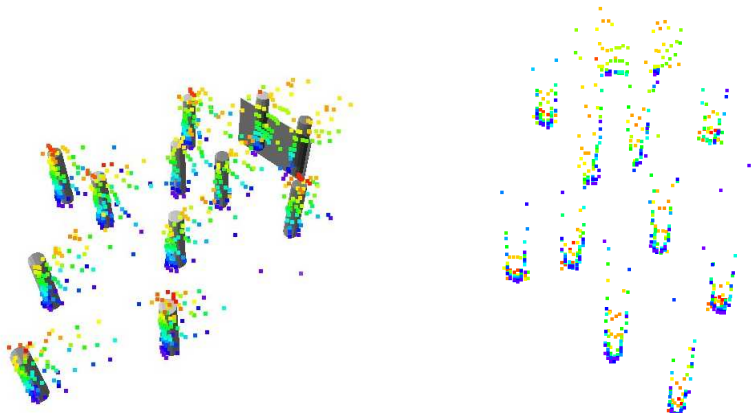
To simulate vegetation-like structures in controlled conditions, small PVC pipes and dowel rods were used in a variety of scenes, both indoors and outdoors. Figures 2, 3, 4 show sample data from each set of experiments. A similar rod set-up was used in [Matthies et al., 2003].

3.2.1 Test series 1: Plywood board and PVC pipes

Plywood board (26 cm tall, 41 cm wide) located behind PVC pipes of diameter 6 cm which were placed randomly between 0 and 150 cm in front of the board. The board itself was located approximately 2 m in front of a solid background, and the laser was situated approximately 6 m directly in front of the board. cans were taken with an increasing number of pipes (from 0 to 12, adding one per scan).

3.2.2 Test series 2: Plywood board located among dowel rods

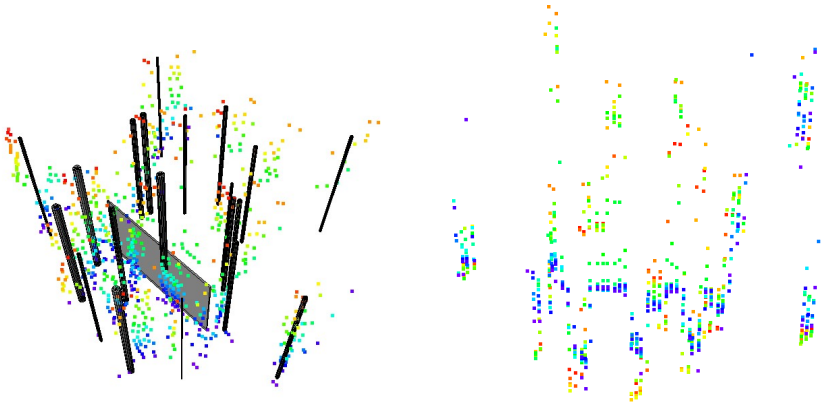
Plywood board located among dowel rods located randomly according to a Poisson distribution around the board (within 61 cm in the x and y directions). The board was placed 1.8 m in front of the background. The poles were 45.7 cm high, and diameters



(a) (b)

Figure 2: **Example data from test series 1**. (a) Side view of the data and model, (b) Top view of data only. Color indicates height (blue is low, red high; scale is relative). Positive directions are: x right, y upward, z out of the page.
.

varied between 3.2 and 22.2 mm (also according to a Poisson distribution). With the laser situated approximately 5.8 m from the board, scans were taken with 0, 8, 16, 24, 32, and 40 poles.



(a) (b)

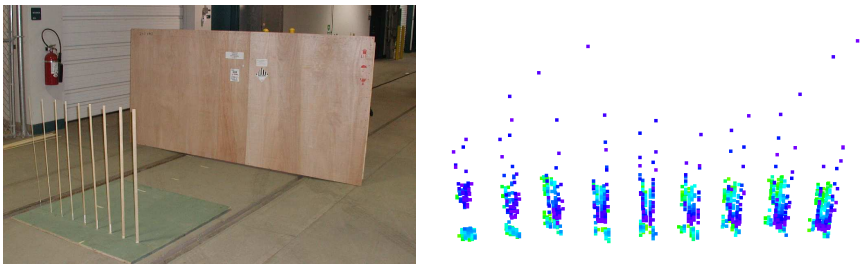
Figure 3: **Example data from test series 2.**(a) Side view of the data and model, (b) Top view of data only. Color indicates height (blue is low, red high; scale is relative). Positive directions are: x right, y upward, z out of the page.

3.2.3 Test series 2: multiple poles with horizontal laser offset

One 92 cm high pole of each of the nine diameters available was set into a plywood base, 12.7 cm apart, approximately 1 m from the background. The laser was positioned directly in front of the target set at 4 and 6 m, and offset approximately 1 m in the positive x-direction, again at 4 and 6 m distant.

3.2.4 Test series 4: poles with varying-distance background

Using a large plywood board as a background, the same nine-pole target as above was set 4 m in front of the laser, with the background moved between 0.33 and 3.38 m distant from the target.



(a) (b)

Figure 4: **Example data from test series 4.**(a) Scene, (b) Top view of data only. Color indicates height (blue is low, red high; scale is relative). Positive directions are: x right, y upward, z out of the page.

3.2.5 Test series 5: thin poles in outdoor setting

Two poles were placed in a landscaped region containing small shrubs, a low stone wall, and trees in the background

3.2.6 Test series 6: systematic variation of target and background ranges

One pole of each of three diameters (3.2, 12.7 and 22.2 mm) was independently scanned in front of a movable background, with distances from laser to target varying across 1.8 to 6.1 m and distances from target to background of 0.3 to 3.1 m.

3.3 Phenomena Observed

The phenomena observed from the data collected can be broken up into several classes: mixed pixels occurring because of overlap of the laser beam onto two objects at the edge of an object; mixed pixels occurring due to target structures with less width than the beam; mixed pixels occurring due to particular characteristics of the SICK laser; and mixed pixels from ground interaction. Figure 6 shows the different phenomena

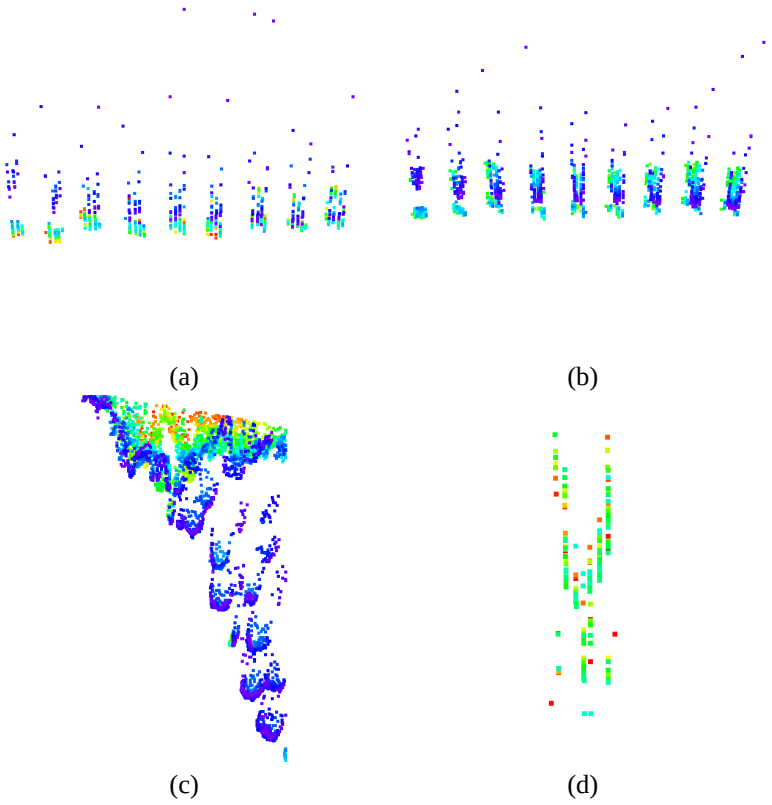


Figure 5: **Example Data.** Top views of: (a) Test series 3 (4 m directly in front of poles). (b) Test series 4 (background 1.85 m from target). (c) Test series 5 (area shown is approximately 2 m wide and 4 m long). (d) Test series 6 (large pole 2.4 m from laser, background at 1.8 m). Color indicates height (blue is low, red high; scale is relative). Positive directions are: x right, y upward, z out of the page.

cited. The LADAR and two rods (the thinnest and the widest) are static. The background is positioned at three different locations. The bottom part of Figure 6 presents a side view of the set-up. The top part of the same figure presents different mixed pixel phenomena.

3.3.1 Edge-occurring mixed pixels

Mixed pixels appearing at the edge of objects are the most common. These occur when a beam is projected across the edge of an object, and (with a TOF laser such as the SICK) will appear between the two objects. The actual position of the return is dependent on the strength of the return from each of the objects and the distance involved. The strength of the return is (as we model it) dependent on distance, reflectivity and

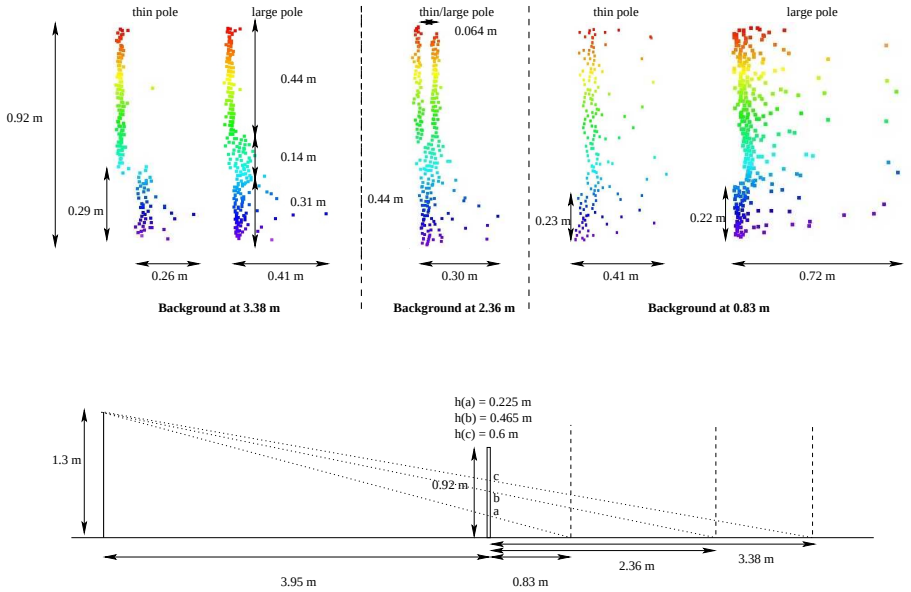


Figure 6: **Mixed pixel phenomena.**

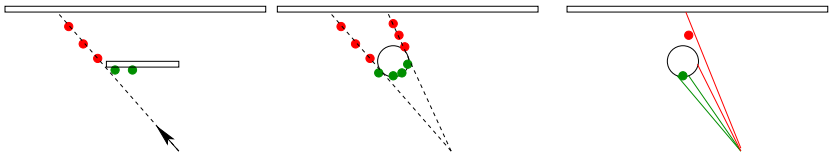


Figure 7: **Mixed pixel taxonomy: edge-occurring.** Beams and the points which they produce are colored red for mixed or green for valid measurement.

orientation. If either the target or background return is sufficiently strong, no mixed measurement occurs.

This type of mixed measurements occurs at the edge of the target's range shadow. For linearly-edged structures in three-dimensional space, these returns exist in a plane forming on the edge of the object and containing the laser's ray of measurement (Figure 7). However, when considering multiple scan lines, the pixels fall in a thin box rather than on a plane due to noise and differing interactions at each beam. For the purpose of clarity of terminology, these will be reference as "V-shaped" mixed pixels, including situations where only one leg of the V is present, such as at the edge of a board.

3.3.2 Thin-structure formed mixed pixels

Thin structures present a special case of the general mixed pixel formation rules. In the case where the object presents less angular width than the width of the beam, the beam is never fully on the object. In these situations, it is possible that no pixels report the

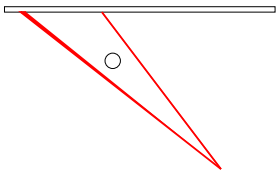


Figure 8: **Mixed pixel taxonomy: thin structure.** A SICK beam (red) overlapping and bracketing a rod.

true position of the target, but instead all points may be considered mixed pixels.

As with most thin structures, the V shape often occurs, but the body of the return may be erroneous, with the edges in the typical V-shape. For example, if the target is a thin rod, the recorded data may contain points which hold the original shape of the rod (including V-shaped edges), but several centimeters farther away than the actual target.

3.3.3 Mixed pixels unique to SICK-type lasers

Due to the nature of the beam projected by the SICK with its regions of high intensity at the edges, it is possible when the beam brackets a thin structure for the high-intensity portions of the beam to fall only on the background object. When this occurs, because of the intensity difference in the beam (and hence the returns) it is possible for the central pixels of a thin target to be incorrectly located as much as 6 - 10 cm from the true location of the object. However, scan lines near the edge of the target remain in the correct location, because one of the two high-intensity regions lands on the target, and one behind. Thus the data, depending on the density of points on the target and size of the target, can appear to have two separate objects where one should exist. Moreover, the second object appears directly in the range shadow of the true target, which is normally an impossibility; it in fact is possible because the center pixels are all at the second location; there are no points obscuring them. This effect, which we have termed a “ghost,” only occurs with thin structures which are perpendicular to the SICK’s scan line. The feature of the ghost that makes it seem more like a real return is that, for all of the scan lines which completely cover the target, the false return behaves similarly to a real return; it has the correct shape and has V-shape structures trailing from behind the center of the ghost return.

A variation on this phenomenon occurs when the target object is nearly as wide as the beam. In this case, the center scan line or lines also produce returns at or near the correct location of the target, so the ghost region becomes bottle-shaped and the translation is less.

3.3.4 Mixed pixels forming from interaction with ground

The mixed pixels formed when a beam is split between an object and the ground (or a similar horizontal surface), while sharing the same cause as the above classes, exhibit a different set of characteristics.

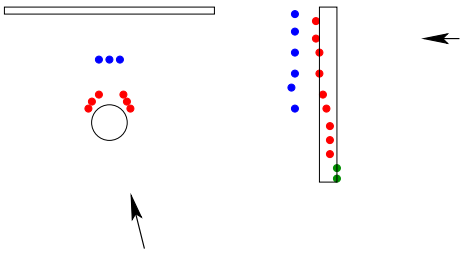


Figure 9: **Mixed pixel nomenclature: SICK specific.** Points corresponding to normal mixed pixels are colored red; ghost pixels are shown in blue. On the left, a top view. On the right a side view.

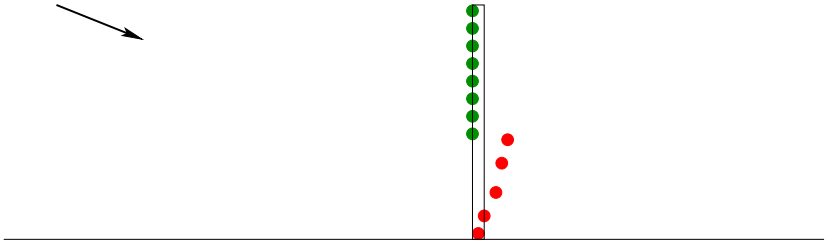


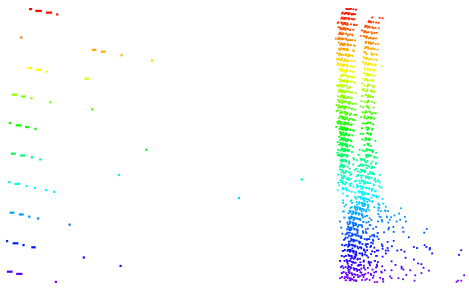
Figure 10: **Mixed pixel nomenclature: ground interaction.** Valid points are shown in green; mixed points in red.

As the sensor scans, the change in distance with each beam is greater for a horizontal structure than a vertical one (since beam is closer to horizontal than vertical in the space under consideration). For instance, if the target is a board, the scan projects beams which first intersect only the floor, then a combination of floor and target, and then only the target. Mixed pixels occur when the beam reflects from both. These measurements describe a slightly curved join between the floor and board. The top edge of the board is also curved, where the beams overlap between the target and the floor beyond.

Thin structures exhibit another characteristic set of mixed pixels when interacting with the ground. The reported structure of the actually-straight poles used in our experiments becomes curved and disjoint when the pixels are formed partly by the ground as in Figure 11(b). This is, of course, caused by the same simple interaction of the laser that causes “normal” mixed pixels, but in this case the effect is different.

3.4 Configurations Producing Mixed Pixels

Mixed pixels do not occur in every scene; only when objects are sufficiently near are mixed pixels present. Our experimentation shows that the foreground and background objects which are hit by the beam must be within 2.5 m for mixed pixels to appear. However, our results do not indicate a lower bound on the occurrence of mixed pixels; on the contrary, when the foreground and background are separated by less than 1 m



(a) (b)

Figure 11: **Side view of mixed pixels.** In both scenes, x comes out of the page, y goes to the right, and z is up. (a) Flat board. Typical V-shaped pixels are visible behind the board. (b) Poles from test series 4. This view clearly shows the ghost effect (the second vertical group of points to the right of the actual poles), V-shaped mixed pixels (the sparse patch of points to the lower right) and the deformation associated with ground interference (the returns begin in the actual location of the poles, but slope backward to where they join with the ghost).

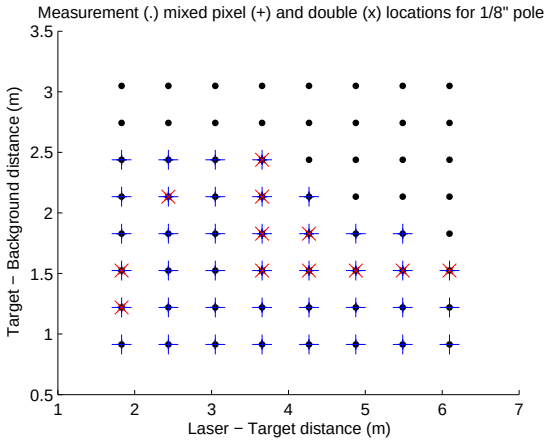
the mixed pixels observed extend completely from the foreground to the background in a solid group, making filtering more difficult. Figure 12 shows where mixed pixels were observed using individual poles of several diameters.

4 Approach to Mixed Pixel Filtering

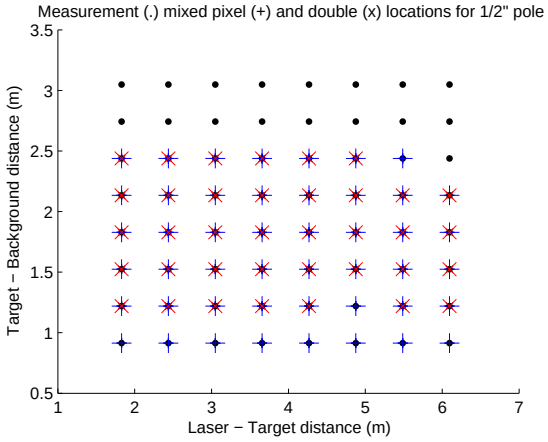
4.1 Philosophy

Because all of the types of mixed pixels observed exhibit the characteristic V in the direction of the laser ray at a point, we filter for mixed pixels only in the region expected to contain mixed pixels. Note that this technique is not capable of displacement of the body of the return (e.g. ghosts, displaced thin structures). These classes of mixed measurements, while erroneous as observed, could be produced by certain objects. Thus we do not wish to automatically remove them from a scene at this time.

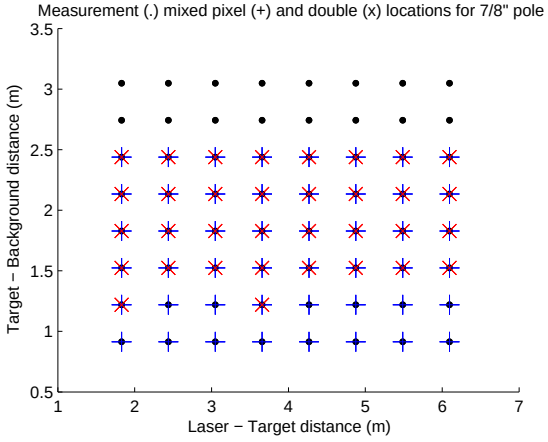
Our algorithm examines each point in the scene in turn. To determine if a point is mixed, we generate a bounding box whose major axis is the ray from the sensor (origin) to the point of interest (Figure 13). The point of interest is at the center of the front face of the box. We compute the saliency features for the points falling in this region. In this regard, we follow the approach of [Vandapel et al., 2004] but with a non-isotropic interest region. The original method looks at the local point distribution in space and uses a Bayes classifier to produce the probability of belonging to three classes – vegetation, solid surface and linear structure. The method takes as input a sparse set of 3-D points. At each point the scatter matrix is computed using a predefined isotropic support region. The decomposition in principal components of this matrix leads to the definition of three saliency features characterizing the 3-D points' spatial distribution as "random", "linear" and "surface". This information is used in two ways



(a)



(b)



(c)

Figure 12: **Appearance of Mixed Pixels.** Dots ($\cdot$) indicate configurations where measurements were taken but no mixed pixels were observed; a cross ($+$) indicates mixed pixels and a star ($*$) denotes ghost-type mixed pixels. (a) 3.2 mm pole. (b) 12.7 mm pole. (c) 22.2 mm pole.

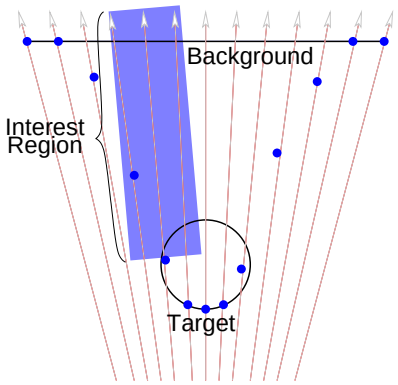


Figure 13: **Filtering of mixed pixels.** Grey lines represent laser beams, blue points are detected points, and the blue rectangle is the interest region used for saliency calculations. Note that this diagram is not to scale and represents a single cross-section of space; incorporating all of the data from each scan line would produce denser points.

to decide whether or not a given point is a mixed measurement. In one variation, the algorithm triggers when the "surfaceness" of the region is higher than the randomness and linearness. The other variation triggers when the normal of the surface described by the region is orthogonal to the laser ray. These methods are described in more detail below.

4.2 Algorithm Variations

4.2.1 Surfaceness triggered

This group of algorithms computes the saliency features for the interest region of a point, and begins marking if the features indicate that the region is a surface. Marking techniques are outlined in Section 4.3 (all marking variations were employed in algorithms with this trigger).

4.2.2 Surface normal triggered

To generalize the classes of mixed pixels which can be filtered (i.e. rely less on the orientation of the interest region), this group of algorithms computes the saliency features but then compares the surface normal vector to the laser ray. If the angle between these vectors is close (according to some pre-determined threshold angle) then the trigger is set and marking occurs. Within this group the naïve, aggressive and Exclusive Blind Window methods of marking were used. This variation was created because we realized that regions where the mixed measurements do not occur in a vertical plane would likely not represent a vertical surface. This trigger mechanism allows for surfaces of all orientations.

4.2.3 Surface-ness and surface normal triggered

Because the surface-ness and surface normal triggered algorithms differ in which types of mixed pixels they are able to filter, one variation examined combined the Aggressive algorithms from both groups (if either triggered, then the interest region was marked as mixed). This algorithm is called Surfaceness/Surface Normal – Aggressive.

4.3 Marking variations

Once the trigger condition is met, one of the following variations is used to mark measurements as mixed (shown in Figure A.2):

- **Naïve** marks only the interest point itself.
- **Aggressive** marks all points in the interest region.
- **Accumulator** counts the number of times each measurement is reported as mixed during the traverse of all points, and then reports measurements as mixed only if some threshold of votes is reached.
- **Exclusive** marks all points in the region except the interest point.
- **Blind window** marks as aggressive, except that it leaves a "blind window" near the original point unmarked.
- **Exclusive blind window** is similar to blind window except that it also leaves the point of interest unmarked.

5 Performance Calculation

The filtering algorithms and their variations described in Section 4.2 exhibited varying degrees of success in detecting mixed pixels and avoiding false detection of good pixels. This section presents the performance data for each variation examined during the course of the project.

It should be noted that the behavior of each algorithm varies significantly with the parameters given to it. The variables which may be altered are:

Cell size of the proto-volume. Since it is the proto-volume that is performing the saliency calculations, this has a significant impact on the precision of the results.

Box width controls the lateral dimension of the box to either side of the measurement ray. Were the ray pointing along the y axis, this would be the full x dimension, centered on the ray.

Box depth measures the maximum distance behind the point of interest that is considered or marked. (This is the size in y of the un-rotated region.)

Box height measures the total height of the interest region (the z size).

Box roll controls the rotation of the box about the laser ray. All other rotations are controlled by the ray itself, which is given by the ray connecting the laser position to the point of interest. Changing the box roll allows the algorithm to detect mixed pixels in non-orthogonal planes.

Blind window depth is an important control on algorithms using this feature; a window which is too small will mark extra good pixels while one which is too large will result in nothing being marked.

Maximum normal deviation applies to algorithms triggered by the surface normals. This is the maximum angle by which a surface's normal may deviate from being orthogonal to the measurement ray.

5.1 Description of Metrics

To evaluate the performance of the various filters, a 25-scene corpus was first manually separated into correct and mixed pixels. The filters were tested against this data with a range of interest region sizes to determine the parameters which gave the best results. After filtering a scene, the number of correctly and incorrectly classified pixels of both types was counted. For the purposes of these metrics, we will refer to classifying a pixel as mixed as a "positive" result (i.e. finding a mixed pixel) and classifying as good a "negative" result (not finding a mixed pixel). Thus the counts of classified pixels are referred to as true positive ($tpos$), false positive ($fpos$), true negative ($tneg$), and false negative ($fneg$). The total number of positive and negative pixels are known as $npos$ and $nneg$ respectively, and the counts for the ground truth are labeled $gpos$ and $gneg$. The total number of pixels classified $N_C = tpos + fpos + tneg + fneg$ and the total number of pixels in the original scene $N_G = gpos + gneg$ also are used as intermediate values in computing other metrics.

The true positive rate (tpr) is given by

$$tpr = \frac{tpos}{tpos + fneg}. \tag{3}$$

This value represents the percentage of actual mixed pixels found. The false positive rate (fpr), is computed by

$$fpr = \frac{fpos}{fpos + tneg} \tag{4}$$

and shows the portion of good pixels which are incorrectly marked as mixed. Accuracy (acc), calculated with

$$acc = \frac{tpos + tneg}{N_C} \tag{5}$$

is the percentage of all correct classifications. Because computing saliency requires at least three points, some interest regions may not contain enough support for the computations. In these cases, the points are discarded and are not reported as either good or mixed, unless part of a region being aggressively marked. The loss rate ($loss$),

$$loss = \frac{N_C}{N_G}, \tag{6}$$

Table 1: **Algorithm performance.** These data represent the best performance each algorithm exhibited during testing.

Algorithm	tpr	fpr	acc	Score
Surfaceness triggered				
Naïve	0.222	0.268	0.713	0.417
Aggressive	0.803	0.643	0.729	0.513
Accumulator	0.516	0.389	0.701	0.466
Blind Window	0.764	0.422	0.750	0.548
Exclusive	0.792	0.609	0.725	0.505
Exclusive Blind Window	0.668	**0.147**	0.812	**0.657**
Surface Normal triggered				
Naïve	0.188	0.217	0.722	0.410
Aggressive	0.811	0.629	0.764	0.573
Exclusive Blind Window	0.686	0.153	**0.814**	0.640
Combination triggered				
Aggressive	**0.936**	0.736	0.744	0.596

is the measure of how many pixels in a scene are not reported. Generally, this is only a factor with Naïve algorithms, since very few are discarded by marking all of the points in the interest regions. Finally, in an effort to condense the analysis, a scoring system was developed by which correctly classified pixels gain a positive score while false reports lose points, as do lost pixels. The number of points per pixel is divided by the maximum possible score for the scene (which results from perfect classification) to give an idea of relative performance. By losing points when making mistakes, an algorithm's score will always be lower than its accuracy (unless both are 1). In the data shown, 5 points were added to a score for a correct classification, but 3 were removed for a false negative, 4 for a false positive, and 1 for a loss.

5.2 Effectiveness

Table 1 shows the best average (over the 25 scenes used in testing) performance for each algorithm. Note that each score in the table does not always come from the same set of parameters to the algorithm, with the exception of tpr, which is presented with its corresponding fpr (thus these are not necessarily the best false positive rates obtainable by each algorithm). For more detailed information about the overall performance of each algorithm, refer to Appendix C. Figure A.3 shows several scenes filtered by Surfaceness – Exclusive Blind Window.

5.3 Efficiency

Written in C++, our proof-of-concept implementation is able to compute localized saliency features for approximately 6000 voxels per second. Naïvely marking when triggering on surfaceness is done in quadratic time and, with our hardware[3] and test

[3] A dual Intel Xeon (2.8 GHz) machine with 3 GB of memory and Red Hat Linux version 9. Note: our software does not make use of both processors.

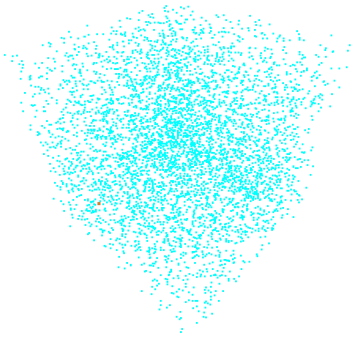


Figure 14: **Filtered dense cloud.** With 15,000 points in a $10 \times 10 \times 10$ m space, surfaceness triggering reports no mixed pixels. The dark dot is the origin.

data (averaging roughly 1000 points per scene, with ground removed *a priori*), can be computed for approximately 1.3 million points per second; aggressive marking can be performed in cubic time on 7500 points per second. When triggering on the surface normals, naïve marking runs on approximately 1 million points per second; running aggressively processes 7700 points per second. For comparison, the GDRS laser on the XUV records on the order of 100,000 points per second.

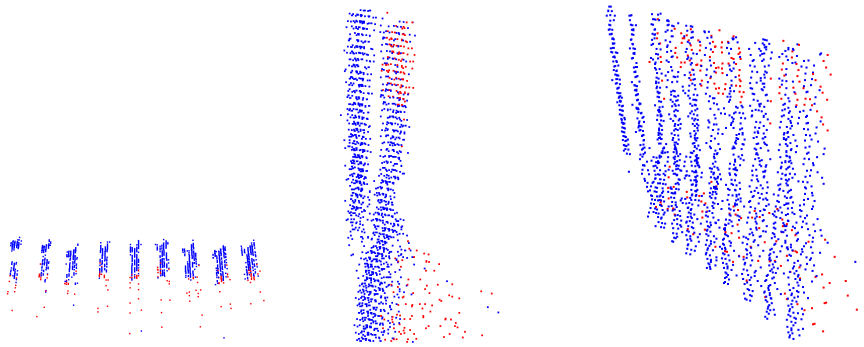
5.4 Performance against Point Clouds

One concern with the approach which we present here is the tendency to always find planar surfaces when considering a near-planar interest region. To demonstrate that this is not a problem exhibited in practice, we test the filter against a space of uniformly distributed random points (Figure 14). (Since vegetation produces very noisy returns, it is important that the filter not mark all point clouds as mixed pixels.) Triggering on surfaceness, no mixed pixels were reported by the Naïve algorithm (which would show only the points for which the trigger condition is present) with the most successful set of parameters for five spaces of 1000 m^3 with 300, 1,500, 2,000, 6,000 or 15,000 points in each region. Testing on the same data sets, surface normal-triggered Naïve reported 9 mixed pixels only on the most dense cloud. (Using aggressive marking would report more mixed pixels but would be a less useful metric since the marking of each pixel is not independent of other pixels.)

5.5 Performance Overview

As can be seen in Table 1, the optimal algorithm examined thus far is the surface normal-triggered Exclusive Blind Window. With an average accuracy of 81% over the scenes tested, with a peak accuracy of 94% its only close competitor is the surfaceness-triggered Exclusive Blind Window, whose advantage is a lower false positive rate (and hence more pixels left in the scene).

With the point density given by our laser, this method is fast enough for real-time scenarios. However, given a peak 80% accuracy rate, with 15% of correct measurements misclassified most of which are at object edges and are unlikely to affect the process of finding objects, though they may affect estimates of size), the process is capable of removing mixed pixels from a scene. We can also say with reasonable certainty that the algorithm behaves as desired and only reports mixed pixels when planar structures exist, and is not so over-sensitive that it detects them in data which cannot be known to contain mixed pixels.



(a) (b) (c)

Figure 15: **Sample filtered data.** Correct measurements denoted in blue, mixed in red. Data taken from test series 4 and filtered with the surfaceness trigger and Exclusive Blind Window marking. (a) Top, (b) side, and (c) perspective views.

6 Conclusions

It is possible to quickly and accurately enhance scene quality by removing mixed measurements. As seen in Figure 15, the filtering techniques presented here are able to remove most of the mixed measurements from a scene involving thin linear structures. Using existing techniques for discovery and removal of ground and vegetation, three-dimensional laser imagery becomes a useful and effective tool for object recognition and feature detection.

Efficiency is as yet an issue. The algorithm is not fast enough to be run on entire scenes, so at this time we first remove the ground plane and crop the scene to our region of interest. When the entire operation of cropping the scene and filtering mixed measurements is performed, a single frame may require one to ten seconds to process. Thus its use in real-time applications is questionable.

As the project progresses, we will continue to improve algorithm efficiency and accuracy. One optimization is to filter by voxels instead of by the raw points themselves. We continue to investigate more effective filtering algorithms and marking methods and the user of other sensors, which would serve to both validate our methods and to allow

us to increase the generality and sensor-independence of our algorithms. Appendix **??** contains a more detailed list of goals for the continuation of the project.

References

[Adams, 1993] Adams, M. (1993). Amplitude modulated optical range data analysis in mobile robotics. In *IEEE International Conference on Robotics and Automation*.

[Adams, 2001] Adams, M. (2001). On-line gradient based surface discontinuity detection for outdoor scanning range sensors. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*.

[Adams and al., 1996] Adams, M. and al. (1996). The interpretation of phase and intensity data from a.m.c.w. light detection sensors for reliable ranging. *International Journal of Robotics Research*, 15(5).

[Ansar et al., 2004] Ansar, A., Huertas, A., Matthies, L., and Goldberg, S. (2004). Enhancement of stereo at range discontinuities. In *CTA conference*.

[Ansar and Matthies, 2003] Ansar, A. and Matthies, L. (2003). Improvement of stereao in complex scenes. In *Collaborative Technology Alliance Conference*.

[Hebert and Krotkov, 1991] Hebert, M. and Krotkov, E. (1991). 3-d measurements from imaging laser radars: how good are they? In *Proceedings of the IEEE/RSJ International Workshop on Intelligent Robots and Systems '91, Intelligence for Mechanical Systems*, volume 1, pages 359 – 364.

[Hebert and Krotkov, 1992] Hebert, M. and Krotkov, E. (1992). 3d measurements from imaging laser radars : how good are they ? *International Journal of Image and Vision Computing*, 10(3):170–178.

[Hebert and Vandapel, 2003] Hebert, M. and Vandapel, N. (2003). Terrain classification techniques from ladar data for autonomous navigation. In *Collaborative Technology Alliances Conference*.

[Kweon et al., 1991] Kweon, I. S., Hoffman, R., and Krotkov, E. (1991). Experimental characterization of the perceptron laser rangefinder. Technical Report CMU-RI-TR-91-01, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA.

[Ludwig et al., 2003] Ludwig, D., Kongable, A., Krywick, S., Albrecht, H. T., Kamrath, G., Milam, J., Brown, D., Fetzer, G. J., and Hanna, K. (2003). Identifying targets under trees: jigsaw 3d ladar test results. In *Laser Radar Technology and Applications VIII*, volume 5086.

[Matthies et al., 2003] Matthies, L., Bergh, C., Castano, A., Macedo, J., and Manduchi, R. (2003). Obstacle detection in foliage with ladar and radar. In *International Symposium on Robotic Research*.

[Murray et al., 2003] Murray, J., Rosasco, R., and Holmes, M. (2003). Active acoustic sensing of stumps in tall grass. In *Collaborative Technology Alliance Conference*.

[Schilling et al., 2002] Schilling, B., Barr, D., Templeton, G., Mizerka, J., and Trussell, C. (2002). Multiple-return laser radar for three-dimensional imaging through obscurations. *Applied Optics*, 41(15).

[Vandapel et al., 2003] Vandapel, N., Donamukkala, R., and Hebert, M. (2003). Experimental results in using aerial ladar data for mobile robot navigation. In *International Conference on Field and Service Robotics*.

[Vandapel et al., 2004] Vandapel, N., Huber, D. F., Kapuria, A., and Hebert, M. (2004). Natural terrain classification using 3-d ladar data. In *IEEE International Conference on Robotics and Automation*.

[Ye and Borenstein, 2002] Ye, C. and Borenstein, J. (2002). Characterization of a 2d laser scanner for mobile robot obstacle negotiation. In *IEEE International Conference on Robotics and Automation*.

A Additional Tables and Figures



Figure A.1: **Sensor cart.** The unit used to collect the data presented in this paper.

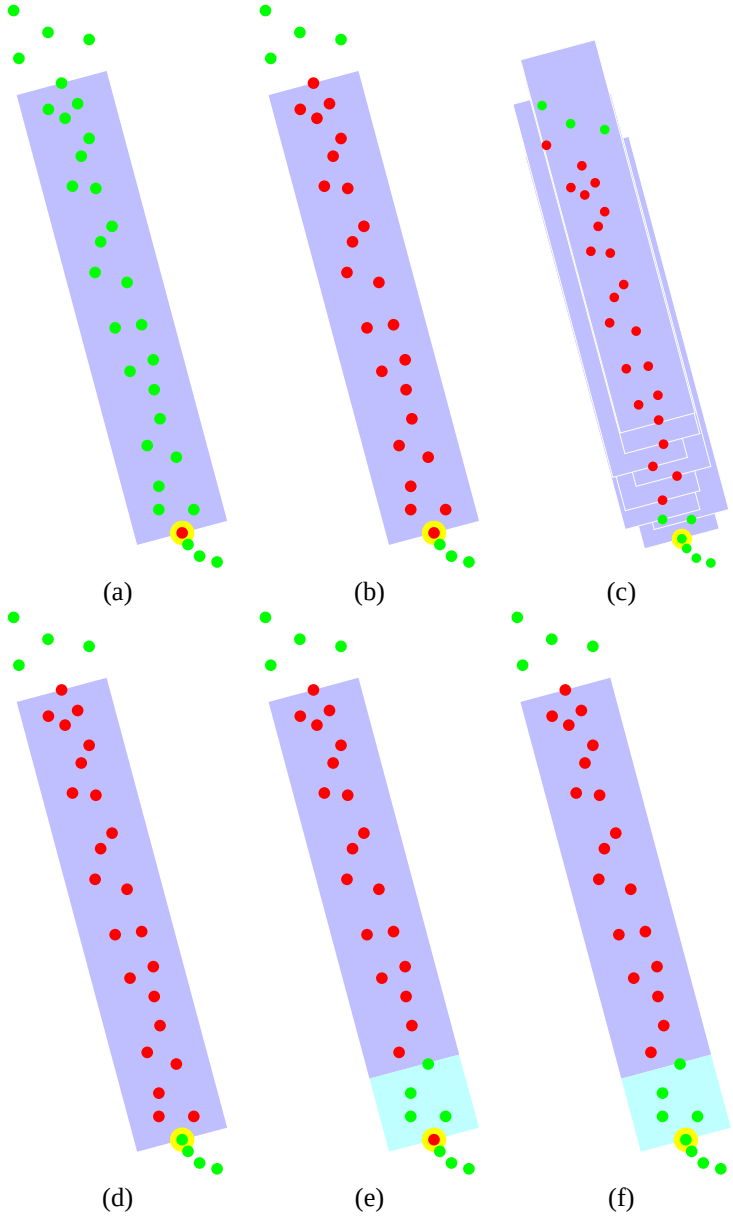


Figure A.2: **Filtering variations.** In each figure, the blue area is the interest region. Green points are left as good points; red are marked as mixed. The point labeled in yellow is the interest point. (a) Naïve. (b) Aggressive. (c) Accumulator (requiring four votes). (d) Exclusive. (e) Blind Window. (f) Exclusive Blind Window.

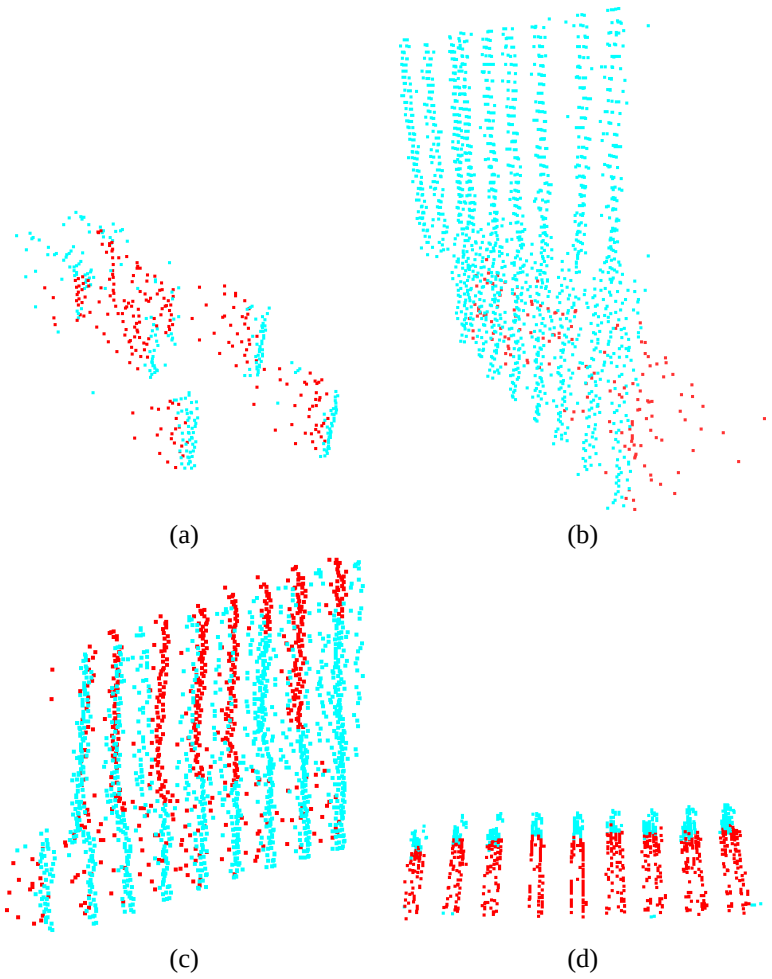


Figure A.3: **Sample filtered scenes.** In each case, mixed pixels are colored red and good pixels blue. Results obtained with the Surfaceness – Exclusive Blind Window filter. (a) Pipes and board from test series 1. (b) Poles from series 4 with background at 2.87 m. (Exhibits only V-shape mixed pixels and ground effect). (c) Poles from series 4 (background 1.85 m) with mirage. (d) Poles from series 4 with background at 0.33 m.