ALGORITHM TO COUT MODERN HOUSES FROM LiDAR DATA SETS OVER RURAL AREAS IN MPUMALANGA

S. Josias *, R. Bekker† and H. Laurie‡

Industry Representative

B. Erasmus[1]

Study group participants

W. Blomerus, P. Dumani, S. Krishnannair, Z. Mkhize, E. Netshiunda, R. Randrianomentsoa and S. Tsiorintsoa

Abstract

The problem is to automate the processing of a large number of aerial surveys to track what appears to be a strong increase in the number of modern houses in rural areas where they were previously rare. Here, "modern house" means a house with several planes in its roof (as opposed to thatched rondavels and flat-roofed shacks).

The data are generated by LiDAR captured at a fairly low altitude. Each data file contains heights (in metres) above ground in a square of 600×600 metres (the actual number of data points vary, typically of the order of 10^5 to 10^6. These are the points at which the laser of the LiDAR apparatus reflected back to the aeroplane and are interpreted to be the height of a solid object, typically a building or a plant (a shrub or a tree). We investigated a simple way of clustering LiDAR data points so that each cluster contains a particular object, be it the roof of a house, or a tree. The clustering is done in the horizontal plane by determining the convex hull of points that are closely clustered. From these it is easy to obtain a rectangular bounding box for the convex hyll. The bounding box is used as a mask to extract LiDAR data for a candidate house. We hope that the Hough transform will suffice to get a reliable measure of the number of planes in the candidate house, thus allowing machine classification of all the houses in the image.

*Department of Mathematical Sciences, University of Stellenbosch, Stellenbosch, South Africa. *email: shane.josias@gamil.com*

†Department of Mathematics and Applied Mathematics, University of Pretoria, Hartfield 0028, Pretoria, South Africa *email: rebeccaabekker@gamil.com*

‡Department of Mathematics and Applied Mathematics, University Cape Town, Cape Town, Rondebosch 7701, South Africa *email:henri.laurie@uct.ac.za*

[1]Global Change Institute, University of the Witwatersrand. *email: barend.erasmus@wits.ac.za*

1 Problem statement and available data

LiDAR measurements provide a very accurate distance between a plane or drone and the light-reflecting surface below. We were presented with processed files of x, y, z data (all in metres). These x, y, z data points refer to points that lie on the surface of objects in some terrain. Only non-zero z-values were reported and they are interpreted as light-reflecting surfaces at the height z above ground level. In other words, ground level is assumed to correspond to $z = 0$ and need not be reported. The x and y data are displacements from a reference point or origin.

These data are of interest because they may furnish a rapid and cost-effective method of keeping track of socio-economic development in rural areas. Specifically, the Global Change Institute are following approximately 1000 households in an area where previously almost all houses were one-roomed structures, such as rondavels and shacks. The assumption is that there is a direct correlation between the socio-economic status of an area and the type of houses that exist in that area. In recent times, more and more households appear to be building bigger houses. The rate of this development could be quantified if a reliable count of these houses were made from time to time. In addition to the above-mentioned quantification, certain strategies may be better formulated to aid socio-economic development.

A multi-roomed house is likely to have a fairly complicated pitched roof comprising several planes. Thus the problem of counting the houses reduces to the problem of identifying houses in the LiDAR data and classifying each house according to the number of planes in its roof. An algorithm for doing so efficiently is what we want.

2 Proposed workflow and algorithms

There are two aspects to an automated solution: the workflow and the algorithms. By workflow we mean the various stages going from input data to a final count.

Here we say very little about the workflow. It will have to be a version of the following (automated, of course):

- pre-process the data files to remove spurious data and make sure all data are in the correct format
- for each file, do filtration, segmentation, determine bounding boxes, and count the number of planes per bounding box
- for each file, identify possible edge effects, that is, houses only part of which cross the edge of the 600x600 tile.
- resolve edge effects, that is, houses that lie across a shared boundary of the 600x600 tile.
- add up all the counts in each file, add houses identified as edge effects

Possibly this could be entirely automated. If not, the number of cases that need to be done by hand may or may not render the overall project infeasible. We note that all image processing in this project was done using a `python` image processing library, `scikit-image` [1].

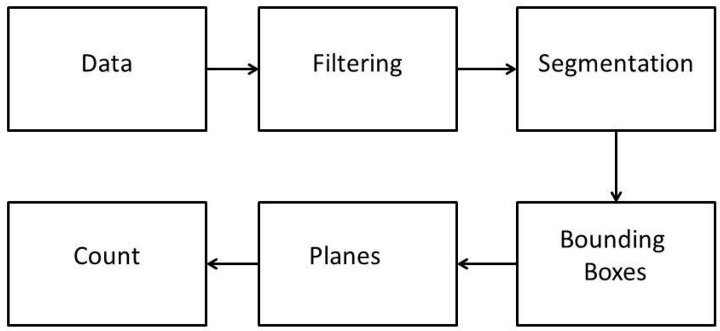

Figure 1: The stages of solving the problem. In this report, we report mostly on the filtering, segmentation and bounding boxes aspects of the process.

2.1 Pre-processing the images to obtain useful height data

This has two stages:

1. cleaning the data, converting the z values to height above the surface, and storing the result as a sparse set of (x, y, z) coordinates of points such that all z-values are nonzero and large erronous z-values are removed, as well as

2. filtering the data so that only points with z-values in the likely range of roof-heights are retained.

For our work during study group meetings, we used files that had already gone through the first of these stages. For the second stage, we used heights between 2 and 5 metres (that is, with $2.25 \leq z \leq 6$). This is not necessarily the very best interval that could be used, but it is likely that an ideal interval does not exist.

2.2 Segmentation algorithm

We used the `heatmap` function of PyPlot to convert the z-values to colour values, and then used a colour segmentation algorithm (RGB thresholding) to determine contiguous blocks of closely related values.

This procedure returned many false positives (but also at least some false negatives, on which more later).

In order to remove the false positives, the clusters were converted to mono-chrome values and subject to morphological filtering. This got rid of a number of small clusters that could not be houses and amalgamated others into a single cluster that could be a house.

Finally, the clusters that remained after morphological filtering were enclosed in rectangular bounding boxes. Each bounding box is then to be separately treated to determine the number of distinct planes in it.

2.2.1 Morphological filtering

Morphological filtering can be described as a set of nonlinear operations that is concerned with shapes or structures. We can think of trees, houses and shrubbery having different structures. This is true for both the LiDAR and aerial photography data sets. Mathematical morphology thus provides an almost natural approach to discerning between the different shapes. The morphological operations are as follows:

1. Erosion: sets pixel at position (i, j) to a minimum over all pixels within some neighbourhood.
2. Dilation: sets pixel at position (i, j) to a maximum over all pixels within some neighbourhood.
3. Opening: erosion followed by dilation.
4. Closing: dilation followed by erosion.
5. Convex hull: smallest convex polygon surrounding a set of white pixels.

Now, erosion reduces the bright regions, and also alters larger structures. The neighbourhood mentioned above may be of arbitrary polygonal shape and size, which affects how white pixels are removed. A larger neighbourhood would remove more white pixels. In the case of the binary images achieved from the heatmap, we use a 3x3 neighbourhood. Once this is chosen, the remaining processes are fully automatic. Conversely, dilation enlarges bright regions.

Erosion and dilation, applied in succession can remove small bright spots and attempt to restore the larger objects to their original form. This process, described above, is called opening. During opening over-dilation occurs. To compensate for this morphological closing is applied. If no over-dilation occurs the larger, brighter regions (i.e. a house) cannot be restored to their original form.

The resulting binary image may still contain streaks, which can be removed by forming a convex hull around objects. This yields objects filled with white pixels, for which we find bounding boxes. To this end we make use of *connected components*.

2.2.2 Bounding boxes: connected components

Once these regions have been obtained, they are labelled by considering pixels that are connected. Two pixels are connected if they are neighbours and share the same value. The neighbourhood used for the binary images is the so called *Moore neighbourhood*. The

Moore neighbourhood considers the eight (8) immediate surrounding pixels of a centre pixel. Once the regions are labelled, we extract the bounding box coordinates (in pixels) and convert these into real-world coordinates, in order to isolate clusters of LiDAR data points. These clusters are then examined to determine whether or not they represent a modern house.

2.3 Finding planes via the 3-D Hough transform

By our definition modern houses have complicated pitched roofs, which can be thought of as planes, while informal/rural houses are comprised of a single plane. One method of categorising houses is then by fitting planes to the LiDAR points of the roots, and if the number of planes fitted is more than one the object under consideration is a modern house. Naturally, a single plane corresponds to a flat roof. To this end we apply the Hough transform, a method of fitting models to data or images. An issue that might arise from this method is when the object under consideration is vegetation. Due to the randomness of heights within trees and shrubs it is possible that a large number of planes can be fitted to this cluster. To overcome this we consider the strength of fit of a plane: the higher the number of cluster points that can be fitted to a plane, the stronger the fit. Planes fitted to tree clusters should therefore be weaker than those fitted to clusters corresponding to houses.

The Hough transform works by considering a three dimensional (θ, ϕ, ρ) accumulator matrix M and the representation of a plane in spherical coordinates

$$\rho = x \cos\theta \sin\phi + y \sin\theta \sin\phi + z \cos\phi. \tag{1}$$

Then for each (x, y, z) point in the cluster previously obtained we iterate through all discrete (θ_i, ϕ_i) pairs and use equation (1) to determine a corresponding value for ρ_i. The (θ, ϕ, ρ) triple is used as an index in M, and the entry at that index is incremented. Since the triple represents a plane, we need only to search the accumulator matrix for the largest values. The index of the maximum value then provides the strongest planes in the cluster of points.

In order for the Hough transform to be applicable we assume that all LiDAR points that lie on the surface of a roof are coplanar. This is not necessarily always the case, as some roofs are slightly concave.

3 Preliminary results

Here, we discuss the processes an image is subjected to. Figure 2 shows a compound image, consisting of an aerial photograph and the corresponding LiDAR data. It should be noted that this is the bottom right corner of a full compound image.

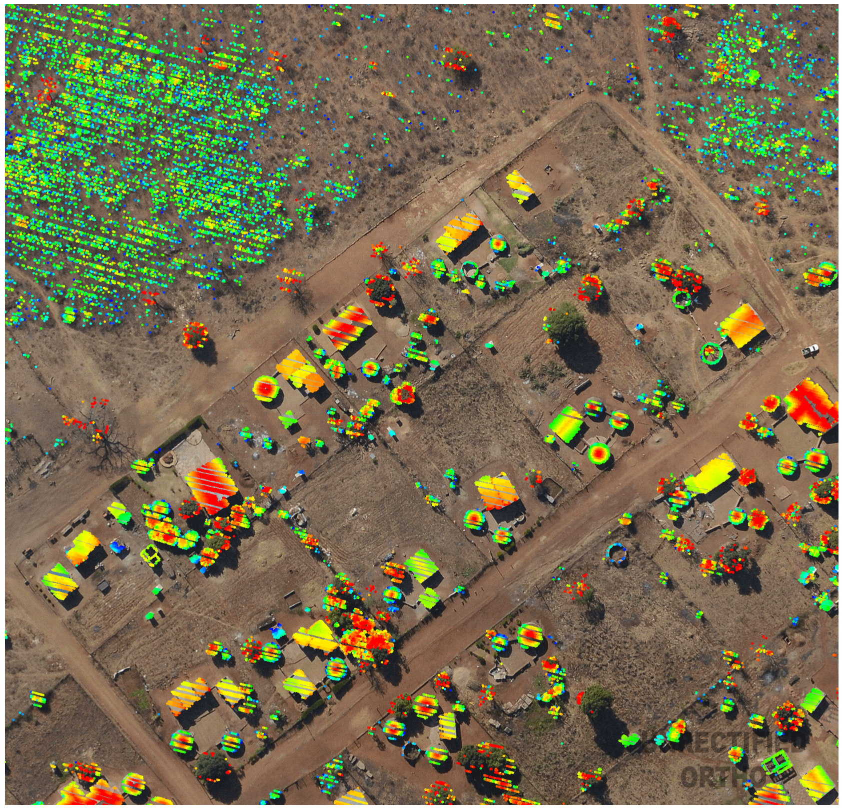

Figure 2: A typical set of LiDAR data, overlain on an aerial photograph. Only heights between 2.25 and 6 metres are shown, with colour indicating height (blue is 2.25 m, red is 6 m). It is clear that the filtered LiDAR data on the whole correspond well to houses. However, some vegetation is retained, and unfortunately there are defects appearing as diagonal stripes (which are artifacts of the data) through some of the houses.

3.1 Segmentation and bounding boxes

After the initial segmentation, conversion to binary data and clustering, we obtain Figure(3). The white shapes are supposed to indicate possible houses. Notice that many dots that clearly are not part of any house remain; they are presumably trees. Also, many of the rectangular shapes that clearly are houses have several diagonal stripes of missing data.

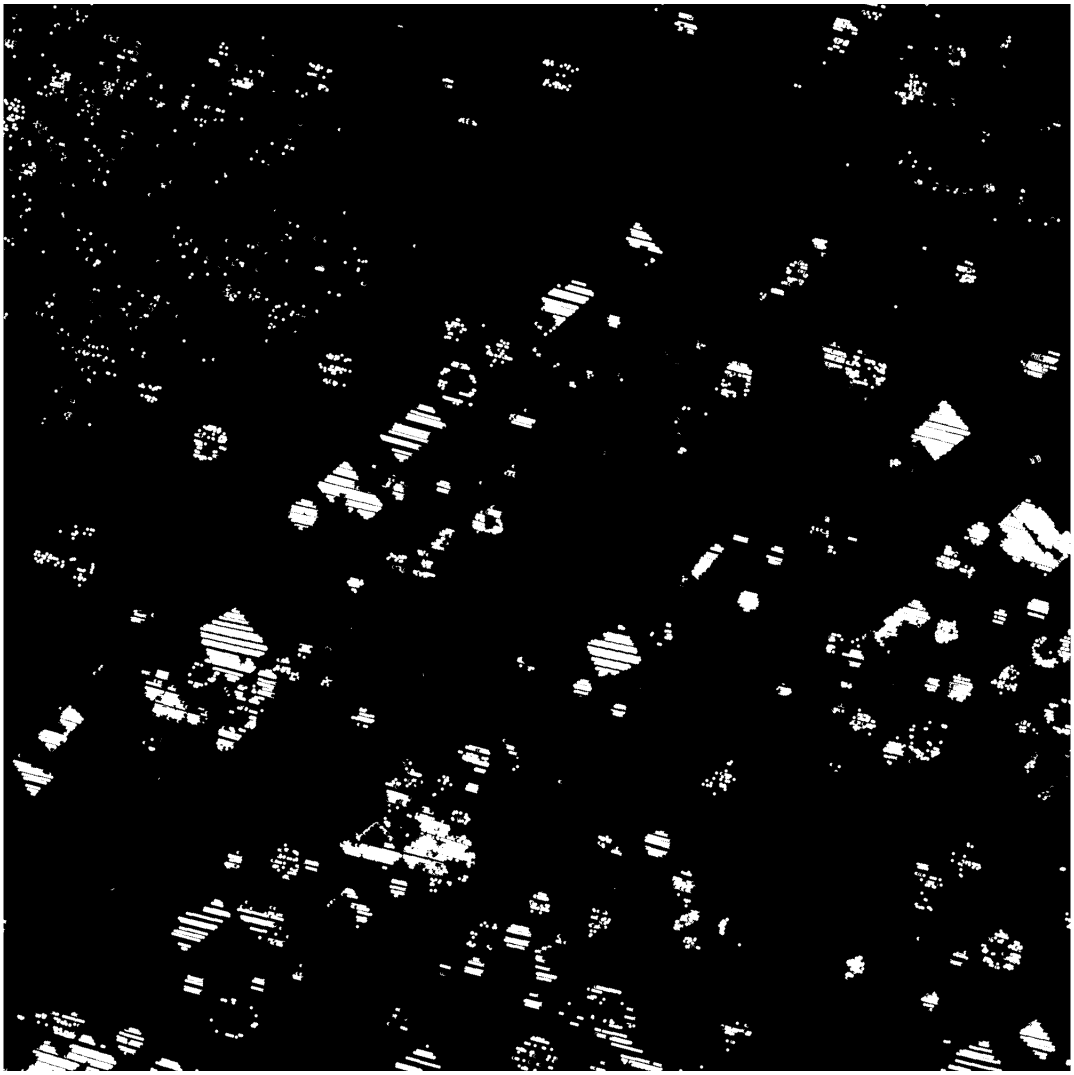

Figure 3: The image after segmentation, clustering and conversion to binary data.

We apply the morphological filter to remove as far as possible the dots that correspond to trees, and to consolidate the rectilinear shapes that seem to come from houses. We obtain Figure (4)

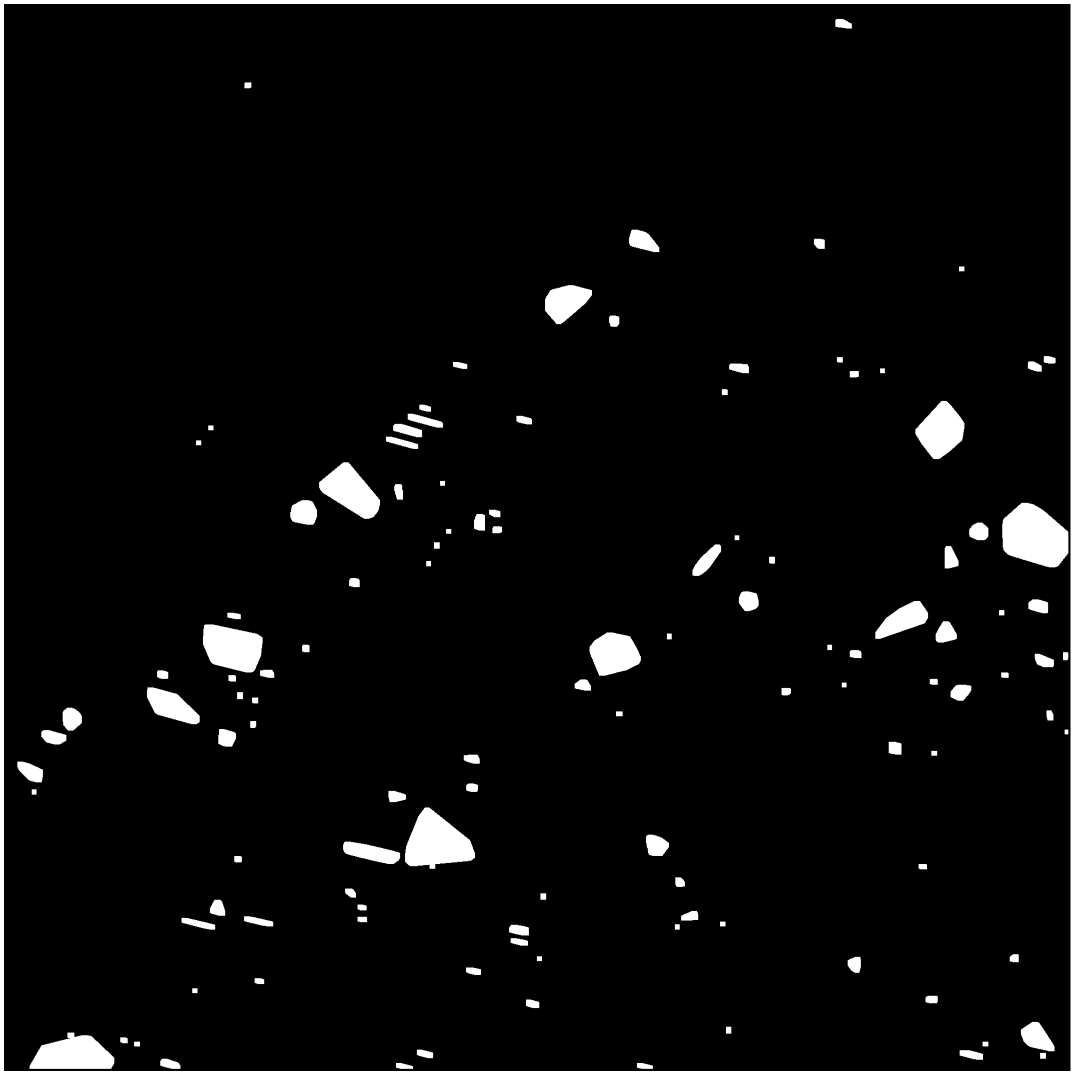

Figure 4: After morphological filtering, the majority of trees are gone as are the majority of the diagonal defects in houses. However, some remain.

Finally, the remaining white shapes are consolidated by finding the convex hull, and a rectangular bounding box, with sides parallel to the boundaries of the image, is drawn around each cluster that is found. Figure 5 is obtained. Sadly, this leads to some false negatives. We would hope that the houses that are included in the false negatives are small, informal houses but this needs to be tested.

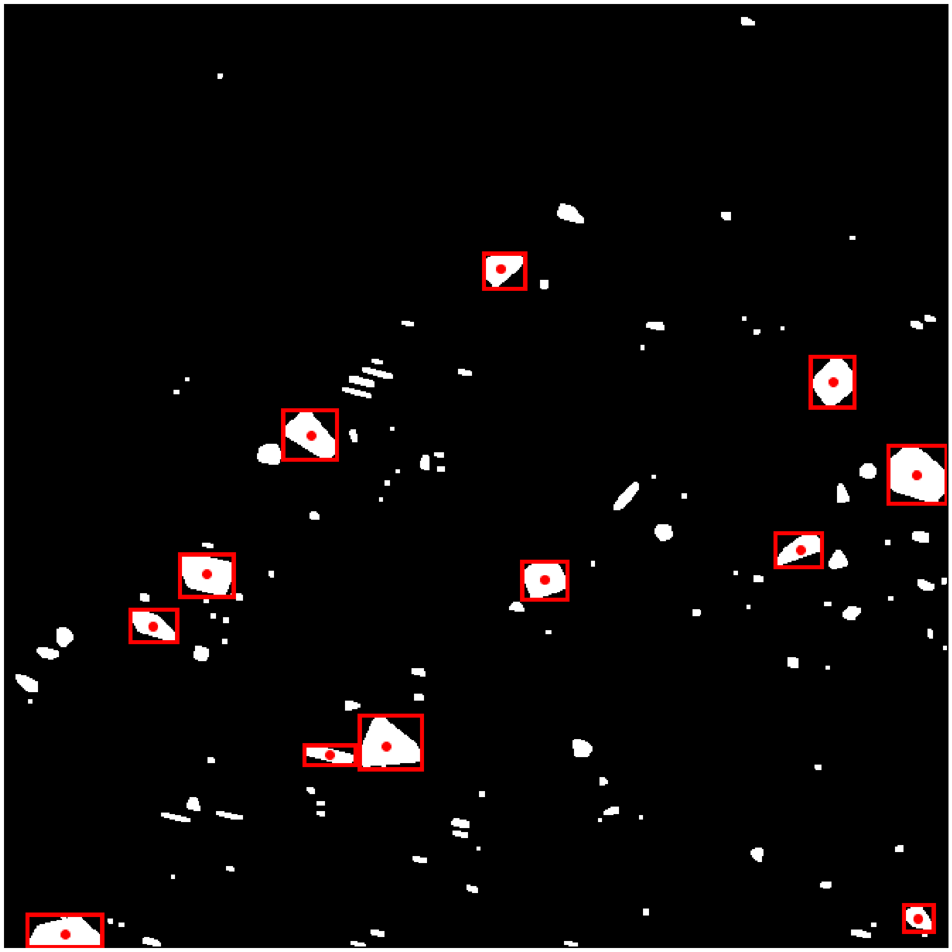

Figure 5: Bounding boxes around candidate houses. Notice that one shape that is likely to be trees is included, while a likely house is excluded because the diagonal defects were not repaired by the morphological filter.

3.2 LiDAR data from within a bounding box

Figure 6 shows examples of how LiDAR points look within the bounding box. Note the regularity of the points and the distinct planar shapes in the first two images. Clearly the last cluster does not correspond to a man made structure due to the randomness of the points; thus this cluster represents vegetation.

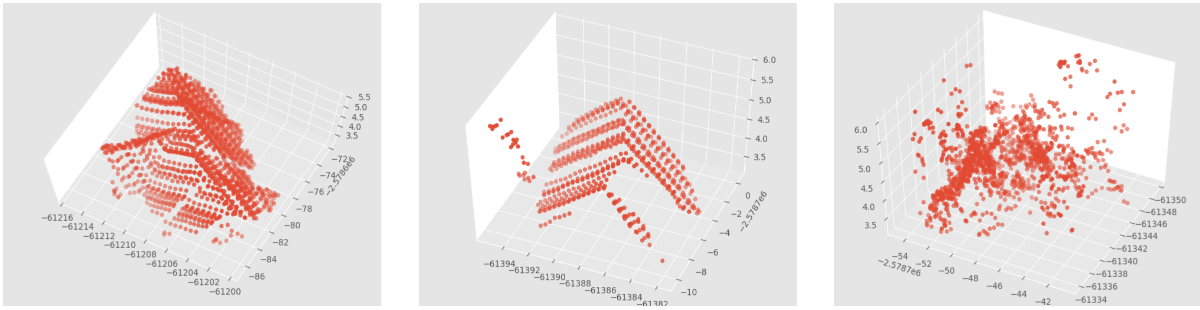

Figure 6: Here we have clustered points that are close in euclidean distance. The first two images are that of a house, whereas the third image represents a cluster of tree points.

3.3 Identifying the number of planes in a bounding box

Unfortunately there was not enough time to complete the suggested workflow, and this is left for future work.

4 Future work

4.1 Fine-tuning: false positives and false negatives

We have seen that some non-houses were included in the bounding boxes. However, there is good reason to think that these will not result in multiple planes being detected, so the false positive rate is likely to be low. This conjecture is based on the assumption that the Hough transform is successful in identifying the number of planar structures in a set of 3D points, which remains to be confirmed.

However, based on this extremely limited test, it is likely that the approach as suggested will miss a number of modern houses. This is likely due to the failure to remove diagonal defects and the possibility of houses occurring on the boundaries of two images. This needs further work.

4.2 Finding planes: geometric multi-model fitting

In addition to implementing the 3-D Hough transform, an additional multi-model fitting algorithm should be investigated. The work done in [2] suggests an approach based on minimising a global energy function.

4.3 A software system for processing the images

Even if all the stages perform as we indicated, there is a significant amount of coding and testing that would need to be done to create an efficient software system. This represents a fairly large investment.

4.4 Principal components analysis

Principal components analysis (PCA) of the image and/or x-y-z points in a single bounding box can be considered. Once we build up a database of a cluster of house points and clusters of tree points, we can use PCA on an unseen cluster of points to attempt to classify it. This would take the form of a nearest neighbour classifier where we have a database of house and tree clusters that the unseen cluster would be compared to.

4.5 Deep learning as an alternative

A convolutional neural network might be able to learn the signature of a modern house directly, from a fairly large set of LiDAR training data (say, about 10 000 houses of all types in about 1000 images). If this works, it could lead to an extremely efficient counting algorithm. Of course, the clustering steps are still required as the classification will be done on the segmented clusters.

Alternatively, we could apply convolutional neural networks on the aerial photographs themselves. First, we need examples (images) of roofs of modern houses as positive samples and informal houses and trees as negative samples. It could be a binary classification task, or we may consider to have three class labels. Then once a neural network is trained, we can move a sliding window across the image and attempt to classify what is contained in the sliding window. The regions of the image likely to contain a modern house can then be investigated and a count can be made.

References

[1] van der Walt S., Schönberger J.L., Nunez-Iglesias J., Boulogne F, Warner J.D., Yager N. et al. scikit-image: image processing in Python. Peer J, 2014 6;2e453. Available from: `http://dx.doi.oorg/10.7717/peerj.453`.

[2] Isack H. and Boykoc Y. Energy-based geometric multi-model fitting. International Journal of Computer Vision, 2012, **97**, 123-147.