

# **AUTO HYDRO BREAK LINE GENERATION USING LIDAR ELEVATION AND INTENSITY DATA**

**George J. Toscano\***,

**Uday Kiran Gopalam,**

**Venkat Devarajan,**

Dept. of Electrical Engineering,

University of Texas at Arlington, Texas, USA,

[george.toscano@mavs.uta.edu](mailto:george.toscano@mavs.uta.edu),

[udaykiran.gopalam@mavs.uta.edu](mailto:udaykiran.gopalam@mavs.uta.edu),

[venkat@uta.edu](mailto:venkat@uta.edu)

## **ABSTRACT:**

In rural areas, hydro break lines are necessary to create LiDAR derivative products such as TINs, contours, DEMs etc. Water body detection and delineation are necessary to generate hydro break lines. Hydro flattening follows the detection and delineation of water bodies (lakes, rivers, ponds, reservoirs, streams etc.). Manual hydro break line generation is time consuming and expensive, especially when there are large number of lakes, rivers and their tributaries. Accuracy and processing time depend on the number of vertices marked for delineation of break lines. Automation with minimal human intervention is desired for this operation. In this paper, we proposed an advanced algorithm to detect water bodies automatically using both LiDAR elevation and intensity data. We used two distinguishing characteristics to detect water bodies. The first one is that the surface of water body is a flat area and the elevation of it is lower than that of the surrounding area. The second one is that spectral reflectance of water is very small in the near infra-red range compared with vegetation and other topographic features. Beyond these two broad strategies, we used additional techniques to fine tune the algorithm. After detecting water bodies, we generated 3D break lines. Finally, we performed a comparison of automatically generated break lines with their semi-automated/manual counterparts and we discussed the results.

**KEYWORDS:** LiDAR, Water body, Classification, Break lines

## **1. INTRODUCTION**

Vast majority of the feature detection work appears to be in urban areas. Building tops, well defined fields, streets etc. are some of the more common features that are detected from overhead imagery and LiDAR. Our investigations unearthed very little work in automated generation of rural features such as water bodies and wetlands. Although manual and semi-automated approaches exist for some rural feature generation, these tend to be time consuming and expensive.

Water body detection is necessary to generate hydro break lines, which are in turn useful in creating derivatives such as TINs, contours, DEMs from LiDAR data. Surface water mapping is also essential in many applications such as monitoring of river corridor for natural hazard management (e.g. French, 2003; Brügelmann and Bollweg, 2004; Hollaus et al., 2005), surveying geomorphological change of floodplains (Thoma et al., 2005, Jones et al., 2007) and to improve understanding of wetland dynamics (Jenkins and Frazier, 2010).

Combining LiDAR elevation and intensity data gives better classification accuracy for water body detection (Antonarakis et al., 2008; Brzank et al., 2008; Höfle et al., 2009). Smeeckaert et al. (2013), classified water in coastal and river bank areas using Support Vector Machine (SVM) method. However, this method requires historical coastline or rough 2D manual plotting of seashore. By supervised fuzzy classification, Brzank et al. (2008) delineated water body in Wadden sea area, where the probability of presence of water bodies was attributed to low intensity and low elevation. However, this method is not applicable in land areas where water bodies do not necessarily have lower elevation. Especially in hilly regions, we can find water bodies located at different altitudes. Höfle et al., (2009) used a seeded region-growing algorithm to delineate water surface from land. However, this method requires significant pre-processing of LiDAR data such as dropout modelling.

## 2. OVERVIEW OF OUR METHOD

In our method, we proposed using LiDAR intensity data as well as an advanced histogram analysis of LiDAR elevation data to automatically detect and delineate water bodies with very little pre-processing. The surface of water body has the same elevation. So, if we create a histogram based on elevation, we generally notice a peak at the corresponding elevation in the histogram. However, small water bodies which typically have fewer pixels with same elevation might not result in a peak in the histogram. Hence, LiDAR intensity data is essential to detect smaller water bodies. Thus, our algorithm combines intensity and elevation in a unique fashion to detect small and large water bodies. The flow chart for the proposed auto hydro break line generation method is shown in figure 1. Our algorithm is comprised of two methods to detect water bodies. In the first method, we detected large flat areas (larger than half acre in size) as probable candidates for water bodies. Then, using the intensity of the probable area and, intensity and elevation of the surrounding areas we eliminated false detections. In the second method, we selected low intensity areas using a threshold as probable candidates for water bodies. Then, using the flatness of the probable area and, intensity and elevation of the surrounding areas we eliminated false detections. After detecting water bodies using the aforesaid methods we merged the detections to get the final detection.

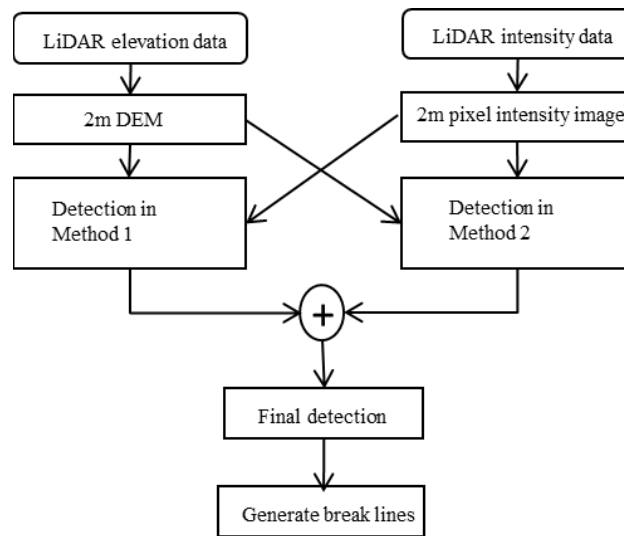


Figure 1. Flow chart for the proposed auto hydro break line generation method

## 3. DETAILED DESCRIPTION OF OUR METHOD

In this section, we describe in detail how we detected water bodies and generated break lines using both intensity and elevation data. We detected only the water bodies that are greater than half acre in size. At first, we trained our algorithm in a few training areas in L'Angeuille river basin, Arkansas, USA. Thereafter, we applied the algorithm in several test areas in the same river basin.

Here we describe our proposed method step by step for a test area. Please note that the specific numbers, units, scales etc. mentioned in this section refer to this specific example. We provide a logical explanation for any general conclusions reached as a result of working with this example.

### 3.1 Creating 2m DEM and 2m pixel intensity image

We rasterized the elevation and intensity data to detect water bodies, since point cloud based approach is complex and time consuming. Point cloud based approach can admittedly give us more accurate result. However, if higher level of accuracy is needed, we can use point cloud data around the hydro break lines generated using the comparatively simpler raster based method.

At first we exported the .las file for the test area to ASCII text format and read it in MATLAB. Then we divided the test area into a 2m x 2m square grid. Now to create a 2m DEM, in each 2m x 2m square, we kept all the single returns and, all the last echoes in case of multiple returns. Thereafter, we calculated the median elevation value of those LiDAR returns and assigned it as the elevation of that square. Interpolation methods such as the inverse distance weighted method, kriging etc., are generally used for DEM generation. However for our water body detection, we took the median elevation value to avoid some high elevation last returns which might not be from ground. We gave zero elevation and zero intensity value to squares, from which no LiDAR return was recorded.

To create the 2m pixel intensity image, we represented each 2m x 2m square by the median of all the intensity values of the single returns of the point cloud in that square and, in case of multiple returns, we took the last returns. Median intensity value can provide a better approximation of the intensity since it is not affected by the few high and low values in the square. The 2m DEM and 2m pixel intensity image for the test area are shown in figure 2 and figure 3 respectively. In the image, each pixel represents a 2m x 2m square.



Figure 2. 2m DEM of the test area in L'Anguille river basin



Figure 3. 2m pixel intensity image in log scale

### 3.2 Creating histogram of elevation and classifying the peaks

The range of elevation could be different for different areas. For the elevation histogram, if we select the number of bins to be proportional to the range of elevation, then for different areas, the number of bins required would be different. However, we wanted to use equal number of bins for all the areas and hence we selected a large number 1024, as the number of bins for better resolution. L'Anguille river basin is a comparatively flat area. For hilly area, the number of bins should be higher for proper detection of water bodies.

In order to create a 1024 bin histogram from the 2m DEM (shown in figure 2), we multiplied the elevation values of the DEM with a constant *scale* where

$$Scale = 1023/\text{highest elevation value in the DEM}$$

We calculated the *scale* for the test area to be  $12.0851\text{ft}^{-1}$ . The resulting DEM now has elevation ranging from 0 to 1023. We rounded the elevation values to the nearest whole number and we got a modified 2m DEM. We created a 1024 bin histogram using this modified 2m DEM. In the histogram, each bin represents an elevation and frequency of a bin represents the number of pixels with same elevation represented by the bin. We smoothed the histogram using a low pass filter with coefficients: [0.25, 0.5, 0.25]. The smoothed histogram is shown in figure 4. In the smoothed histogram, we determined the peak locations. If the frequency of any bin is not significantly higher than its adjacent or alternate bin, then we didn't consider it to be a peak. Therefore, if in the histogram, any bin has a frequency greater than that of the adjacent bins and, if the frequency is greater than 20 from adjacent bins or greater than 50 from alternate bins on both sides, then we defined the bin location as peak location. From the histogram, we calculated the peak locations, which are: [1 50 129 138 143 148 152 168 176]

If  $f(i)$  is the frequency of  $i^{\text{th}}$  bin then  $i^{\text{th}}$  bin location will be considered as a peak location if the following criteria are fulfilled

$$f(i) > f(i \pm 1) \text{ AND } [f(i) - f(i-1) > 20 \text{ OR } f(i) - f(i-2) > 50] \text{ AND } [f(i) - f(i+1) > 20 \text{ OR } f(i) - f(i+2) > 50]$$

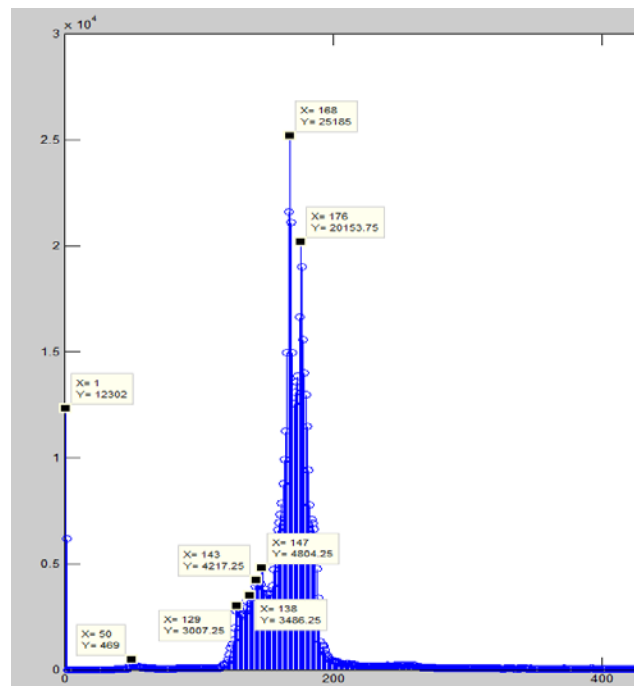


Figure 4. Smoothed histogram of the modified DEM

We classified the peaks into three main classes. We considered a peak in  $i^{\text{th}}$  bin location as class 2 peak if it fulfills the following criteria

$$f(i-1)-f(i-2)>200 \text{ AND } f(i-1)-f(i+1)>200 \text{ AND } scale>5.5$$

We considered a peak in  $i^{\text{th}}$  bin location as class 3 peak if it fulfills the following criteria

$$f(i+1)-f(i+2)>200 \text{ AND } f(i+1)-f(i-1)>200 \text{ AND } scale>5.5$$

We considered all the rest of the peaks as class 1 peak. In figure 5, we showed different types of peaks. We classified the peaks and varied the range of elevation taken for water body detection based on the class of the peak to increase the accuracy of detection.

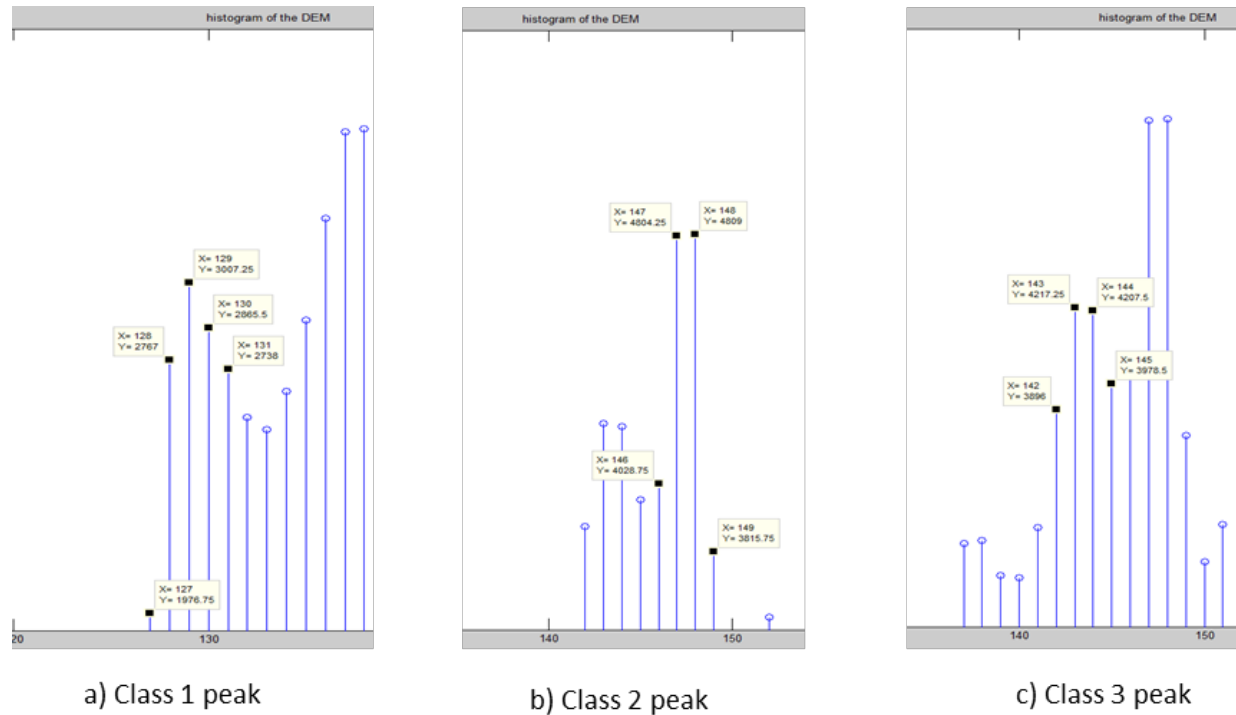


Figure 5. Different classes of peaks

### 3.3 Detecting water bodies using method one

The flow chart of the water body detection in method 1 is shown in figure 6. In the histogram shown in figure 4, each bin represents an elevation. It is highly probable to find surface of some water bodies in a certain range of elevation around peak bins. For three different classes of peaks, we calculated the range of elevation differently to detect water bodies.

For class 1 peak, the range is:  $(peak\ elevation - \left\lfloor \frac{scale}{3} \right\rfloor)$  to  $(peak\ elevation + \left\lfloor \frac{scale}{4} \right\rfloor)$

For class 2 peak, the range is:  $(peak\ elevation - \left\lfloor \frac{scale}{3} \right\rfloor - 1)$  to  $(peak\ elevation + \left\lfloor \frac{scale}{4} \right\rfloor)$

For class 3 peak, the range is:  $(peak\ elevation - \left\lfloor \frac{scale}{3} \right\rfloor)$  to  $(peak\ elevation + \left\lfloor \frac{scale}{4} \right\rfloor + 1)$

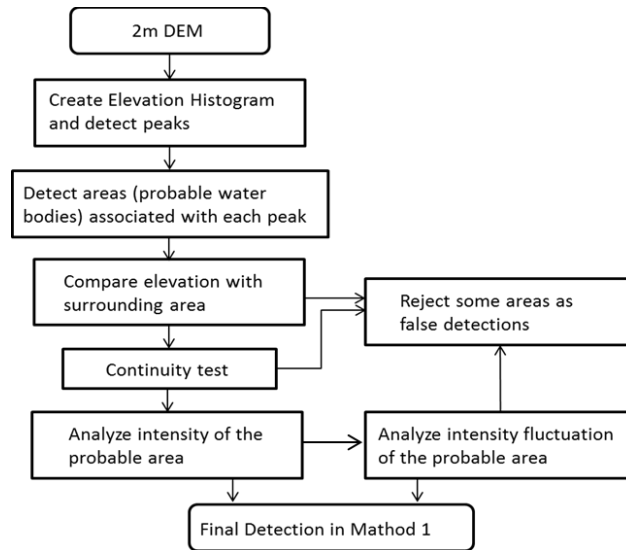


Figure 6. Flow chart for Water body detection in method 1

For each of the peaks, we detected all the pixels in the modified DEM that falls in the corresponding elevation range. For example, for the LiDAR data shown in figure 2 following are the specific steps. For the peak (figure 4) at bin location 143 (which is a class 3 peak) we detected all the pixels that fall in the range  $(143 - \lfloor \frac{12.0851}{3} \rfloor)$  to  $(143 + \lfloor \frac{12.0851}{4} \rfloor + 1)$  or 139 to 147. The pixels detected for peak at bin location 143 are shown in figure 7(a). We assigned zero elevation to all the 2m x 2m squares from which no LiDAR return was recorded, so we got a peak at zero elevation (bin #1) in the histogram. The pixels detected for peak at bin location 1 are shown in figure 7(b). We added the areas with zero elevation to the areas detected around peak at 143. After adding the areas, the resultant image is shown in figure 7(c). Then we performed a connected component analysis (4-connectivity based) and rejected all the areas less than half acre in size. The resultant probable water body areas are shown in figure 7(d). Many false detections also appeared in figure 7(d) and to filter out those false detections, we used the elevation and intensity of detected areas and their surrounding areas as well.

In our earlier reported work (Toscano et al. 2013), less number of sharp peaks were chosen for probable water body detection and thus a few water bodies were not detected. In the method proposed in this paper, more peaks were selected. In addition, the peaks were also classified into three different classes. For each of these classes different elevation range was assigned. This resulted in almost all the water bodies to be detected with more precision. However the number of false detection also increased. To eliminate those false detections, we used strong four stage nonlinear filtering algorithm, described in the following section.

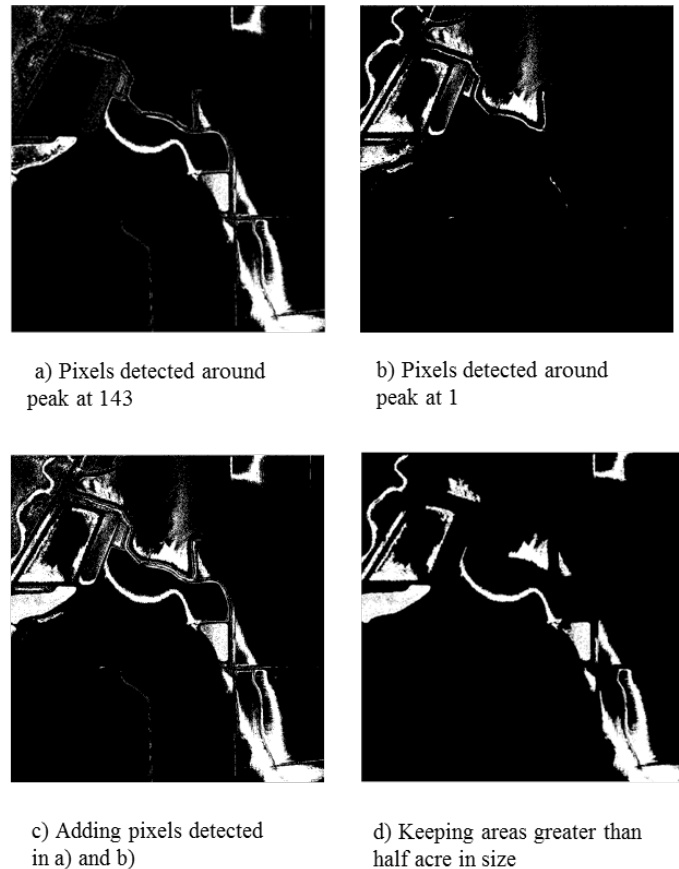


Figure 7. Steps of detecting probable water bodies before eliminating false detections for a particular peak

### 3.3.1. Stage 1 of filtering: Comparing elevation with surrounding area

The elevation of the surrounding area of a water body must be higher than that of the water body. We determined the elevation of each probable water body by calculating the median of all the elevation values of the pixels, where LiDAR returns were available. We postulated that a sampling of the surrounding area could be obtained from a chain of pixels approximately 4 meters away from the probable area. If more than 80% of this chain of pixels have elevation greater than the median elevation of the probable water body and, if the median elevation of the chain of pixels is greater than the median elevation of the probable water body by at least  $\frac{2}{3} \times \text{scale}$  (almost equivalent to 8 inches) then we did a continuity test on the probable area in the next stage for filtering. Ideally 100% of the pixels in the surrounding area should have greater elevation than the water surface. However, the LiDAR data we used has a 9.25 cm RMSE and so we lowered our criteria to 80%.

### 3.3.2. Stage 2 of filtering: Continuity test

Water bodies must be continuous and should not contain too many non-water body pixels. The non-water body pixels inside a probable water body area appear as voids. Closing the voids inside the water bodies will increase the size of the water body. We closed all the voids in probable water bodies which have radii less than 4 meters. After closing the voids if the area of the probable water body increases by more than 20%, then we rejected that area as false detection. We took all the other probable water body areas for the next stage of filtering. Inside a water body there might be large islands, and closing those voids might increase the size of the water body by more than 20%. So, we closed only the voids which have smaller radii.

### 3.3.3. Stage 3 of filtering: Comparing intensity with surrounding area

Water absorbs a significant portion of light energy in the near infra-red range. So the LiDAR intensity from water body is low compared with that of surrounding areas, especially when the deflection angle is high. We calculated the median intensity of every probable water body using the 2m pixel intensity image. We took a 10 meter wide (10 meters is an optimum value, neither too narrow nor too wide) buffer around each of the probable water bodies and calculated the median intensities of the surrounding areas. If the median intensity of the surrounding area exceeded the median intensity of a probable water body by a threshold  $t$  then we took that area to the next stage of filtering and rejected the rest of the areas. We calculated the threshold  $t$  by equation 1.

$$\begin{aligned} t &= m/12 && \text{if } m \geq 18 \\ &= 1.5 && \text{if } m < 18 \end{aligned} \quad (1)$$

where  $m$  is the median intensity of the 2 m pixel intensity image.

### 3.3.4. Stage 4 of filtering: Analyzing intensity of the probable area

If the median intensity of a probable water body is less than the median intensity of the 2m pixel intensity image, then we considered it as a water body. Sometimes, median intensity of the water body might be higher than the median intensity of the 2m pixel intensity image, when intensity values fluctuate significantly inside the probable area (this high fluctuation is observed when the deflection angle is low). At this stage, we calculated intensity fluctuation of each probable water body using the 2m pixel intensity image to classify probable candidate areas as water bodies that don't fall in the criterion mentioned above. To calculate the intensity fluctuation, we did a 2-level decomposition of the image using biorthogonal 4.4 wavelet filters and reconstructed the diagonal coefficients. We set a threshold of 4 to extract high fluctuation points from the diagonal coefficients. If, in any probable water body, more than 20% of the pixels are found to be high fluctuation points, then we considered that to be a water body.

Using these four stages of filtering we eliminated false detections and detected the water body for a particular peak. For each peak, we detected water bodies in the same way described in section 3.3 to obtain the final detection in method 1 as shown in figure 8.

## 3.4 Detecting water bodies using method 2

Sometimes small water bodies might not get detected by the method we proposed using method 1. For example, if the size of the water body is too small (which means that there are very few pixels with the same elevation), we might not get a peak. Hence, LiDAR intensity data should be used for detecting small water bodies. To detect a water body from intensity data, we identified all the pixels in the 2m-pixel intensity image, the intensity of which falls below 20<sup>th</sup> percentile. Using connected component analysis (4-connectivity based), we rejected areas less than half acre in size and assumed rest of the connected areas to be probable water bodies. We took a chain of pixels surrounding each of the probable water bodies and calculated the median elevation of the chain as described in section 3.3.1. If the median elevation of the chain is greater than that of a probable water body by at least  $2/3 * scale$ , we performed a continuity test on the probable area as described in section 3.3.2. If the probable area passes the continuity test, we considered the area to be a water body. If it fails the continuity test, we compared the median intensities of the probable water body and its surrounding area as described in section 3.3.3 and set the threshold  $t$  to be the median intensity value  $m$  of the 2m pixel intensity image, instead of the threshold described by equation 1. If the median intensity of the probable area was less than that of the surrounding area by at least the threshold  $t$ , we considered it to be a water body. Water body detected for the test area using method 2 is shown in figure 9.





Figure 8. Water body detected using method 1.



Figure 9. Water bodies detected using method 2

### 3.5 Merging detections from both methods

After detecting water bodies using both method 1 and method 2, we merged the detections using a logical OR operation. We rejected areas less than half acre in size using connected component analysis and for the rest of the areas, we performed a continuity test as described in section 3.3.2. Then using morphological operations, we closed small gaps or voids in the detected water bodies. The final detected water bodies are shown in figure 10.



Figure 10. Water body detected after merging detection from method 1 and method 2

### 3.6 Creating 3D break lines for detected water bodies

After detecting water bodies, we identified vertices from the edges of the detected water bodies to create vectors for the break lines. To generate 3D break lines, the median elevation of each water body (calculated from 2m DEM) is chosen as the  $z$  value. 3D break lines for the detected water bodies are shown in figure 11.



Figure 11. 3D break lines for water bodies shown in figure 10

## 4. RESULTS AND DISCUSSIONS

We compared the hydro break lines generated automatically by our method with the break lines drawn in a semi-automated way by a vendor. In figure 12, our break lines are shown by thin, black lines and the break lines drawn in semi-automated way are shown by thick, blue colored lines. From the comparison in figure 12, we found that the auto generated hydro break lines closely matched with the break lines drawn in semi-automated way. All the water bodies greater than half acre in size were detected by our method. In case of the two water bodies shown by green arrows, some part of the land along the edge of the water bodies was misclassified as water body because of the small elevation difference between the water body and adjacent land area.

We applied our method in another test area in L'Anguille river basin and compared our auto generated break lines with their semi-automated counterparts, which is shown in figure 13. At the bottom right corner of the large water body we found that the break lines did not match well. And we also found that at the middle of the small water body, there is a little mismatch between the two break lines. The mismatched areas are shown by green arrows. In

this case also the mismatch occurred due to the small elevation difference between the water body and its adjacent surrounding area.

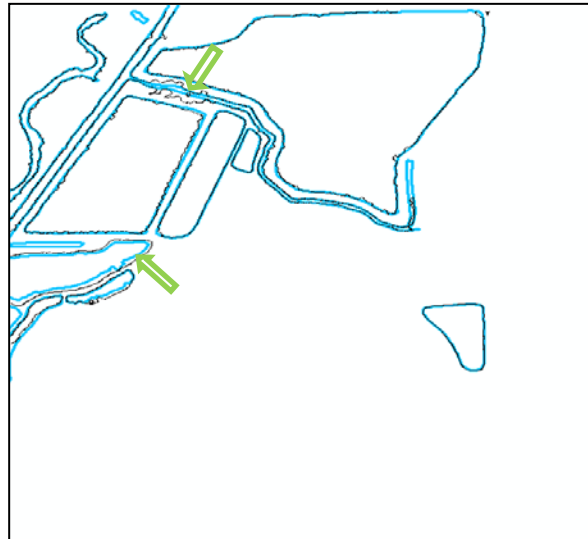


Figure 12. Comparison between break lines generated in our fully automated method (thin black line) and break lines drawn in semi-automated method (thick blue line).

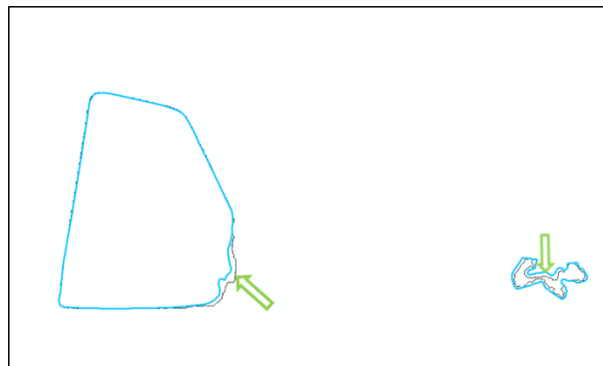


Figure 13. Break line comparison between our fully automated method (thin black line) and semi-automated method (thick blue line).

## 5. CONCLUSIONS

In general, feature detection work in rural areas is somewhat sparse. Detecting water bodies from LiDAR data is a very challenging task due to high dropout rate from both water bodies and mudflats and due to high intensity variation caused by rough water surface and specular reflection from water bodies.

In this paper, we used a non-linear heuristic method to detect and delineate water bodies and generate break lines primarily in a rural area, using both LiDAR elevation and intensity data. We found the break lines generated using our fully automated method closely matched with the break lines drawn in a semi-automated way. As a future work, we propose the use of LiDAR point cloud data in the areas around the hydro break lines generated by our method to improve accuracy and remove much of the drudgery associated with the manual elements of this work.

## REFERENCES

- Antonarakis, A.S., Richards, K.S., Brasington, J. 2008. Objects-based land cover classification using airborne LiDAR. *Remote Sensing of Environment*, 112(6), pp. 2988–2998.
- Brügelmann, R., & Bollweg, A., 2004. Laser altimetry for river management. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 35, pp. 234–239
- Brzank, A., Heipke, C., Goepfert, J., Soergel, U. 2008. Aspects of generating precise digital terrain models in the Wadden Sea from lidar-water classification and structure line extraction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63(5), pp. 510–528.
- French, J.R., 2003. Airborne LiDAR in support of geomorphological and hydraulic modelling. *Earth Surface Processes and Landforms*, 28(3), 321–335
- Höfle, B., Vetter, M., Pfeifer, N., Mandlbürger, G., & Stötter, J., 2009. Water surface mapping from airborne laser scanning using signal intensity and elevation data. *Earth Surface Processes and Landforms*, 34(12), pp. 1635–1649
- Hollaus, M., Wagner, W., and Kraus, K., 2005. Airborne Laser Scanning and Usefulness for Hydrological Models. *Advances in Geosciences*, 5, pp. 57–63
- Jenkins, R.B. and Frazier, P.S., 2010. High-Resolution Remote Sensing of Upland Swamp Boundaries and Vegetation for Baseline Mapping and Monitoring. *Wetlands*, 30(3): pp. 531–540
- Jones, A.F., Brewer, P.A., Johnstone, E., Macklin, M.G.. 2007. High resolution interpretative geomorphological mapping of river valley environments using airborne LiDAR data. *Earth Surface Processes and Landforms*, 32(10), 1574–1592.
- Smeeckaert, J., Mallet, C., David, N., Chehata, N., & Ferraz, A. (2013). Large-scale classification of water areas using airborne topographic lidar data. *Remote Sensing of Environment*, 138, 134–148.
- Thoma, D.P., Gupta, S.C., Bauer, M.E., Kirchoff, C.E., 2005. Airborne laser scanning for riverbank erosion assessment. *Remote Sensing of Environment*, 95(4), pp. 493–501
- Toscano, G. J., Gopalam, U., and Devarajan, V., 2013. A novel method for automation of 3D hydro break line generation from LiDAR data using MATLAB, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-2/W2, pp. 99–104

## ACKNOWLEDGEMENT

This research was funded by NRCS (Natural Resources Conservation Service), Ft.Worth, TX. We are thankful to Mr. Steven Nechero, Collin McCormick and William Marken of NRCS for providing us LiDAR data along with useful technical guidance.