Cross-Country Course Elevation Analysis

Design Document

David Kirshenbaum
Andrew Mumm
Connor Smith
Thomas Chambers
Ryan Hillby
Jacob Feldman

Table of Contents

1. Reference Materials

a. List of Figures

b. List of tables

c. List of definitions

LIDAR (also LiDAR, LIDAR, or LADAR) - Light detection and ranging. A method of measuring distance in which lasers are aimed at the target, and the return time and wavelength is measured in order to calculate distance to the target.

Differential GPS - An improvement on the traditional GPS which uses a network of known ground-based stations such that the base stations apply a correction to the GPS data received from the satellites. This can improve accuracy from about a 10 m error to as little as several centimeters.

GIS - Geographic Information Systems. A framework for gathering, analyzing, and viewing data related to Earth including topography, roads, terrain, etc.

2. Introductory materials

a. Acknowledgement

i. **Dr. Amy Kaleita:** We would like to thank Dr. Amy Kaleita for lending us the differential GPS equipment that is crucial for the ground truth team to collect data for validation.

ii. **Dr. Bradley Miller:** We would like to thank Dr. Bradley Miller for teaching us more about GPS and topographical data. What we learned from Dr. Miller is greatly beneficial for data analysis.

iii. **Dr. Brian Hornbuckle:** We would like to thank Dr. Brian Hornbuckle for creating the project and helping us find the resources we need to succeed.

b. Problem statement

General Problem Statement: There is a suspicion that cross country courses are becoming less hilly over time. Hilly and difficult courses are integral to the spirit of cross country as a sport. If courses are becoming less hilly then there is a growing deviation from the true spirit of the sport. This can also present a great disparity regarding what is expected on a competitive level for cross country courses. We are attempting to answer the question, “Are cross country courses becoming less hilly?”

General Solution Approach: Our project consists of three main parts with a potential fourth part granted that we have already answered the main question of this project. The first phase of the project is data collection. We will collect GPS data and topographical information from various sources including manual collection from GPS units, Google Maps data, and LIDAR data. The next step is to convert this data into formats that we can easily use and relate to each other. When the data is compiled in a uniform format, we will be able to verify what source of data is the most reliable for analysis. The third phase of the project is to compare this data with the topographical data we collect from courses over time. Through this comparison we will be able to determine if the hilliness of courses has generally reduced over time. We will also develop software that allows the user to supply the program with an existing course, and the program will analyze the hilliness of the course and give it a rating. The fourth potential part of this project is to develop software that would allow course designers to supply a program with a region on which they wish to build a course. They would then be able to specify various details about the course such as degree of difficulty (hilliness), length, number of turns, and a start and end of the course. The software would then generate a course for the user over the region that they supplied.

c. Operating environment

By the end of the project our main deliverables will be entirely software. There will not be any environment concerns for the use of our project deliverables. However, our end project solution needs to be viable in areas of heavy tree cover and other foliage.

d. Intended user(s) and intended use(s)

Our software will be used by any officials and course planners when evaluating or creating new cross-country courses. Our intention is that the software will be useful to officials at every level of the sport ranging from small 1A Iowa high school athletic directors to the highest levels of the NCAA. The goal of our product is that it will provide users with the ability to accurately evaluate and create courses. This will help to bring some form of standardization to the sport while staying true to the spirit of cross country.

e. Assumptions and limitations

One assumption is that people are actually going to use and benefit from the software we create. We believe that once this product is created there will be enough interest in whether or not particular courses are hilly enough that officials and course planners will use this software.

A second assumption is that we are going to be able to locate a database that is accurate and plentiful enough to be able to create our end product. If we can't find a large enough source of accurate data, then we will not be able to reliably use our product to test the hilliness of courses.

Currently we are only certain of high-resolution, comprehensive LIDAR data being available in Iowa. If we want our product to be able to be used in other states or even other countries, then we would need an accurate and plentiful data source for the area in which the product is to be used.

f. Expected end product and other deliverables

Source of Truth: We will be analyzing several data sources including LIDAR, Google Maps, and GPS in order to find one data source that is accurate enough to measure elevation in a cross-country course.

Cross Country Course Creator and Evaluator: We will develop software that gives the user the ability to supply the program with an existing course by drawing the course on a map in the program. The program will analyze the hilliness of the course, subsequently giving it a rating. The software would also allow course makers to supply a program with a region that they wish to have a course on. They would then be able to specify various details about the course such as degree of difficulty (hilliness), length, amount of turns, and a start and end of the course. The software would then generate a course for the user over the region that they supplied.

3. Specifications and Analysis

a. Proposed design

Following our analysis of different data sources' precision of elevation measurements, we have decided to design our application to work with the bare-earth model LIDAR data set from the Iowa Department of Natural Resources. This LIDAR model of the state of Iowa will serve as the elevation ground truth by the application. The LIDAR data itself will be stored on a central server to be queried by the application on a county-by-county need basis.

In conjunction with the LIDAR model, the application features three different methods of accepting XY waypoint user inputs to represent a cross country course path. The primary tool for user input of courses is a path drawing tool atop Google Maps' satellite imagery for the course area. The path drawn on the satellite imagery is stored as a series of latitude and longitude coordinates. Alternatively, the user can also generate this equivalent datafile by making use of a GPS tracking feature of the application. This tracking feature operates by simply making measurement requests to the device's GPS unit in one-second intervals. It should be noted that a more precise dedicated GPS unit such as a Garmin Montana-class device could also be used for XY mapping of the route, and the application will also feature an option to input a .csv file where such coordinates would be stored.

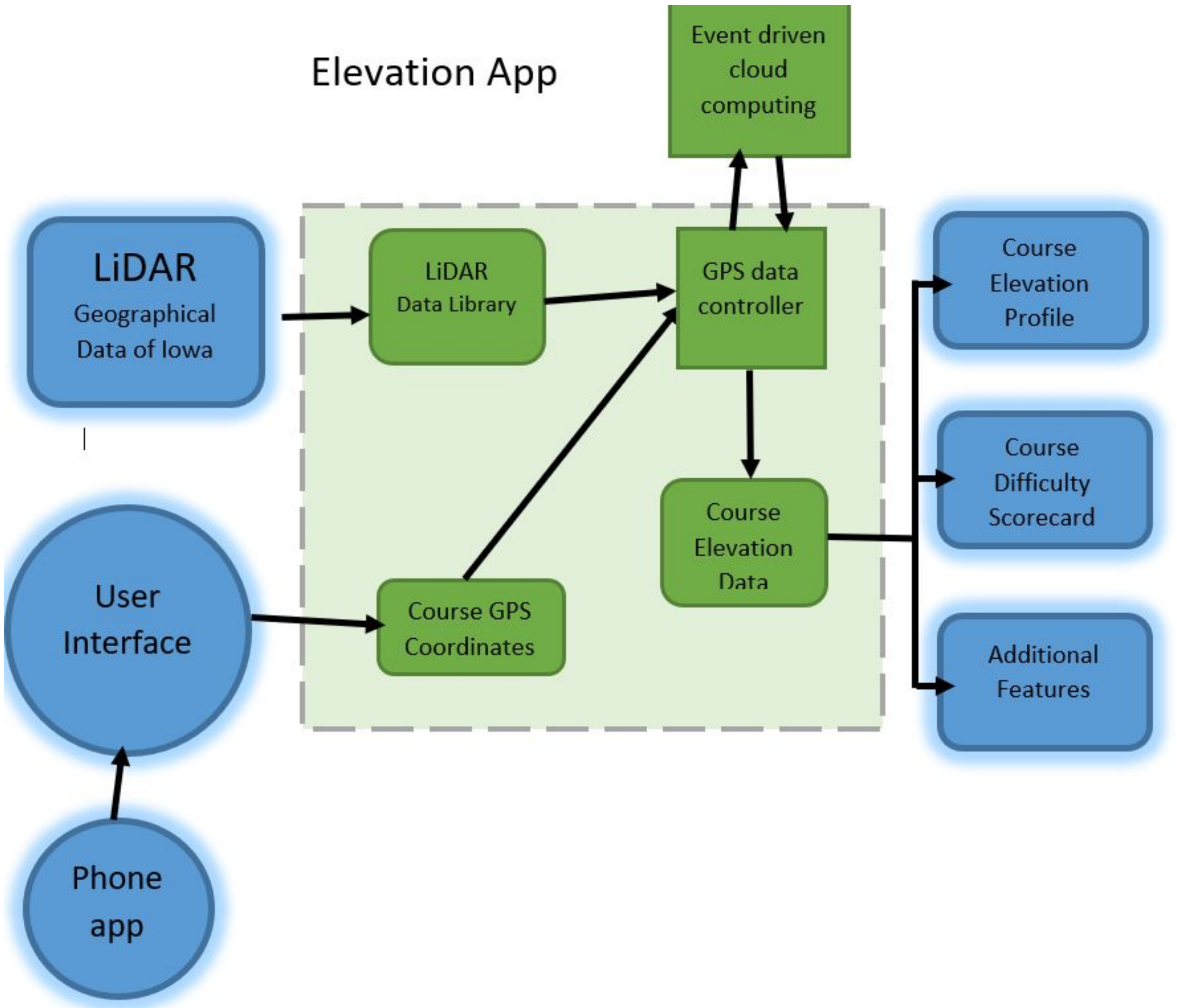
Regardless of the approach, the XY coordinate file should be displayed in an R Shiny application through Shiny's "Interactive Map" tool. This display of the route serves as a data validation stage for the user, as they are able to visually guarantee that the course path they intended to capture with either the drawing tool or their own GPS unit is properly represented. Before continuing on to the next screen, the user is asked if the route looks accurate to their eye. If the user answers in the negative, they will then be prompted to re-input the XY coordinates. This cycle will continue until the user answers in the affirmative that, yes, they are satisfied with the course as visualized on the Shiny "Interactive Map" tool.

With the course now being digitally captured in waypoint form, the XY coordinates is next superimposed on the LIDAR point cloud as accessed from an online server. Given that the LIDAR point cloud has a 3 meter horizontal resolution, the XY coordinates are grouped on an individual basis to the LIDAR point with the minimal vector magnitude distance difference the LIDAR point's XY address. Once this association process is completed, the course is effectively represented in the LIDAR datum, allowing us to extract the elevation (Z coordinates) for the path into a .csv file.

This Z coordinate .csv file is then read in by an R script to commence the analysis of the course elevation profile. This R script will perform a series of calculations on the elevation data ranging from merely generating a two-dimensional elevation profile chart to measuring the lengths of different trends in the elevation data and displaying the classifications in a histogram representation, illustrating the quantity and quality of different hill/downhill elements along the course route.

Once the various analyses are completed via R, their subsequent visualizations will be displayed on a generated “course scorecard” printout along with the top-down perspective of the route overtop satellite imagery of the course area. The user is then presented with options to share the printout via social media or download as a PDF for printing.

An illustration regarding the general structure of the app’s processes is seen below.



b. Design analysis

So far, we have done a few runs of data collection using GPS data from multiple phones and a differential GPS unit in an attempt to determine the accuracy of these units and the accuracy of LIDAR data itself. Thus far, we have determined that LIDAR is significantly more accurate and consistent. Additionally, we have begun writing a script to compare our datasets against one another directly to quantify exactly how accurate each of the data sources are. Originally we tried to use a Python library that only worked with point clouds, but we have since decided to use a raster file instead (the bare-earth model described in the previous section). In the future, as mentioned in the last section, we will be using R to finish this script. We have also begun writing code to obtain waypoints via a user-drawn course map. This code is being written to be a part of our future app, and it uses the Google Maps API.

Thus far, all the code we've written is in very early stages, so almost none of it is fully functional yet. Data collection as well is not complete; we have collected some preliminary data that is already indicating some very clear trends of accuracy to us. We still need to formalize this data collection, however, and we have also decided to collect data from at least 3 cross country courses. This data still needs collected.

Our design, as it stands, it not finalized. We may decide that a particular method of inputting the x and y coordinates of the proposed course is good enough that having multiple methods of input is pointless. We will also need to flesh out what kinds of statistical analysis we will be doing with R on our data, once we have our data. This would include how we are going to create a 'elevation profile' for a course.

4. Testing and Implementation

a. Interface specifications

For the application we are creating, there is not much interfacing necessary. The user will input their course either by drawing the course on their computer or by walking through the course with their phone's GPS. If they choose the phone GPS option, we will create a simple app that gathers latitude and longitude coordinates as they walk so that the file type will be the same for all users. They will then send that data, which will simply be an ordered set of latitude and longitude coordinates, to our server. Our server will then figure out which county those coordinates lie. Then our server will send the set of coordinates and the file containing LIDAR data to a cloud computing Lambda, due to the LIDAR file being over 1 GB in size per county. The cloud computing will return a set of coordinates containing latitude, longitude, and elevation to the server, where the server will plot on graphs, eventually returning the graphs to the users computer for display.

b. Hardware and software

During the testing phase, GPS hardware and software will be utilized to collect geographical data from courses. This includes two dedicated GPS devices - a Garmin Montana 680t handheld unit and a Thales Navigation Promark2 differential GPS system. The GPS units are being used as a solid source of data with which to compare LIDAR data. The project also includes the use of two mobile phones running the free app "GPS essentials". This app is not being used with the intention of it being a reliable data source, but rather it will be used as a comparison to more reliable data sources to show why our planned app will be necessary and useful to users. The phones running the app will be a cheaper model Moto G Play and a more expensive model Google Pixel.

ArcGIS software is being used for data processing in the testing phase as well. “GIS” stands for “geographic information system”, which means the software is designed for mapping and geographical tasks. It is being used to view and manipulate the LIDAR data, as well as output the data in a .geotif format that can be used in more applications.

R Shiny is another software package being tested now that is being used to create the final app. It is a straightforward tool for designing apps that allows for the display of interactive visualizations. With this tool an application will be designed that lets users enter points on a map and send that data to be processed for elevation.

c. Process

- For the grading of courses, as stated earlier, we will be using the LIDAR data for the z-value of the coordinates. We will validate the accuracy of the LIDAR data z-value in a few different ways. One thing we have done is to check for validity is if the z-value differs from one source to another then we will verify if the z-value is consistently different. For example if a data source we are checking against is 30 feet higher at every point we check against then this data source is a valid data source to check against. If the z-value is 10 feet off, then 50 feet off for another point, and then spot on for another, then it is probably an invalid source to test against.
  - The first way that we have attempted to validate the LIDAR data is by plugging in a test xy coordinate into google maps and checking how the z-value matches up. This is not a great source to test the data as it is from a third party, but it does give an additional source of validation. We have read reports that google maps elevation is not a great source, so it will not be our main validator in this project.
  - The second way was using physical GPS devices to collect elevation data. These data points are being collected on known elevation points with the assistance of the GNSS Survey DOP planning tool. We are currently evaluating the accuracy of this data. However, the preliminary results show the phones are not very precise compared to the Garmin Montana 680t and the Thales Navigation Promark2 differential GPS system.

d. Results

A number of surveys have been completed to collect GPS data. One survey produced a comparison between phone GPS and dedicated GPS device data, which was made to demonstrate that phone GPS is not as accurate. The results came out as expected but more tests are needed in this area to satisfactorily demonstrate the conclusion that phone GPS is not accurate enough for the task of cross country course elevation.

A second survey using a differential GPS unit did not create satisfactory results. Two hardware problems plagued the process. The unit experienced signal interruption, which prevented any collection of data when under trees and near buildings. This is a problem associated with the way it collects data and can only be overcome by remaining in more open areas. The device also is an older model which uploads data using a program that existed on older versions of Windows, and uploading the data to a computer running a modern operating system is a task that still needs to be overcome.

5. Closure materials

b. Closing Summary

Our team and client, Dr. Hornbuckle, believe that cross country courses are becoming less "hilly." We also believe that this is contrary to the spirit of the sport of cross country. In order to prove this, we will create a way to rank the "hilliness" of a course using various sources of elevation data, namely LIDAR, Google Maps, and GPS units. For our findings to hold any water, we must assert that the data we are using is accurate, so we will compare the differences between all our data sources to determine their accuracy. We will also write some software to automatically generate a "hilliness" profile for a course, and it will also generate a course when provided a user specified "hilliness" profile and an area of land.

c. References

i. GNSS Survey DOP planning
https://www.gnssplanning.com/

ii. NGS Geodetic Marker Datasheets
https://www.ngs.noaa.gov/NGSDataExplorer/

iii. Gantlab - chart making tool
https://live.ganttlab.org/

iv. Iowa LIDAR Mapping Project - source of unprocessed LIDAR data for Iowa
http://www.geotree.uni.edu/LIDAR/

v. Three Meter Digital Elevation Model of Iowa, Derived from LIDAR - the same LIDAR data from the Iowa LIDAR Mapping Project, but smoothed out and with trees and buildings removed.
https://geodata.iowa.gov/dataset/three-meter-digital-elevation-model-iowa-derived-LIDAR

vi. PDAL - Point Data Abstraction Library we will be using to process the LIDAR data stored as point clouds
https://pdal.io/

vii. Google Maps JavaScript API
https://developers.google.com/maps/documentation/javascript/tutorial