Cross-Country Course Elevation Analysis

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 part of the spirit of cross country. 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. Proposed approach and statement of work

A. Functional requirements

The initial ground truth validation studies need to provide definitive information regarding the accuracies and, subsequently, the viability of using topographic data sources available that are also feasible and scalable to a wider deployment.

The software tool needs to be able accept .las data files as inputs to easily allow users to provide the source data themselves.

The software tool must be able to run classification algorithms on the elevation profiles and classify hill-like topography in to subclassifications as well as quantify the curviness of routes.

The classifications of elevation profiles and route curviness must be presented in a visually appealing manner and in an easy-to-interpret format.

B. Constraints considerations

The chief constraint (which has an uncertainty currently revolving around it) is the accuracy of the obtainable topographic data for rural, isolated cross country courses. If it is determined that LIDAR data is inadequate to produce accurate classifications of course topography, the time-intensive nature of on-site GPS surveying methods will reduce the feasibility of a wide-scale adoption of the final software deliverable. Furthermore, Iowa's harsh winter climate will constrain our ability to easily conduct GPS surveys past the end of autumn.

The technological savviness of the software's users is also critical to the project. It is key that the user interface is as simplistic as possible in order to lower the intimidation of the entry barrier for cross country officials at every tier of the sport including older, small town athletic directors. Since the use case of this software has never been tested before, it is only reasonable to expect the users to be apprehensive about giving it a try.

C. Technology considerations

In order to produce the best analysis software under the limited time scale of the next two semesters, the construction of the software is beginning while simultaneously conducting empirical research to determine the best data source for our application. This "empirical research" is occurring in a methodical process of directly comparing a variety of elevation-measuring technologies over varying topographies. These technologies include two cell phones (Moto G Play and Google Pixel), a Garmin Montana 680t GPS, an Ashtech Promark2 differential GPS, and LIDAR data from both a 2017 survey conducted by the Iowa DNR and Google Earth's mix of LIDAR and topographic elevation data.
Cross country courses are defined by their vast variety of physical features including but not limited to large open fields, densely wooded valleys, zig-zag routes up hills, and proximity to mountains. All of these geographic features are detrimental to the dilution of precision (DOP)

value that is used in calculating the reliability of a GPS XYZ calculation. To account for this, we will test each of the measurement technologies on a variety of different terrains and routes with the intent of observing disparities between technological adequacy across course elements. It is only through this process that the truly best data source can be chosen for elevation signal processing.

The aforementioned disparities will be computed using a point-by-point RMS error calculation of the vertical distance component of each data point to the exactly straight lines from verified starting and ending point XYZ coordinates for each survey.

Given the wide price variation of the physical measurement devices (from several hundred dollars for the phones to several thousand dollars for the differential GPS unit), it is expected to see a correlating variation in these devices' performances. The state of Iowa is lucky enough to be entirely mapped with LIDAR at a 3 meter resolution, likely doing away with the need for ground based GPS course data generation. However, if the software is to be in unmapped states, it will be crucial to determine the cost-benefit relationship of all ground-based GPS devices. While a differential GPS unit may provide marginally better data than the Garmin Montana 680t unit, the Garmin is much easier to operate and, therefore, much more scalable as a course mapping procedure. Our ability to find the "sweet spot" of devices between low cost/ease of use and data integrity will be paramount to the adoption of the software.

In regards to the software we will be using, we are trying to use the best available tool for each part of the project. For data validation as well as creating visual aids to accompany the data, we will be using ARCGIS. ARCGIS is extremely powerful and one of its main use cases is to display changes in terrain just as we are doing. In regards to the application which will grade a cross-country course, we will be using PDAL, a Python "Point Data Abstraction" library, to handle our .las and .img files on the backend. The front end will be using Google Maps Javascript API so the users may input their course routes to be analyzed. Bear in mind that we are NOT using Google maps as the data source for elevation, simply just as a way to input.

D. Testing requirements considerations

As previously mentioned in the technology considerations section, we plan to methodically compare all data from each source in different conditions to determine the best data source for the software's input. While all devices will always be tested at the same time on each set of new terrain tests, special care will be taken to conduct these site surveys on exclusively sunny days at hours when the alignment of the GNSS constellation allows for the lowest DOP value.

As part of the algorithm development process for topographic classification, it will be necessary to test our ideas of hill classifications etc. against qualitative perspectives of those same physical features. This will potentially require the running of different routes by team members to verify that the divisions in classifications made by the algorithm are intuitive from a runner's perspective.

E. Security considerations

Our project will have very little to no security considerations. We are dealing with all public information, and we aren't planning on saving any user data through our software. The only possible (although unlikely) area that could be a security risk is if we decide to have users of our software make accounts that contain their personal information to make the account. Even then, the most sensitive information that would be stored is an email, a name, and maybe a school name. If we do decide to have users make accounts, then we will need to encrypt their data when it's stored.

F. Safety considerations

There are minimal safety considerations for this project. End-user activity will either be using computer software or walking a cross-country course with a handheld unit. There are no concerns beyond standard every-day safety guidelines.

G. Previous work / literature review

The biggest competition for our software comes from the mobile phone app markets. There are many different GPS programs which can provide information about elevation. "Elevation Profile" and "GPS Essentials" are two general purpose apps that offer elevation information. There are also many running apps, such as "Runkeeper" and "Strava" and these market directly to this project's intended audience of runners.

An important task for the project is to make a comparison between the data collected by a phone GPS or handheld GPS and the output generated by our own process. The goal is to show that gauging elevation variation on a track is a task too precise for standard GPS devices, which often have a large margin of error or do a large amount of averaging.

H. Possible risks and risk management

Risk is relatively minimal in this project. There is not significant capital investment. The work revolves around developing a process based on existing software and devices previously in possession of the team.

One consideration, however, is expensive equipment on loan to the team. They will be making use of two GPS devices that are of significant cost. Care will be shown to ensure any danger of theft or damage is minimalized. Specifically, when the team is surveying with the differential GPS unit there is a standing policy to never leave the stationary 'base' unit unattended while the mobile 'rover' unit is collecting data.

I. Project proposed milestones and evaluation criteria

The first milestone is determining what data source is accurate. The goal is to verify Iowa's LiDAR data, but also to have a general understanding of the accuracy of different options. Since there exist utilities that can produce an elevation profile, it is necessary to determine how effective they are at the task of mapping a track.

The second milestone is to produce an accurate elevation profile of a cross-country course located in Ames using one of the methods available. This will involve collecting and/or processing the data to accurately represent the changing altitudes experienced by a person running the course.

The third milestone is to devise a system to produce an analytical report from the data for the target audience that shows the course characteristics. This could involve a number of different measurements, from a maximum to minimum calculation to number of peaks of a certain height. The ability to produce a general measurement or grouping for the course by difficulty or “hilliness” is also a goal.

The fourth milestone is to produce an accessible method by which this process could be replicated by users on other courses.

J. Project tracking procedures

The project will be tracked using the repository software Gitlab. Additionally, weekly status reports will be filed by the team that detail progress made toward each milestone. This includes weekly contributions of each member and goals moving forward. The client will meet with the team weekly to discuss progress and secure any needed contacts and resources. Finally, the team is continually using Slack for all discussions related to the project and individual members' status reports.

4. Estimated resources and project timeline

A. Financial Requirements

In terms of hardware costs, a Garmin Montana 680t and an Ashtech Promark2 has been loaned to us by Dr. Bradley and Dr. Kaleita for our initial data collection. We also used a Moto G and a Google Pixel from our team to conduct our initial analysis. Thus, our costs of using this GPS equipment is $0.

In terms of software, we plan to use open-source tools wherever possible to process our data. The ArcGIS software to initially extract elevation points from LIDAR is provided for free to students at the GIS computer labs. Thus, our costs using the software is currently $0. However, if the current free open-source tools are not adequate to process the LIDAR data, we will need to explore other paid options.

As of right now we will also be using the Google Maps API in order to display maps to users and have them trace their courses. The free tier of Google maps is $300 per month, which we should not exceed during the development phase, thus making the cost of using the API $0.

Potentially, our team could take a trip to collect data from two XC courses out of town. We have not formally discussed this plan as it was thrown out as an idea. If we would do this trip, we would need to account costs for travel. Otherwise, our current financial requirements are $0.

B. Other Requirements

In order to analyze XC courses, we must verify our LIDAR data is accurate. Thus, we are using the following GPS equipment to verify the accuracy and precision of the LIDAR data collected:

Garmin Montana 680t, Ashtech Promark2, Moto G, Google Pixel.

These devices will help determine the best and most feasible source to use for our project. Since we need to filter our data points, we plan to use a visual point cloud to constrain our data. For the software tools we are going to develop, we plan to use open-source tools such as Liblas to import and manipulate LIDAR data and the Google Maps API to allow users to draw their XC courses on the map.

C. Personnel Requirements

The table below lists the major tasks that needed to be completed for our project.

Task	Description of Task	Estimated Time
Comparison of GPS and LIDAR data	Collect elevation and coordinate information using several GPS devices and	60 hours

	compare results to the Iowa LIDAR mapping project	
Obtaining and reading LIDAR data	Obtain LIDAR data and convert it into a format to be read easily by our own tools	20 hours
Extracting elevation and coordinates from LIDAR	Convert LIDAR information into points with elevation values at each coordinate	25 hours
Course Map Drawing Tool	Create tool using Google Maps API to allow users to draw course maps	40 hours
Deriving Profile of XC Courses	Formulate and quantify difficulty of courses based on elevation, slope, frequency, etc..	140 hours
Gather feedback	Talk to XC coaches, athletes, and IHSAA officials to understand what they want out of the tools	30 hours
Course Generator	Develop tools based on feedback and profiles to generate XC courses based on an area	120 hours
Documentation	Create tutorials and documents of the tools we created	40 hours
Final Analysis	Answer our original hypotheses	40 hours

Table 1: Tasks to complete

D. Project Timeline

The first half of our project will be dedicated to collecting data through various GPS sources to determine the accuracy of these sources along with the LIDAR data. Once the accuracy and precision of the our data sources are determined, we will move to building our tools to analyze the LIDAR points and to generate profiles of XC courses. The second semester is mainly focusing on creating statistical analytics of XC courses to present to coaches and athletes. We will also focus on creating a course generator for users to create a course based on the difficulty they choose. Lastly, we would report on the conclusions we found in our final report.

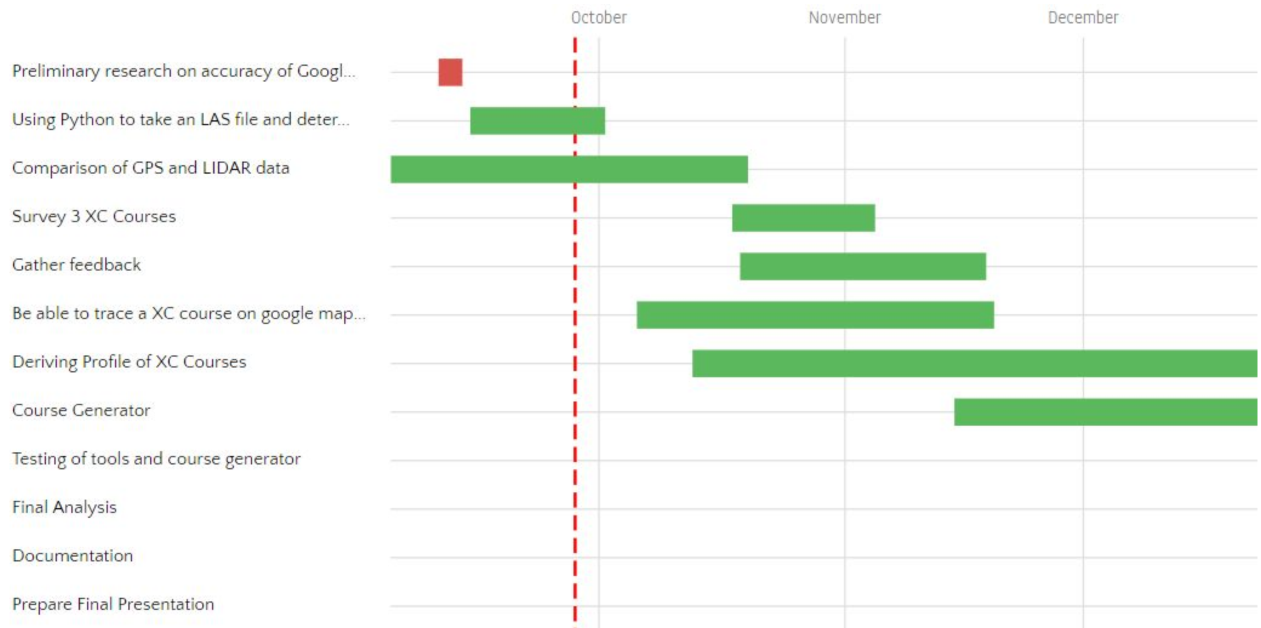

Figure 1: Gantt chart of project timeline of first semester

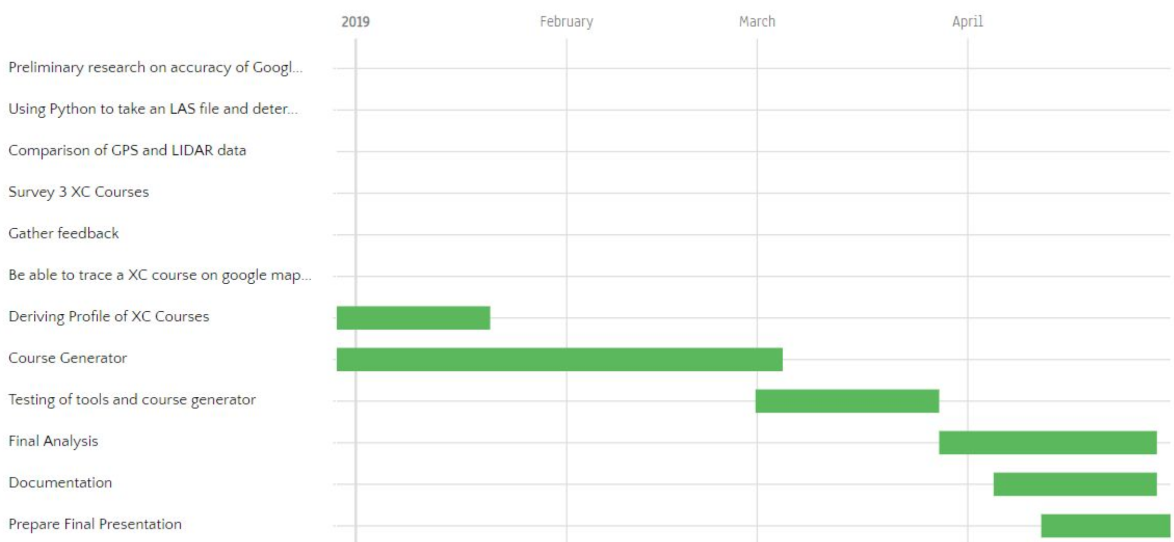

Figure 2: Gantt chart of project timeline of second semester

Tool used to create Gantt chart: https://live.ganttlab.org/

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