ANALYZING VOLUNTEERED GEOGRAPHIC INFORMATION ACCURACY AND DETERMINING ITS CAPABILITIES FOR SCIENTIFIC RESEARCH DATA

An Undergraduate Research Scholars Thesis

by

MICHAEL SCHWIND, KELSI DAVIS, PAYTON BALDRIDGE

Submitted to Honors and Undergraduate Research
Texas A&M University
in partial fulfillment of the requirements for the designation as an

UNDERGRADUATE RESEARCH SCHOLAR

Approved by
Research Advisor: Dr. Daniel Goldberg

May 2014

Major: Geography

TABLE OF CONTENTS

ABSTRACT

Analyzing Volunteer Geographic Information Accuracy and Determining its Capabilities for Scientific Research Data. (May 2014)

Michael Schwind, Kelsi Davis, Payton Baldridge
Department of Geography
Texas A&M University

Research Advisor: Dr. Daniel Goldberg
Department of Geography

The primary purpose of this research project is to develop, test, and evaluate a volunteered geographic information (VGI) based approach for collecting data in order to assess its accuracy and relativity to a field of scientific research. As such, this research project is composed of three specific aims: 1) Develop a series of mobile VGI applications, accompanying websites, and databases to enable the collection, storage, visualization, and analysis of VGI data; 2) perform usability studies with specific groups of end-users to evaluate the appropriateness of data model and application design; and 3) determine fitness-for-use of the data within a limited set of commonly-employed of spatial analysis techniques sought after by our test audience.

ACKNOWLEDGEMENTS

We would like to thank Dr. Daniel Goldberg for his support on this project. Because of the instruction and resources which he has provided, this thesis and our knowledge of volunteered geographic information has greatly improved.

We would also like to thank Zhongxia Li for his assistance in the development of the CatMapper. Because of his help we were able successfully create this mobile application which will hopefully aid in the understanding of VGI, as well as map cats.

NOMENCLATURE

GIS	Geographical Information System
VGI	Volunteered Geographic Information
SDK	Software Development Kit
GPS	Geographical Positioning System
PDOP	Positional Dilution of Precision

CHAPTER I

INTRODUCTION

The uncertainty of volunteered geographic information

Volunteered Geographic Information (VGI) has become a popular tool in mobile data collection scenarios because of its convenience and an open market of millions of users worldwide. VGI is geographic data that is provided voluntarily by individuals through a variety of means. Many of these users that employ these applications to gather geographic data have no formal geographic or geospatial training. Inaccuracies therefore are real concerns when using voluntarily collected data. But does the possibility of using tainted data outweigh what could be gained from a world-wide and ever-changing data source? We plan to perform research in order to help answer this question. Geographic points of people, infrastructure, and other sources can be locationally tracked or publicly conveyed by users of nearly all phone or computer applications, i.e. Facebook, Twitter, Yelp and Foursquare. When evaluating literature focused on VGI, most conclude that it is a valuable tool for research. "VGI has the potential to be a significant source of geographers' understanding of the surface of the earth" (Goodchild, 2007). Other studies have applied numerous methods to understand accuracy, "..relative, but not absolute, quality of data from different volunteers could be characterized accurately" (See, 2013). Researchers can use VGI to amass the knowledge of the public, who are often best positioned to provide information that requires indigenous experience, esoteric understanding of a physical environment, or up-to-date information about local conditions (Flanagin and Metzger 2008). We will attempt to replicate and extend the literature in this area by utilizing our own prior work on the development of CatMapper, a VGI application for both Android and iPhone. We will partner with the Audubon Society to act as a test end-user group for piloting prototypes and receiving feedback. We will use the analysis goals of this group – studying feral cat locations and restoring natural bird populations – as case for evaluating the appropriateness of the data collected. Analyzing data input, interviewing potential or current customers of the application, and field work to inspect coordinate outputs versus known benchmarks will all be practiced in the research process. We hope the application obtains accurate information that can serve as a representation of how we believe VGI networks can potentially play a more significant role in data collection for scientific research. The project we will undertake is important to the fields of GIScience, GIS programming through development of geographical mobile phone application, and the study of VGI by further evaluating its quality and value.

CHAPTER II

METHODS

In recent years enterprises have developed that have the ability to mine data from the public. Geo-tagged images, Facebook asking you to "check in", or Yelp asking for permission to use your current location are all examples of publically provided locational data. For this project we designed and developed our own data mining tool called the CatMapper. The CatMapper is a mobile application which is operable across mobile platforms. The purpose of this mobile application is to collect locational data of feral cats so that all who are interested may analyze their spatial distribution.

The CatMapper was developed for both Android and iPhone mobile devices. Both utilize the Esri software development kit (SDK). The Esri SDK allows the application to display a map and perform spatial analysis within it. Users are able to take a picture of a feral cat and input additional information such as if it appeared that the cat was being fed, how long they had seen the cat in the area, and the number of cats that were observed. This information is then sent to a database and hosted on an ArcServer where it is displayed on a web map. The application also has the ability to allow a user to draw a polygon on the map, known as a geofence, and be notified if a cat is reported within that geofence.

Once the CatMapper was released to our study group we were able to begin analysis on the quality of data which was collected. The web map and database were closely monitored for obscurities and curious entries. Coordinate points were compared to known benchmarks to determine accuracy. Other data such as cat images, number of cats etc., were closely monitored to ensure it was relevant data. At the end of our research we were able to calculate the maximum, minimum and average error of the points that users entered. Our application utilizes the mobile devices GPS capabilities. It is known that GPS units are prone to error and that higher quality GPS units are less prone to error and more accurate. What was not known, however; was the level of inaccuracy of the mobile phone's GPS. A benchmark, which is a standard measurement that provides the basis for comparison, was used to determine the inaccuracies of the GPS. Since the benchmark is a known location, we collected a series of waypoints with the mobile device directly above the benchmark. These waypoints were then averaged together. We then took the average waypoint and measured the distance between it and the benchmark. This measurement showed us the error of our mobile devices GPS. Because we have an interest in the development of GIS enabled mobile applications we needed input from our sample group on the design of the CatMapper. We were unable able to directly verify the accuracy of other data such as number of cats due to the dynamic nature of the data, but we did monitor uploaded images to ensure they were relevant and appropriate to the subject matter. The final purpose of the project then was to determine the user friendliness of the CatMapper. Surveys were sent out to the sample group which asked the users how user-friendly the CatMapper was, if they found it worth their time using, and if they would contribute again to research through VGI if asked to do so in the future or for their own projects. This survey helped us to determine how easy or difficult the application was to use, how relevant it was to the actual problem, and how well the users believed it performed the task that it was developed to do. This information will be valuable not only for future development of the CatMapper but also for the development of optimal VGI and GIS related mobile applications.

CHAPTER III

RESULTS

Once the application was developed to a point where it could be beta tested, we disseminated the downloadable version to our test user group, made up of twenty-five individuals. With this, the first challenge of testing our product was made apparent. Without an incentive, very few people took the time to download the application and even fewer used it once it was downloaded. Out of the twenty-five people that we sent the downloadable version of the application to, we only had six people test it. Despite the little amount of participation, we were still able to yield some results. From our test user group, we were able to test the CatMapper in Texas, California, and internationally in Australia. As the testing was in progress, we continually analyzed the web map in which the mobile application was posting to. When a new point was posted we were able to determine which user posted it, and after speaking with them, determine if the point was posted in the approximate location. Based on the points taken by our test-user group, the greatest amount of error was determined to be 24 meters from where the user took the point. The average error was 6.47 meters.

Once our user group had a sufficient amount of time testing the application, surveys were conducted to gain insight of the test-user's experience using the CatMapper. In the survey, participants were asked to rank statements based off of their experience on a scale of strongly agree to strongly disagree. These results are displayed in Figure 1 and Figure 2.

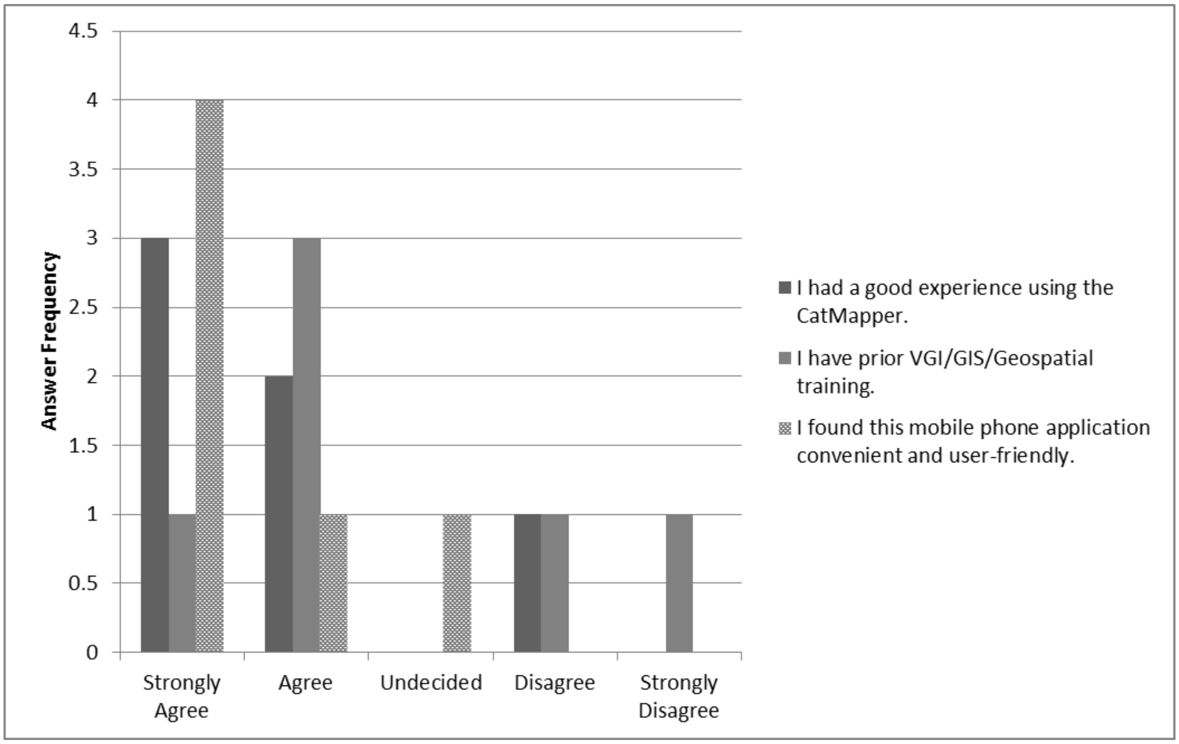

Figure 1. Survey Questions Part 1. This figure illustrates user's responses to questions concerning use of the CatMapper application.

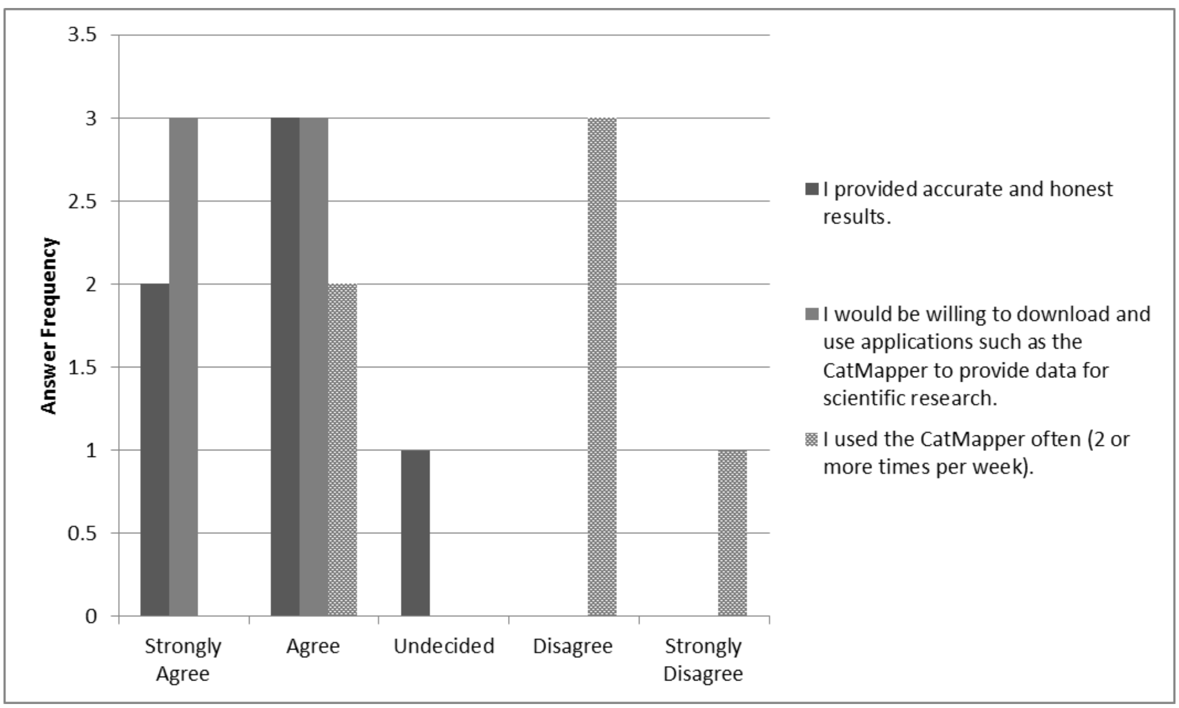

Figure 2. Survey Questions Part 2. This figure illustrates user's responses to questions concerning use of the CatMapper application.

The feedback that we received from this survey will help with the continued development of the CatMapper as well as future mobile applications. The participants of the survey and usage of the mobile phone application seemed to have a good experience overall using the CatMapper. Free-response comments provided both positive feedback and constructive criticism. For example, Respondent #1 claimed, “Keep up the good work! I really liked the ease of the program”, while criticism included, “The points that post to the map are very small. It is difficult to select them” (Respondent #2).

The final stage of testing for the CatMapper was to test the mobile device’s location service. To test this, we located a USGS benchmark on the Texas A&M University campus and recorded a series of sixty points with both the iPhone and the Android while the phone was sitting on top of the benchmark. The benchmark provided a constant and corrected position which gave us the ability to measure the distance from the benchmark to the collected points. When conducting this research we were aware of two apparent errors. The first of these errors is called multipath. Multipath is caused when the direct path from the satellite to the receiver is blocked by things such as buildings, trees, or your body. The satellite signals reflect off of these objects which results in a signal that has traveled farther to get to your receiver than it should have. This can result in the GPS miscalculating your position. The benchmark that we used happened to be located underneath a tree and 15 meters from a large building. We intentionally introduced this error into the research because it will be used highly in urban settings, therefore; we chose to simulate the environment in which the CatMapper would most likely be exposed to.

The second apparent error we were aware of is Positional Dilution of Precision (PDOP). PDOP indicates the accuracy of a GPS position based on the number of satellites visible to the receiver and the geometry of satellite position. A low PDOP results in greater accuracy while a higher PDOP results in greater inaccuracy. The level of PDOP is forecasted and made available to the public. By checking this forecast, an individual could plan to use his GPS at the most optimal time of day. For our case, however; we needed to test the application as if it were being used at various times with various PDOP conditions. This was done by determining when the PDOP was lowest on one day and when it was highest on the next and taking sample points on both days to average together. Figures 3 and 4 depict the forecasted PDOP for both days, represented by the green line. The first sample points were taken at 11:30 AM with a PDOP of 1.16 which is shown in Figure 3. The second set was taken at 5:45 PM with a PDOP of 1.49 which is shown in Figure 4.

PDOP for Day 1 (Green Line)

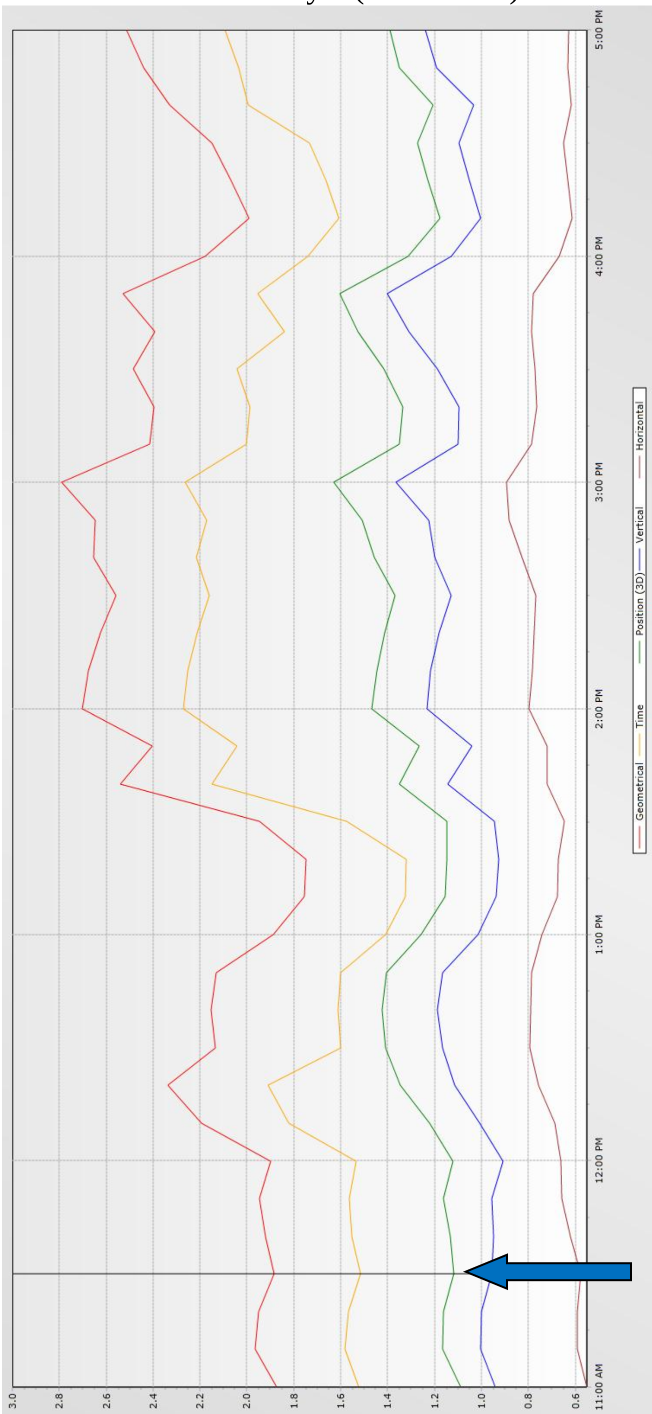

Figure 3. PDOP Error Analysis Day 1. This figure illustrates the GPS accuracy on Day 1 based on the Positional Dilution of Precision.

PDOP for Day 2 (Green Line)

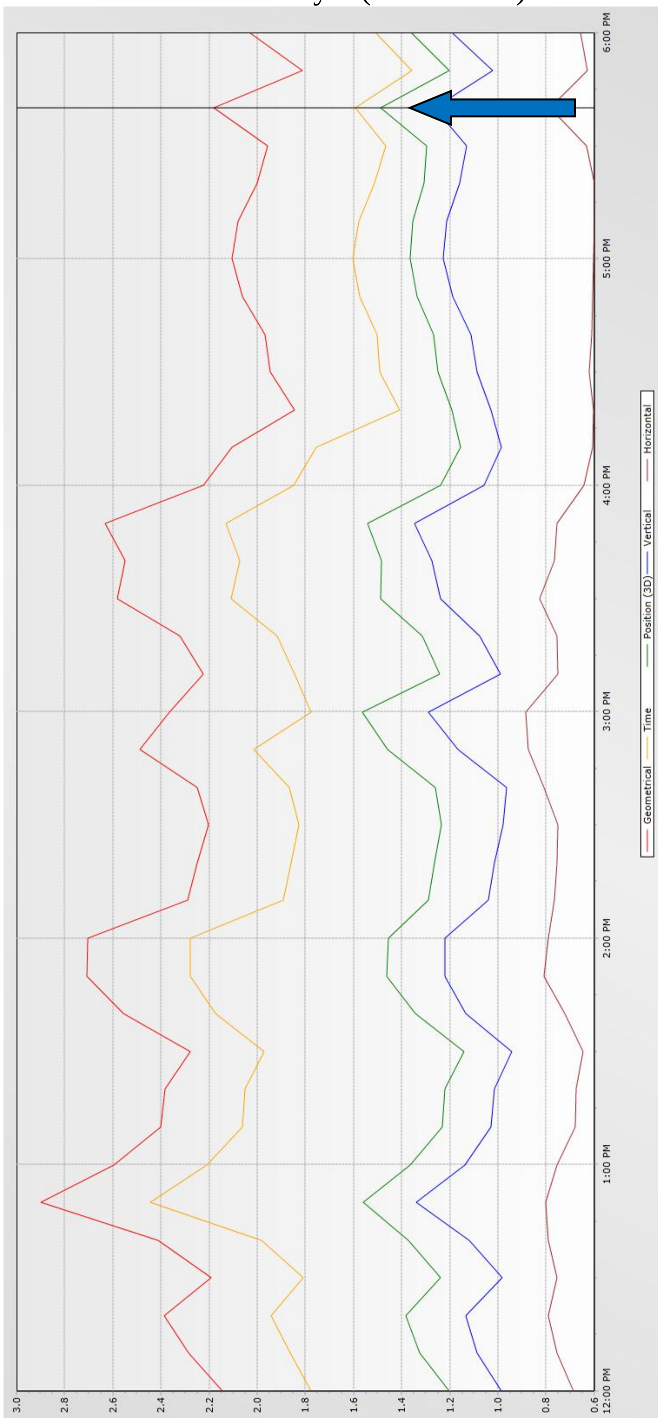

Figure 4. PDOP Error Analysis Day 2. This figure illustrates the GPS accuracy on Day 2 based on the Positional Dilution of Precision.

Once sixty data output points were gathered for both the iPhone and Android, each set was averaged. By averaging the latitude and longitude output this gave us the most statistically accurate point that could be obtained from the devices. We then measured the distance from the benchmark to the averaged points to obtain the amount of error that can be expected from the CatMapper. The distance from the benchmark to the iPhone average was 3.65 meters and the distance from the benchmark to the Android was 8.65 meters. Figure 5 shows the environment that the points were obtained in as well as the locations of the iPhone and Android averages relative to the benchmark.

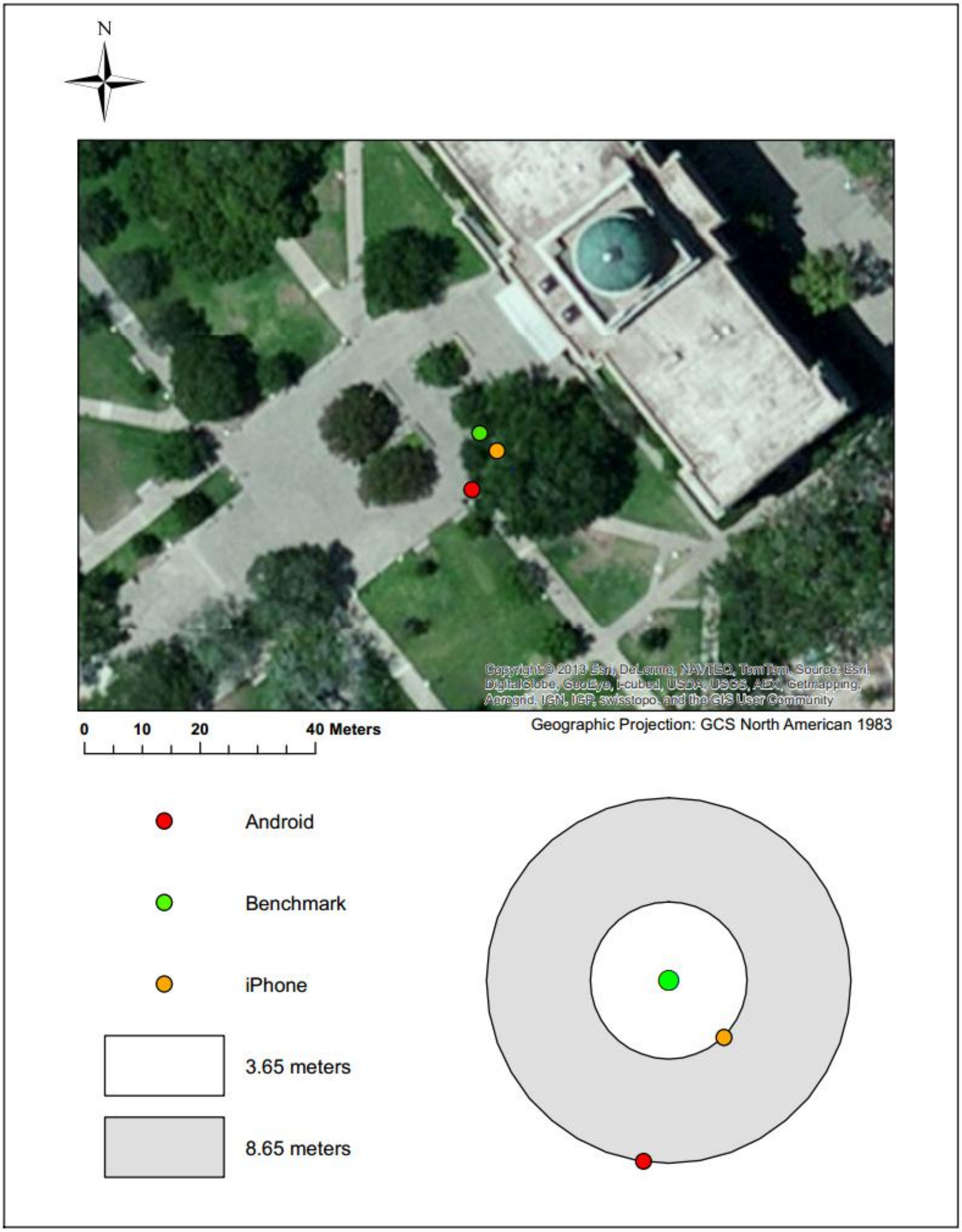

Figure 5. Locational Analysis of GPS points. This illustration shows the actual location of test points for both Android and iPhone in relation to a known benchmark.

CHAPTER IV

CONCLUSION

The conclusion of this project resulted in a functional mobile application intended for the purpose of mapping the density of feral cats. We aimed to create, test, and analyze the VGI that was collected from this mobile phone application to determine if it could provide accurate geospatial results for scientific research. With the convenience of VGI users worldwide, we suggested that this large data source has the potential to provide invaluable data, of good quality, for scientific research.

Our first results concluded that it can be difficult to persuade a substantial population of participants to gather data without some sort of incentive. This can propose that marketing may have to be implemented to reach populations if heavy geospatial data is needed for a project. We discovered this when we had not reached a test-group as large as we anticipated. It does not seem to be a concern of whether there are enough people out there employing VGI, but who would be willing to contribute to research.

Due to the minimal number of test-users, we were unable to confidently determine if the mobile application collects quality data. Because our user group was so small we were able to heavily monitor the points which were being posted to the map. In a situation where VGI has the potential to impact a real world event, the data would be inputted at a rate that would not allow for it to be screened as intensively as ours, resulting in lesser quality data entering the database. In this type of scenario, statistical analysis could be conducted to gain a more accurate understanding of VGI quality. Despite the lack of large scale quality testing, we were able to test the accuracy of the individual devices in the environment which they will be expected to operate in. Collecting a series of points with a variety of variable errors such as multipath, atmospheric error, and high PDOP, showed us how this application can be expected to perform in the field. This test leads to the conclusion that the mobile phone’s location service will not record the exact location of the user but will be within approximately 9 meters of that location. While this level of accuracy may not be suitable for sciences requiring absolute locational data, it is extremely valuable for the research that can utilize relative locational data.

This project should be continued and developed further. In the future the CatMapper should be distributed to a larger test-group so to increase the chances of introducing lesser quality data. A larger test-group would better simulate a real world situation.

REFERENCES

Connors, J. P., Lei, S., & Kelly, M. (2012). Citizen science in the age of neogeography: Utilizing volunteered geographic information for environmental monitoring. *Annals of the Association of American Geographers*, *102*(6), 1267-1289.

Foody, G. M., See, L., Fritz, S., Van der Velde, M., Perger, C., Schill, C., & Boyd, D. S. (2013). Assessing the Accuracy of Volunteered Geographic Information arising from Multiple Contributors to an Internet Based Collaborative Project. *Transactions in GIS*.

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, *69*(4), 211-221.