**Title**: Multi-Method Approach to Geocoding Point Features in GIS.

R. Baffour, J. Bohn, P. Digirolamo, A. Kalinski, K. Russell, J. Cleaves.

**Abstract**: Advances in remote sensing technologies together with better street databases have made geocoding of community facilities simpler and cost effective. Traditional methods of capturing point data for GIS applications have relied mostly on GPS. In updating community facilities database, the Atlanta Regional Commission in Atlanta, Georgia utilized a multi source approach to minimize data acquisition time and cost while maintaining the accuracy of the data. This paper addresses the use of multiple methods of address matching, DLGF, and GPS to update the database. The paper also presents a simple statistical analysis conducted to check the reliability of the methods used.

**Background.** Developing a GIS databases to support community planning activities has become very popular with the increasing knowledge of GIS and also with the ease of using GIS software such as ARCVIEW. Many communities are using GIS as a platform to present data and also to support various planning activities such as landuse planning, community development projects such as zoning, and other neighborhood initiative projects such as crime analysis and reporting. In most cases, manipulating the data is not the critical problem but obtaining the data and also getting the data at the desire accuracy.

There are all kinds of data collection methods used in gathering data for GIS application development. Such methods include the use of remote sensing images such as NASA's Landsat images, the use of digital aerial photographs, the use of GPS especially for point and line features, and the use of GIS processing modules such as address geocoding or address matching. To determine the best data collection methods to adopt for a project, it is always important that the applications' data requirements are very well understood. Data requirement such as accuracy determines the methods used to gather data and ultimately contribute significantly to the overall cost of the project.

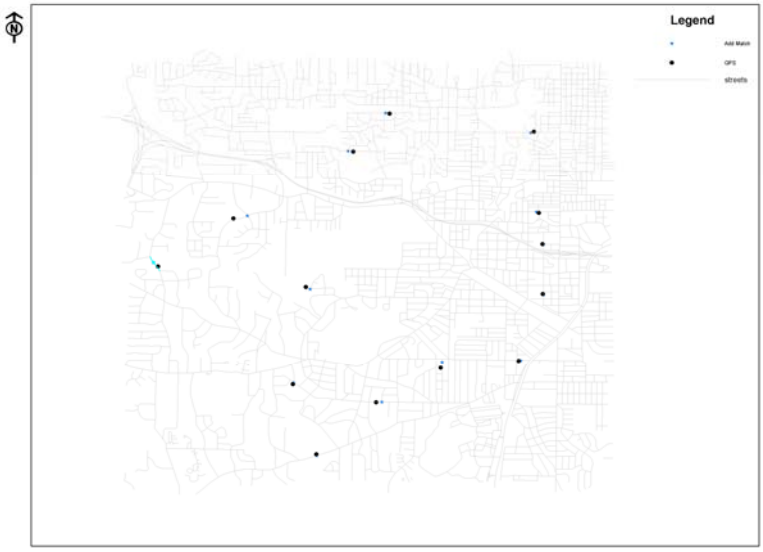
In other to provide information on community facilities such as hospitals, schools, daycare centers, Fire stations, and police facilities to the planning community, the Georgia Department of Community Affairs (DCA), working through such agencies as the Atlanta Regional Commission (ARC) collected and developed a GIS database of these facilities in 1996. A GPS survey was used to collect the locations of these facilities together with associated attributes.

In the summer of 2002, DCA decided to update the database. Updating the database meant adding new locations and also checking whether old facilities were still in place and performing the same functions. The challenge was to determine the best approach to use for this project.

The objective of this paper is to present some of the methods used in the project and also present some basic comparisons on the relative accuracies of some of the key methods used.

**Methods Used**. After series of discussions, two data collection methods were adopted for the project. These methods are; 1) Address matching, and 2) using GPS surveys. The following sections present how the methods described above were used in the project.

***Using Address Matching***: Almost all the features in the database together with the new ones to be added have some sort of address. Address matching was therefore an obvious choice for most of these locations. Since the old database was created with GPS survey to an accuracy of 3 feet, DCA wanted to make sure any methods used will not degrade the accuracy of the original database. Judging from the needs of the data and the accuracy of the base maps, it was obvious that address marching could be useful especially when one considers the cost of doing a GPS survey for the points being considered. To convince ourselves that address geocoding will not significantly affect the accuracy of the data, we conducted a survey of a few points using GPS. We also used address matching to geocode all the points surveyed. Figure 1 is a map showing the relative positions of the points using the two methods of address matching and GPS. Even on such a large scale map of 1 inch to 4000 feet, one can see that the two methods are very agreeable.




1in to 4000 feet

Figure 1: Address Matching and GPS Survey Comparison

A quantitative approach to get a good relationship between the two methods was also conducted. Table 1 below depicts the GPS and Geocoding (GC) coordinates and the relative distances computed between the GPS points and the address matching points.

Table 1: Address Matching and GPS Survey Comparison

| FAC_ID | GPS X (ft) | GPS Y(ft) | GC X (ft) | GC Y (ft) | Distance (ft) |
|---|---|---|---|---|---|
| ET1 | 2210929.11 | 1282655.38 | 2211066.50 | 1282713.28 | 149.09 |
| ET4 | 2185126.72 | 1513858.68 | 2186571.65 | 1512927.13 | 1719.19 |
| ET6 | 2286864.41 | 1381900.52 | 2285896.68 | 1382362.27 | 1072.25 |
| TP1 | 2206955.92 | 1347968.24 | 2206968.49 | 1347894.74 | 74.57 |
| TP10 | 2205683.41 | 1352541.30 | 2205714.94 | 1352658.09 | 120.96 |
| TP2 | 2209031.08 | 1367860.57 | 2208744.11 | 1367928.18 | 294.82 |
| TP20 | 2218935.68 | 1369141.03 | 2218760.83 | 1369102.15 | 179.12 |
| TP23 | 2221868.05 | 1368297.08 | 2221827.13 | 1368108.34 | 193.13 |
| TP27 | 2219180.92 | 1363785.16 | 2219056.70 | 1363844.71 | 137.76 |
| TP3 | 2213791.19 | 1353605.06 | 2213852.86 | 1354000.55 | 400.27 |
| TP34 | 2202440.97 | 1363455.17 | 2203220.65 | 1363642.28 | 801.82 |
| TP36 | 2218049.05 | 1354056.23 | 2218166.97 | 1354100.47 | 125.95 |
| TP37 | 2219382.31 | 1361732.64 | 2219423.07 | 1361724.01 | 41.66 |
| TP41 | 2210232.27 | 1351342.28 | 2210539.67 | 1351375.61 | 309.20 |
| TP44 | 2198301.27 | 1360323.15 | 2198050.60 | 1360588.52 | 365.05 |
| TP45 | 2211023.72 | 1370355.59 | 2210808.15 | 1370404.05 | 220.95 |
| TP46 | 2196157.18 | 1518033.64 | 2195782.07 | 1518287.12 | 452.72 |
| TP48 | 2211659.09 | 1254313.88 | 2212162.24 | 1255110.55 | 942.26 |
| TP49 | 2201665.20 | 1255413.91 | 2201077.88 | 1255599.42 | 615.92 |
| TP51 | 2221990.25 | 1281104.09 | 2221877.50 | 1281190.90 | 142.30 |
| TP51B | 2222009.83 | 1281311.88 | 2221792.54 | 1281199.01 | 244.86 |
| TP52 | 2206094.34 | 1290764.57 | 2206868.00 | 1290148.36 | 989.07 |
| TP53 | 2166818.47 | 1333307.96 | 2166279.62 | 1333792.16 | 724.44 |
| TP56 | 2260828.51 | 1348073.22 | 2261017.34 | 1348015.86 | 197.35 |
| TP57 | 2262454.59 | 1356011.59 | 2262300.69 | 1356082.28 | 169.36 |
| TP58 | 2261725.95 | 1357551.25 | 2261688.83 | 1357501.27 | 62.26 |
| TP59 | 2275285.10 | 1381160.73 | 2275312.12 | 1381339.85 | 181.15 |
| TP6 | 2206399.18 | 1358959.96 | 2206639.81 | 1358833.98 | 271.61 |
| TP60 | 2278756.46 | 1385827.33 | 2278879.70 | 1386177.82 | 371.53 |
| TP61 | 2284914.27 | 1385518.11 | 2284698.03 | 1386319.77 | 830.32 |
| TP62 | 2288134.89 | 1376126.52 | 2288284.01 | 1376288.79 | 220.39 |
| TP63 | 2286957.74 | 1377962.12 | 2286860.88 | 1378228.92 | 283.84 |
| TP64 | 2283812.18 | 1377200.24 | 2283865.54 | 1377435.29 | 241.03 |
| TP65 | 2283110.21 | 1374787.92 | 2283108.90 | 1374728.60 | 59.34 |
| TP66 | 2277788.41 | 1374864.17 | 2278233.10 | 1374787.12 | 451.32 |
| TP67 | 2310999.57 | 1348848.13 | 2311307.55 | 1348995.21 | 341.30 |
| TP69 | 2295529.62 | 1356294.58 | 2296504.32 | 1355577.68 | 1209.95 |
| TP70 | 2291874.12 | 1354735.33 | 2291965.51 | 1354699.56 | 98.14 |
| TP8 | 2219391.91 | 1358465.92 | 2219421.27 | 1358419.03 | 55.32 |
| TP99 | 2259417.66 | 1356040.13 | 2257824.04 | 1355354.04 | 1735.04 |

An average distance calculated was 428 feet with the minimum and maximum being 42 feet and 1735 feet respectively. Based on this simple analysis, cost, time and also based on the fact that there is usually one of such facilities in a neighborhood, we concluded that address matching will be good enough in creating the required database. Address matching was therefore used for all addresses that could match with a t least a 70% score. In matching addresses, two different databases were used: the conventional street map of the Atlanta area that has been edited by the Atlanta Regional Commission and the Georgia Department of Transportation GIS transportation network layer (GLDF) and the associated road characteristics file (RC).

***Using GPS***. Facilities that had P O Box addresses and those that did not score well in the address matching process were surveyed using GPS.

**Conclusion**. This project demonstrated the ability of using multiple methods of acquiring GIS data to keep project cost and time at a minimum. We cut the GPS survey time by more that 80% which allowed a lot of the work to be done in house. In essence less than 20% of the job was done with field work. It is also very apparent that for planning purposes, GPS and address matching are all very useful in determining the positions of points of interest.

**Acknowledgement.** We would like to acknowledge the Georgia Department of Community Affairs (DCA) for their support and also the staff of the GIS section of the Atlanta Regional Commission.

**Authors' Information**. Robert Baffour-Associate Professor, Department of Engineering, Clark Atlanta University, 223 James P. Brawley Dr. SW, Atlanta, GA 30314. Jim Bohn, Paul Digirolamo, Art Kalinski-GIS Section, Atlanta Regional Commission, Kirkwood Russell and Jasmyn Cleaves, Undergraduate Students, Department of Engineering, Clark Atlanta university.