Localization and 3D Reconstruction of Urban Scenes Using GPS

Kihwan Kim, Jay Summet, Thad Starner, Daniel Ashbrook, Mrunal Kapade and Irfan Essa
College of Computing, Georgia Institute of Technology
{kihwan23,summetj,anjiro,thad,mrunal,irfan}@cc.gatech.edu

Abstract

Using off-the-shelf Global Positioning System (GPS) units, we reconstruct buildings in 3D by exploiting the reduction in signal to noise ratio (SNR) that occurs when the buildings obstruct the line-of-sight between the moving units and the orbiting satellites. We measure the size and height of skyscrapers as well as automatically constructing a density map representing the location of multiple buildings in an urban landscape. If deployed on a large scale, via a cellular service provider's GPS-enabled mobile phones or GPS-tracked delivery vehicles, the system could provide an inexpensive means of continuously creating and updating 3D maps of urban environments.

1. Introduction

3D data sets of major buildings in urban areas are produced and sold commercially. These data sets are constructed by using expensive aerial photography, purchasing access to building permit and architectural plan databases, and performing street level laser surveying. Perhaps the most widely known use of these data sets are in GIS visualization tools such as the popular Google Earth. Users of these visualization tools have discovered many types of new applications of such data sets such as community planning, virtual tourism, education, and augmented reality. Similar commercial tools are used for urban planning, inspection of utility lines, humanitarian aid, emergency relief, and traffic monitoring. 3D models of buildings in cities can also be used to improve GPS-based location systems by modeling the effects of urban canyons on signal propagation [2]. Here, we present an inexpensive way to produce 3D building data sets using only GPS signals.

Private individuals are collecting GPS information every day. Linked by the internet, groups of cooperating users have already begun to share GPS data for the purpose of creating free open-source maps [1]. The advantage of worn GPS enabled devices (e.g. mobile phones) is that the data collection can include relatively fine features such as

(a)   (b)   (c)

(d)   (e)   (f)
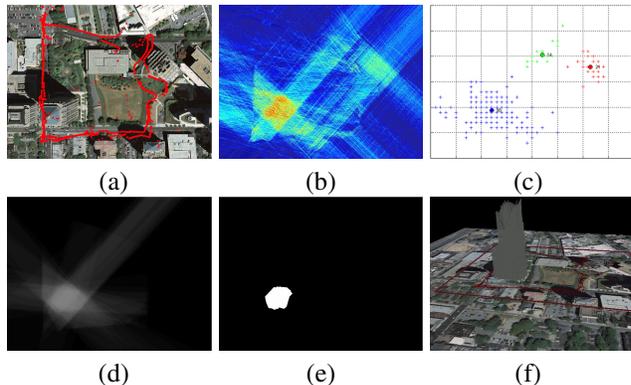
Figure 1. Building density maps from GPS data and clustering to find building centers (a-c), estimating building region and volume from the dominant cluster (d-f).

walkways between buildings, footpaths in parks, and entrances to buildings. Here, we demonstrate that the addition of 3D building data-sets to these already existing efforts would only require a minimal amount of additional computation. Our technique requires many trips to collect data, and we imagine its main advantage is supporting such "crowdsourcing" or open-source data collection system. However, a motivated mobile phone service provider could anonymize the location information provided by their users' GPS-enabled mobile phones to create a proprietary 3D map that is updated daily as its users move about a city. Fleets of vehicles equipped with GPS receivers are already driving routes that cover the majority of urban cities. Any major delivery company, car rental company, taxi service, or corporate fleet management office already has access to GPS data that covers major cities. With the technique described in this paper, that data might be extended to include 3D models of the urban centers such companies service.

In this paper we will demonstrate the feasibility of using GPS signals to determine the location, size, and height of buildings using GPS signals. Our approach takes advantage of the fact that when a building obstructs the line-of-sight between a satellite and a GPS receiver, it causes a de-

tectable drop in the signal to noise ratio (SNR). For example, Figure 2 shows two traces from a GPS receiver in motion around a skyscraper. As the line-of-sight between the GPS receiver and a satellite was obstructed by the building the SNR is significantly reduced. Our technique combines data samples from many GPS positions to detect buildings and determine their location, size, and height.

Our approach has two main stages: (1) making density maps to detect multiple buildings in the test site, (2) estimating the region and size of a specific building and generating a 3D reconstruction. Figure 1 shows graphical output from several stages of our approach. Figure 1(a) is an overhead view of one of our testing sites. The red paths show GPS ground tracks where data was collected. Using multiple samples, we first generate a density map showing the probability that each location is a building obstructing GPS signals (Figure 1(b)). Once the density map is determined, we apply mean shift clustering [5] to cluster and find peak points of each cluster (Figure 1(c)) corresponding to the center of our detected buildings. To measure the largest building, we choose the dominant cluster (blue area in Figure 1(c)) and then apply a threshold to estimate the building region. See Figure 1(d&e). Finally, we use the estimated building region to reconstruct the volume of the building (Figure 1(f)) using a voxel rendering algorithm.

The structure of the paper is as follows: Section 2 explains our data collection procedure. In Section 3, we show how we create the density maps. In Section 4, we describe how to estimate the region of the dominant building and use this region to produce a building volume and height estimate via modified voxel carving. In Section 5, we will summarize the quality of our results and show how increasing the number of GPS sample points affect accuracy. In Section 6, we summarize related work.

2. Data Collection and Feasibility

We initially collected data by walking at a slower than normal pace around the Bank of America (BOA) Plaza in Atlanta, GA, a 55 story skyscraper that is 1023 ft (311.8m) tall. We used two different off-the-shelf consumer grade GPS receivers: a Garmin GPSmap-60CSx and a Garmin GPS 35-LVS. The devices were carried mounted on a wooden support above a backpack. By walking a complete circuit around the building on three days we collected NMEA data samples at a rate of one sample per second which we cropped to use exactly 4000 samples.

After developing our technique, we also collected data from around a different Atlanta skyscraper, the One Atlantic Center (OAC) building, and verified that the same technique worked with a second set of data gathered at a different location. We gathered data on two days using the Garmin GPS 35-LVS (walking two circuits around the building on each
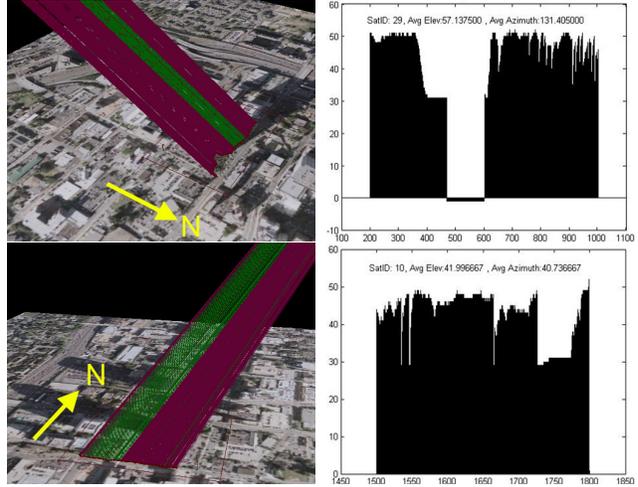

Figure 2. Example of SNR reduction due to the obstruction of a building. Green indicates low SNR.
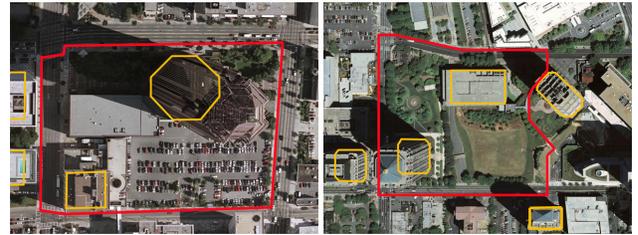
Figure 3. Top view of testing environments.

day) to reach 4000 samples. Unlike the slow pace of data sampling for the Bank of America building, when circling the One Atlantic Center building the researcher walked at a medium to fast walking pace, and each circuit took just over fifteen minutes.

Figure 3 shows an overhead view of each data collection site, with the path taken by the researcher marked in red. The researcher walked on public sidewalks around the buildings. All buildings higher than five stories are marked with orange lines. Note that due to warping of the overhead photography, the base of the building (which we use as the ground truth) appears at a slightly different location than the top.

To demonstrate initial feasibility of our technique, we present the following evidence. Figure 2 demonstrates the SNR values for two satellites as the researcher walks along two sides of the Bank of America (BOA) building, occluding the direct line-of-sight path to the satellites with the building. We observe that the SNR drops dramatically when the vector from the GPS receiver to the satellite intersects the target building. These results demonstrate that our assumption –occlusions between the GPS receiver and satel-
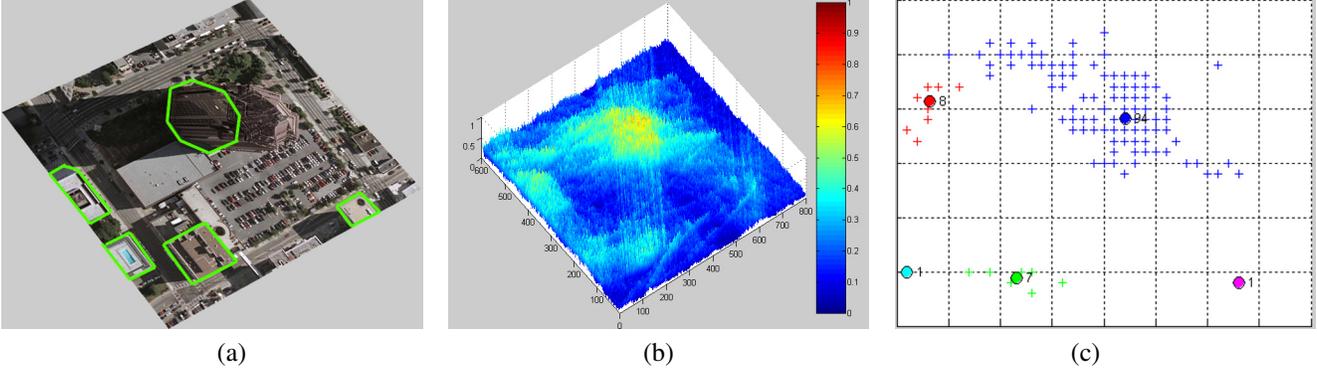
Figure 4. (a) BOA testing site, (b) generated density map, (c) clustering to detect building centers.

lite will lower the SNR– is valid.

Note that for a GPS receiver to achieve a positional fix, it must have a reasonable signal from at least four satellites. In many conditions, GPS receivers detect signals from more than four satellites but do not use the data from satellites with very weak or reflected signals. Our technique makes use of these signals as an indication that the direct line-of-sight path to that satellite may be occluded.

So far, we have ignored the fact that vectors between a GPS receiver and each satellite have both an azimuth and elevation. Some signals (to satellites that are on or near the horizon) have an elevation that is too low to be of use. For example, a vector with an extremely low elevation could be occluded by a four story building several miles away, but we have no way to determine the distance to the occluding object. In addition, signals with low elevations tend to have much more variability in their SNR's, as signals tend to be reflected more near the ground. We found that ignoring all vectors with an elevation below 15 degrees improved our results significantly.

Our techniques make use of statistical averages, so minor errors in GPS position do not greatly harm our results. Our GPS receivers did occasionally drift from the ground-truth path of travel (see traces in Figure 9(a)&(c)). Limited baselines and multi-path reflections reduce the accuracy of GPS units in urban canyons such as our test sites. Even in optimal conditions, consumer GPS units can be off by several meters. Various methods have been demonstrated to use inertial tracking combined with GPS readings to generate a more accurate position [4] which could slightly improve our results.

3   Detecting Multiple Buildings

In this section, we demonstrate a method to calculate a *density map* of each site and decide the location of dominant buildings within the testing site. Each NMEA sample collected has multiple signals (each representing a vector to a specific satellite's elevation and azimuth from the current

position of the GPS receiver) that have different SNR values. We convert the vector data into a 2D density map by projecting each vector onto the ground plane. Each vector that has a weak SNR value (possibly due to an occlusion) contributes to the values along a line on the ground plane. Multiple "occluded" vectors passing through the same volume have lines that pass through the same space on the 2D ground plane, and, similar to a Hough transform, probability values combine at the true locations of occlusions while noise values are randomly distributed.

The proceeding overview of our method ignores several key points. Obviously, as we are looking for weak signals (signifying occlusions), we invert the SNR values when adding it to the accumulation buffer. Also, SNRs vary between satellites based upon both their ID and distance. If we add all signal vectors into the buffer directly, the results are biased towards satellites that have overall weaker signals. Additionally, vectors from satellites that are projected onto lines close to the path taken by the GPS receiver are combined with similar vectors from the same satellite multiple times. Without normalization, this tends to generate artifacts based upon the path taken by the GPS receiver.

To compensate for satellites with different signal levels and the path taken by the GPS receiver, we first calculate the maximum, minimum and average SNR values of all signals from each satellite in the data set. We define $S_j^{\max}$, $S_j^{\min}$ and S_j^{avg} as the maximum, minimum and average SNR respectively of the *j th* satellite. SNR values of individual vectors are labeled as V_{ij} which is the SNR at the *ith* sample along the traveled path to the *jth* satellite. Note that all vectors from the *ith* position share the same real-world origin $i|_{x,y}$, which is determined by the position of the GPS receiver in real-world coordinates. These x, y coordinates, combined with the vectors' azimuth and elevation, determine the projection of the vector into the accumulation buffer.

At every point in the accumulation buffer, where the vector is projected onto the ground plane, we add the value $L(i, j)$. It is calculated from all samples that fall below the

(a) (b)

(c) (d)
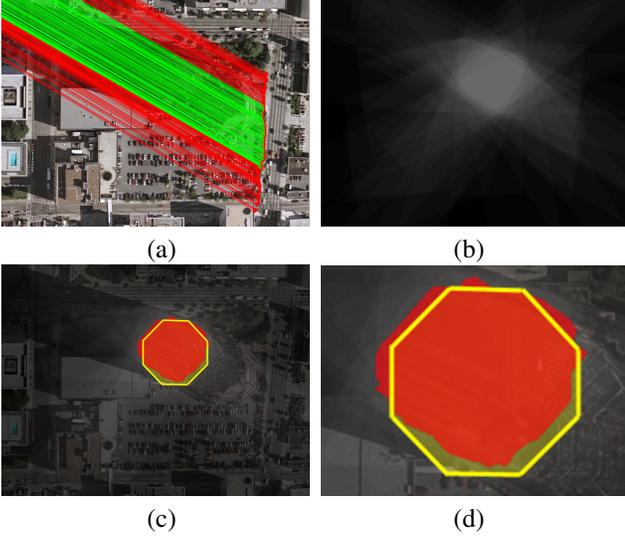
Figure 5. Estimating the region of the BOA building using accumulation of multiple occluded vectors.

S_j^{avg} as follows:

$$L(i,j) = \begin{cases} \left| \frac{S_j^{\max} - V_{ij}}{S_j^{\max} - S_j^{\min}} \right| & , \text{if } V_{ij} \leq S_j^{avg} \\ 0 & , \text{if } V_{ij} > S_j^{avg} \end{cases}$$

Once we process all signals over every *i* and *j*, we normalize the accumulation buffer entries from 0 to 1. Now, the accumulation buffer becomes the *density map*.

Figure 4 shows an overhead view of the Bank of America testing site and the density map generated using our technique. Buildings higher than five stories are marked with green lines in Figure 4(a). The density map shows all five tall buildings with higher than 50% probability. Figure 4(c) shows the clustered result. We applied the mean shift algorithm [5] to find the peak points. These clusters identify candidate regions where buildings are likely to exist. We use the dominant cluster (the blue cluster, which has probabilities larger than 60% on average) as the target building to reconstruct.

4   Building Reconstruction

In the previous section, we automatically found a dominant cluster which indicates the presence of a building. In this section, we will demonstrate that by using multiple signal vectors passing near and through the dominant cluster target region we can determine the approximate 3D size and shape of the building.

As an example of how we estimate the building region, Figure 5(a) shows "occluded" vectors (in green) from a sin-


(a)

(b) (c)
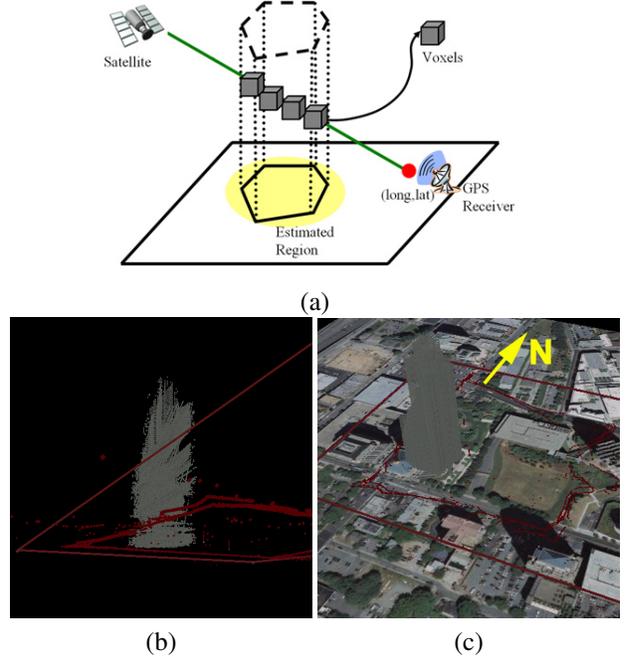
Figure 6. Building 3D models using voxels.

gle satellite that pass through a dominant cluster. First we select all vectors that pass through the cluster. Of these vectors, the occluded vectors are those whose SNR falls under a manually selected threshold. By combining all occluded vectors from multiple satellites (Figure 5(b)) we can determine the 2D region of the building. Because each satellite produces slices of data at different heights in 3D space, and this simplification is being projected onto a 2D ground plane, it helps to use multiple slices that pass through the area of interest at close to the same height as possible, or not require a complete boolean union to select your estimated region. We determine our final building region estimate by selecting the top 10% of the accumulation buffer.

Our result, denoted in red in Figure 5(c&d), shows our method's estimate of the BOA building's region. The ground truth region is denoted with yellow lines. In this results the difference between the center point of the ground truth region and that of our estimate is only 12 ft (4m). To evaluate our method quantitatively, we calculated the error between our results and ground truth by measuring area difference.

Although visually our result is quite impressive, the actual area difference reports both false-positive regions outside of the ground truth region, and false-negative areas inside of the ground truth region. This result still represents a 23% error in region match. We will analyze the effect that using fewer data samples has on estimation accuracy in Section 5.

Once we estimate the region of a building we can gen-
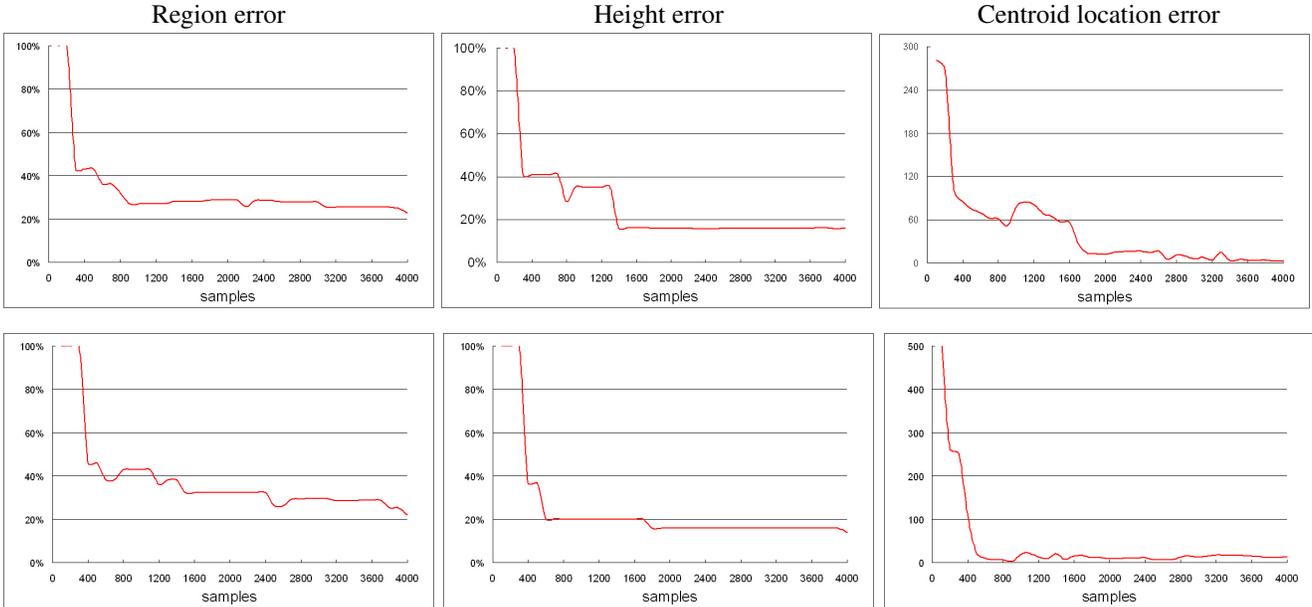
Figure 7. Error metrics as the number of samples increase. (Top: BOA and Bottom: OAC.)

erate its volume based on the collected vectors and their SNRs. Although each vector does not provide range information (the occluder can be anywhere between the satellite and GPS receiver) the estimated region can be used to determine the most likely position of the occluding buildings. Thus, we can reconstruct the building by constructing a 3D volume where the occluded vectors pass over the estimated region. This relies on the assumption that the estimated region on the ground plane accurately reflects the overall shape of the building. When calculating the building region estimate, we use a combination of signals that pass by the building at different heights. This results in a region estimate that is an average of multiple cross sections of the building. Using enough data, it should be possible to calculated separate region estimates at different heights of the building.

Sophisticated volume rendering algorithms exist which would improve our results ([13, 6, 8]), but their implementation is not only outside of the scope of this proof-of-concept research but would also require attenuation or depth information. Thus, we reconstruct a 3D model of the building by using a grid-based voxelization algorithm [15]. As shown in Figure 6, we render voxels that intersect with an occluded vector when they pass above the estimated region.

So far our work has used the law of large numbers to average data samples, removing noise and reducing the effect of outliers. When using individual vectors to voxelize buildings, near-vertical outliers adversely affect the height estimate. To avoid this problem, we exclude outliers near the top of the building by checking for agreement between at least 10 vectors at each 30 ft increment. This

Site	Centroid dist(ft)	Region err(%)	Real height(ft)	Estimated height(ft)	Height err(%)
BOA	7.398	22.73	1023	862.80	15.65
OAC	22.59	22.20	820	705.32	13.98

Table 1. Error results using 4000 samples from each site.

procedure tends to underestimate the height of the building slightly, however, not removing outliers sometimes resulted in a massive height over-estimate.

Figure 6(b) shows the result of voxel reconstruction. In Figure 6(c) we rendered a building object by filling in all voxels that fall under a detected voxel, that is, we assume buildings have no overhangs. We use this reconstructed building model to estimate the height of the target buildings, as reported in Section 5.

5 Results and Error Analysis

Our approach produces reasonably accurate results using only 4000 sample points (approximately sixty-six minutes of data collection). Table 1 shows the final result using 4000 samples for both sites. We have determined the building centroids to within 7 and 23 ft when compared to commercially available data sets. Region error approaches 22%, and as shown in Figure 5(d), the visual region is quite close to our comparison data set. Although rough, the automatically generated 3D building model gives a good representation of the building volume and height.

Although the accuracy of building reconstruction may be improved by the use of more sophisticated radio-wave prop-
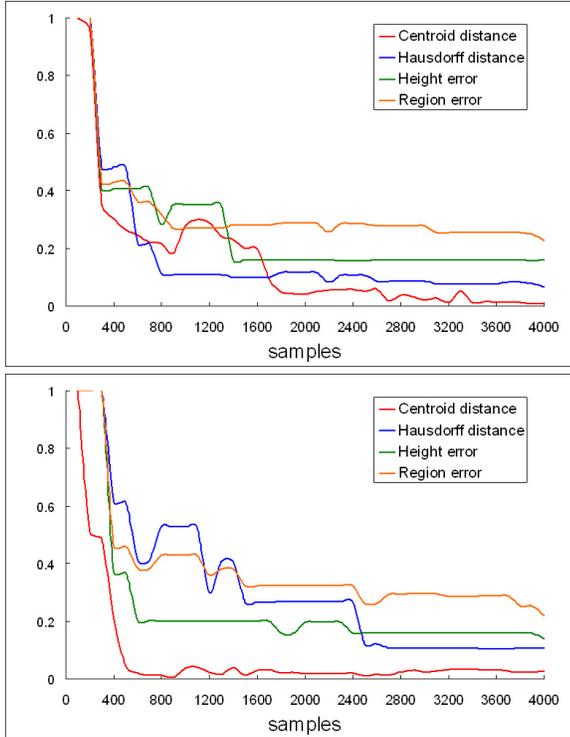
Figure 8. Normalized error metrics as the number of samples increase. (Top: BOA and Bottom: OAC).

agation models [19] or better algorithms, it does not appear that our current method can be improved upon by collecting more data samples. The graphs in Figure 7 show the behavior of our error metrics (estimated building centroid, building region estimation, and building height estimation) with respect to the number of samples collected. The results clearly show that collecting more data samples markedly improves the results from zero to 1000 samples. An additional 1000 samples (2000 in total) can double the accuracy. Two-thousand samples can be collected in 33 minutes with a consumer GPS device, and in our case, consisted of two circuits around the surveyed area. By the time 2000 samples are used, most results have converged, and additional data produces diminishing returns.

Obviously, the geographical coverage of data samples also plays an important role. In our data collection, 1000 samples corresponds to a complete circuit by the researcher. It may be that a smaller number of samples equally spaced along a complete circuit of the area/building under survey will provide equivalent results. Our density map, region estimation, and voxel filling approaches all require widely spaced data around the area/building under examination. [1]

---

[1] We observed that the cluster centroid errors improved dramatically near the 400-500 sample range, corresponding to one-half of a circuit.

Luckily, collecting data from all sides of most large buildings is not usually a problem.

In addition to the error metrics reported above, we also measured region similarity by a modified hausdorff distance [7], a relative distance metric commonly using in matching problems. Figure 8 shows all error metrics overlaid for direct comparison, normalized on a 0 to 1 scale. Figure 10 graphically shows the rendered volumes as the number of data samples is increased.

6 Related Work

GPS signals have a long history of being used to detect user motion and make predictions of their goals and context. Additionally, the lack of GPS signals has been used to infer the existence (but not location or size) of tall buildings [12] and as an indoor context detector [3]. Weiss et al. used existing environmental 3D models to predict multipath reflections, but did not use GPS signals to produce 3D models [19].

In addition to passive stereo vision techniques, robotics researchers have used active ultrasonic and laser rangefinding techniques to build 3D environmental models [14, 10]. These techniques require expensive equipment and active transmission of ranging pulses, which require additional power and could attract unwanted attention for some applications.

Volumetric ray-tracing techniques similar to our approach but using ultrasonics have been used on a small scale with ceiling mounted ultrasonic transmitters and a mobile robot based receiver to build a room-scale environmental model [9]. Volume rendering using data from CT scanners and ultra-sound is widely studied in computer graphics and diverse research efforts already have been proposed [6, 8, 13].

GPS signal data has been used to monitor large scale global atmospheric effects. Atmospheric researchers have used global positioning system radio occlusion (GPS RO) techniques to measure global attributes of the upper atmosphere. The CHAMP satellite uses a GPS receiver that detects radio occlusion to measure temperature and water vapor profiles in the troposphere[16, 11]. Ground based GPS stations have been used to determine electron density fields of the ionosphere[17].

In the only similar work using GPS signals for building detection, Swinford experimented with using nonavailability of expected GPS signals to suggest intervening obstacles in the environment. However, this work used GPS data only to add height data to building outlines derived from cartographic maps. Our approach uses only GPS

From this result we can infer the minimum required trip for estimating the centroid of a single building or obstruction could be at least half of a cycle around it.

data to detect and measure buildings and does not rely on pre-existing cartographic data. Additionally, Swinford did not present measurements of the accuracy he obtained [18].

7 Conclusion and Future Work

We have shown that a standard GPS receiver can detect and localize buildings by measuring reduction of SNR caused when an object comes between the receiver and one or more satellites. We introduced an approach that can estimate the region of a target building and reconstruct its location, volume and height using only GPS signals. Our approach generated reasonable estimates with around 14-22% errors in region and height. We also demonstrated a method to calculate a density map that predicts where buildings exist in a given area.

This approach is well suited for swarm or crowd-sourcing applications, where multiple cooperative agents roam throughout a space sharing GPS data to generate an overview of buildings and other large objects. Even though our approach is less accurate than carefully calibrated computer vision or laser scanning, it has the advantage that GPS receivers are passive, do not require active aiming, automatically self-calibrate, and are inexpensive. Although we have demonstrated the technique using person carried GPS devices, we anticipate that vehicle mounted GPS units could also be used to successfully collect data.

Our future work will investigate more sophisticated signal processing and extend the use of density maps for building detection to larger scale areas. We hope to automate all phases of our method to produce a fully automated process that will identify likely locations of buildings and then measure them individually without human assistance.

Acknowlegements

References

[1] OpenStreetMap: http://www.OpenStreetMap.org.

[2] Pole Star NAOCity: http://www.polestar-corporate.com/NAOCity.pdf.

[3] D. Ashbrook and T. Starner. Using GPS to learn significant locations and predict movement across multiple users. *Personal Ubiquitous Computing*, 7(5):275–286, 2003.

[4] S. Beauregard and H. Haas. Pedestrian dead reckoning: A basis for personal positioning. In *In Proc' of the Workshop on Positioning, Navigation and Communication (WPNC'06)*, pages 27–36, 2006.

[5] Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE Pattern Analysis and Machine Intelligence*, 17(5):790–799, 1995.

[6] R. A. Drebin, L. Carpenter, and P. Hanrahan. Volume rendering. In *SIGGRAPH '88*, pages 65–74, New York, NY, USA, 1988. ACM Press.

[7] M.-P. Dubuisson and A. K. Jain. A modified hausdorff distance for object matching. In *Proc. 12th Internatioanl Conference on Pattern Recognition*, pages 566–569. IEEE, 1994.

[8] T. T. Elvins. A survey of algorithms for volume visualization. *SIGGRAPH 92*, 26(3):194–201, 1992.

[9] R. K. Harle and A. Hopper. Building world models by ray-tracing within ceiling-mounted positioning systems. In *In Proc' of Ubicomp*, pages 1–17, 2003.

[10] M. Hebert. Outdoor scene analysis using range data. In *Proceedings of IEEE International Conference on Robotics and Automation.*, pages 1426–1432, 1985.

[11] S. Heise, J. Wickert, G. Beyerle, T. Schmidt, and C. Reigber. Global monitoring of tropospheric water vapor with GPS radio occultation aboard CHAMP. *Advances in Space Research*, pages 2222–2227, 2006.

[12] E. J. Horvitz and J. C. Krumm. Learning, storing, analyzing, and reasoning about the loss of location-identifying signals. US patent 20070005243.

[13] R. Huang, K.-L. Ma, P. McCormick, and W. Ward. Visualizing industrial CT volume data for nondestructive testing applications. In *Proc' of the Visualization IEEE*, page 72, 2003.

[14] D. J. Kriegman, E. Triendl, and T. O. Binford. Stereo vision and navigation in buildings for mobile robots. *IEEE Transactions on Robotics and Automation*, 5(6), December 1989.

[15] S. M., B. N., and S. A. Hardware accelerated voxel carving. In *Proc. 1st Iberoamerican Symposium in Computer Graphics*, pages 289–297, 2002.

[16] T. Schmidt, J. Wickert, C. Marquardt, G. Beyerle, C. Reigber, R. Galas, and R. König. GPS radio occultation with CHAMP: an innovative remote sensing method of the atmosphere. *Advances in Space Research*, 33:1036–1040, 2004.

[17] C. Stolle, S. Schlüter, S. Heise, C. Jacobi, N. Jakowski, and A. Raabe. A GPS based three-dimensional ionospheric imaging tool: Process and assessment. *Advances in Space Research*, pages 2313–2317, 2006.

[18] R. Swinford. Building on-the-fly world models for pervasive gaming and other ubicomp applications using GPS availability data. In *The IEEE International Workshop on Intelligent Environments*, pages 133–142, 2005.

[19] J. P. Weiss, P. Axelrad, A. G. Dempster, C. Rizos, and S. Lim. Estimation of simplified reflection coefficients for improved modeling of urban multipath. In *Proc. ION 63rd Annual Meeting*, volume 7, pages 635–643, 2007.
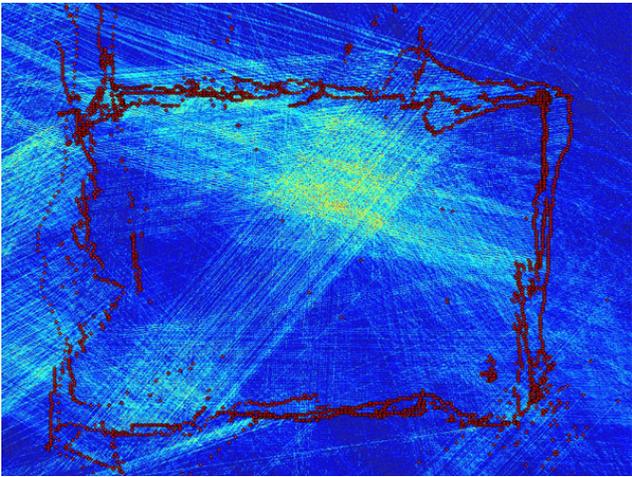
(a)
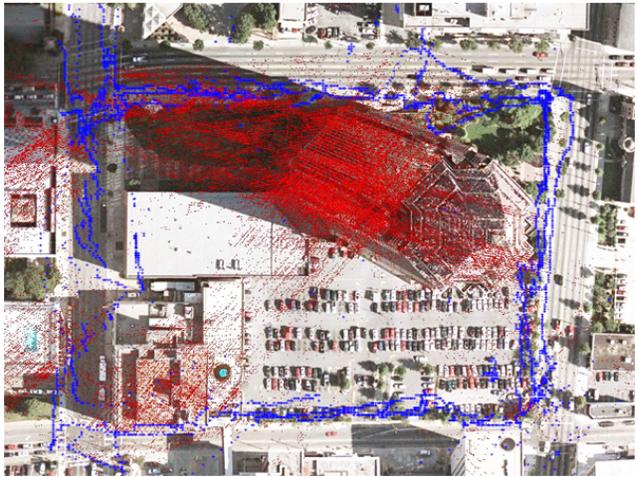
(b)
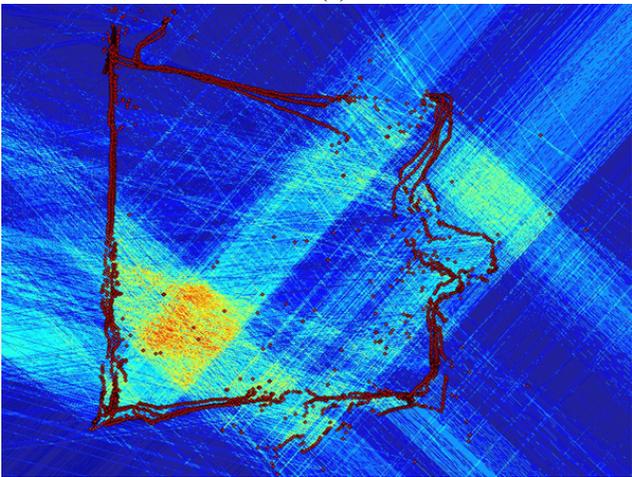
(c)
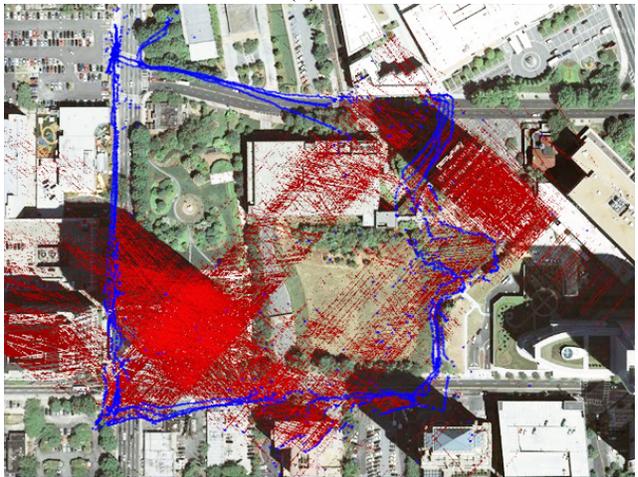
(d)

Figure 9. Density map of each site using 4000 samples (a&c), Compared to map imagery (blue traces are sample positions) (b&d).

| 500 samples
Half trip | 900 samples
One trip | 1100 samples
1.5 trips | 2300 samples
Two trips | 3200 samples
2.5 trips | 4000 Samples
Three trips |

| 400 samples
Half trip | 800 samples
One trip | 1300 samples
1.5 trips | 2400 samples
Two trips | 3800 samples
Three trips | 4000 Samples
Four trips |
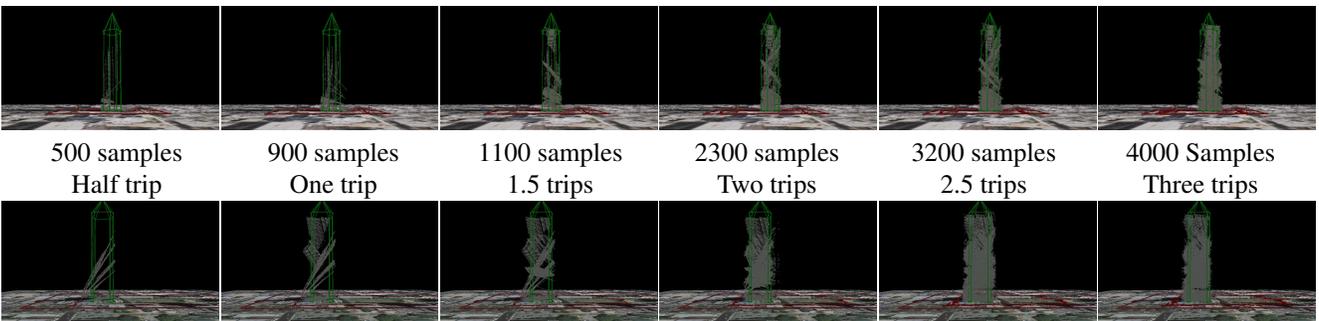
Figure 10. Voxel rendering as the number of data samples increases. Top: BOA and Bottom: OAC.