A Systems Approach to Sizing of Co-operative High-Accuracy Location (C-HALO) services validated by experiments in San Francisco

Venkatesan. N. Ekambaram*, Christian Manasseh**, Adam Goodliss**, Raja Sengupta**, Kannan Ramchandran*
**Department of Electrical Engineering and Computer Science, University of California, Berkeley*
***Institute for Transportation Studies, University of California, Berkeley*

BIOGRAPHY

Venkatesan Ekambaram is a PhD student at the Department of EECS, University of California, Berkeley since August 2009. He received his Masters from the Indian Institute of Science, Bangalore, India in July 2008. His research interests are in the area of peer-to-peer networks, signal processing for wireless communications and intelligent transportation systems. He is a co-recipient of the best paper award in the algorithms track IEEE/ACM DCOSS'08.

Dr. Christian Manasseh graduated with a PhD in December 2010 from the University of California, Berkeley's Institute for Transportation Studies encompassed by the Civil and Environmental Engineering Department. He currently works for a privately owned start-up company.

Adam Goodliss graduated with a Masters of Engineering Degree in December 2010 from the University of California, Berkeley's Institute for Transportation Studies encompassed by the Civil and Environmental Engineering Department. Currently Adam is a Management Consultant at Oliver Wyman and has worked on engagements within the Aviation and Retail industries.

Dr. Raja Sengupta is currently Associate Professor in the Systems program of the department of Civil and Environmental Engineering at the University of California at Berkeley. He received his Ph.D. from the EECS department of the University of Michigan, at Ann Arbor. His current research interests are in DSRC, networked estimation and control, vision based control of unmanned air vehicles, and collaborative behavior in robotic systems. He has served as Associate Editor of the IEEE Control Systems magazine and of the Journal of Intelligent Transportation Systems. He was Program Chair of the IEEE Conference on Autonomous Intelligent Networked Systems 2003 and Co-General Chair of the first ACM MOBICOM Workshop on Vehicular Ad-hoc Networks held in 2004, Co-Chair of the Program Committee for the second ACM MOBICOM Workshop on Vehicular Ad-hoc Networks held in 2005, Program Chair for the First International Symposium on Vehicular Computing Systems 2008, and will be Co-General Chair or IEEE WIVEC 2011.

Dr. Kannan Ramchandran received his Ph.D. from Columbia University in 1993. He is a Professor in the Department of Electrical Engineering and Computer Science of the University of California at Berkeley. His research group is the BASiCS group. Between 1993 and 1999 he was on the faculty of the Department of Electrical and Computer Engineering and the Coordinated Science Lab (CSL) at UIUC and a full-time Beckman Institute faculty member in the Image Formation and Processing Group. His fields of professional interest are communication and information theory, networking, image and video compression and modeling, multirate and multiresolutional signal processing, wavelets, robust image and video communication, packet video, and fast algorithms for signal and image processing.

ABSTRACT

This paper presents a methodology for assessing the accuracy gap between the current level-of-service provided by the US GPS constellation and DGPS, and the level-of-service envisaged by new Cooperative High-Accuracy LOcation (C-HALO) services that achieve decimeter accuracy. We present a novel GIS-based Hidden-Markov Model (HMM) predictive framework to estimate the fraction of roads with low satellite visibility counts that we call as the "dark area", where C-HALO cannot be realized. Out of the total area of San Francisco (121 sq.km), 0.3 to 4% of the San Francisco streets are predicted to fall in the "dark" area with 95% confidence. Today, the DoD provides guarantees of 7m accuracies 97% of the time using the GPS constellation. Our dark area falls within this 3%. If it were covered by systematic new C-HALO investment, the nation might realize a similar performance guarantee at the decimeter level that would be meaningful for safe and efficient travel on the national roadway system.

INTRODUCTION

Several high-accuracy positioning services exist in the market today, yet some of the national road area is still under-covered by these services. The choice of technology and the role that the government would play in rolling out an infrastructure to enable high-accuracy positioning will have an effect on the overall cost of the system. This work is motivated by a parallel work on a cost benefit study [1] of deploying Cooperative High-Accuracy LOcation (C-HALO) for applications driven by Intelligent Transportation Systems, that requires a verifiable systems model for estimating the costs of a large scale deployment of C-HALO infrastructure. We present a methodology for assessing the accuracy gap between the current level-of-service provided by the US GPS constellation and DGPS, and the level-of-service envisaged by new C-HALO services that achieve decimeter accuracy.

Our systems approach to implementing C-HALO relies on a GIS-based Hidden-Markov Model (HMM) [2] predictive framework to estimate the fraction of roads guaranteed not to have C-HALO with the current technologies. We call this gap the "dark area", i.e., the road area that does not have the decimeter level accuracy required to realize benefits from safer, greener, and more efficient travel. C-HALO in the dark area is not provided by the current GPS constellation. In this paper, the dark area is defined as streets with low satellite visibility counts. Since new augmentation technologies such as N-RTK [3] or HA-NDGPS [4] usually provide high accuracy with higher satellite counts (e.g. six or more), they will likely be unable to enhance accuracy in the dark area. We use the method to compute the dark area in the city of San Francisco.

There is existing work in the literature that predicts the GPS availability in urban areas. Alcantarilla et. al. [5] conduct a simulation of an urban environment and contend that with GPS & Galileo 65% of the area is covered by more than 3 satellites, while 20% is covered by 3, and 15% by less than 3. They then go on to qualitatively discuss the principal pieces of a future GPS system along with the envisioned benefits of multi-constellation GNSS SBAS augmentations. Similar analysis is carried out by Zabic et. al. [6] but with actual data in Copenhagen. They estimate the average satellite availabity in Copenhagen through extensive data collection and use simulation tools to predict the improvement in satellite availability with the addition of Galileo. Taylor et. al. [7] estimate the GPS availability using precise LiDAR data and digital surface maps. Our approach differs from these existing models in the following aspect that we only use the GIS data that is publicly available and develop a stochastic model to estimate the satellite counts. We compare our results with those in the literature in the section where we present our model.

Out of the total area of San Francisco (121 sq.km), 0.3 to 4% of the San Francisco streets are predicted to fall in the "dark" area with 95% confidence. The dark area is defined as the area where less than 7 satellites are seen from street level. These correspond to roads in which C-HALO is not available without new investment. Knowing the size of the dark area is essential to quantifying the magnitude of C-HALO investment required by any technology. In the sections to follow, we present the dark area estimation methodology.

DARK AREA ESTIMATION

Several reports exist on the causes of errors when measuring position on the ground using the GPS system [8]. These reports address the theoretical values of the various types of errors. Table 1 below shows the possible values for the different errors attributed to locating objects on the ground using GPS.

Emerging technologies such as the penetration of INS systems in vehicles could mitigate the accuracy gap. For example, we know from our prior work [9] that GPS augmented with INS can dead reckon to lane level precision for about 20 seconds if there are no sudden lane changes or turns at intersections [9,10].

Source	Effect (m)
Signal Arrival C/A	±3
Signal Arrival P(Y)	±0.3
Ionospheric effects	±5
Ephemeris errors	±2.5
Satellite clock errors	±2
Multipath distortion	±1
Tropospheric effects	±0.5
σ_R C/A	±6.7
σ_R P(Y)	±6.0

Table 1: Sources of GPS Errors.

As part of this study, we set out to estimate the size of the "gap" using empirical and data modeling techniques to arrive at a more accurate assessment of GPS accuracy on the ground. Our method for doing this relies on understanding the satellite coverage and the visibility of satellites at a Point-of-Interest (POI) on the ground. When the POI is in an open space environment, the GPS receiver is capable of communicating with several satellites (6 or more) and is able to locate the POI with good accuracy (1-3m). When comparing this POI with another POI in an urban setting with several high-rise buildings, the number of satellites viewed drops significantly resulting in lower location accuracy.

Our effort rests on modeling the relation between position accuracy and number of satellites-in-view by incorporating the Position Dilution of Precision (PDOP) values, the height of buildings near the POI, and the open-space area - as represented by street widths - into the model. The method is tested on data from the city of San Francisco. We collected validation data in the San Francisco Downtown blocks highlighted in Figure 1. This area encompasses 2 sq.km of buildings of various heights providing us with a variety of satellite counts. Figure 2 is a Google Earth 3D rendering showing the structures in this area as of 2009.

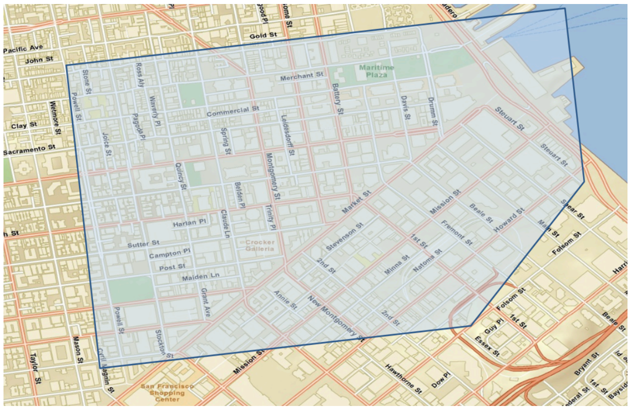

Figure 1: Shaded area represents study area.

DATA FOR MODELING SATELLITE COUNT

In order for us to systematically replicate the modeling of the gap across various cities, we rely on data that is easily accessible in the public domain. The International Association of Assessing Officers (IAAO) in collaboration with the Urban and Regional Information Systems Association (URISA) have been among the leading efforts in enabling GIS use by cities all over the world. As a result almost all major cities in the US have implemented GIS and are affiliated with either of those two organizations. In San Francisco, the assessor office manages the SFParcel GIS system [11], which holds information on close to 198,000 parcels. Of those we were able to obtain clean data on 160,000 parcels covering approximately 86% of the built area of San Francisco. In the remaining 14%, the height data could not be verified. These are dropped from the model (visualized as grey points in the plots below). The 86% that is used covers only the parts of San Francisco that are registered with the assessor's office. This does not include open spaces, public gardens, etc. Those areas (aka Park Acres) are estimated by the San Francisco County's office to be 0.19% of the total 121sq.km area of the City of San Francisco. So for the purposes of this model, we will assume them to be negligible, and the clean data we have on San Francisco from the assessor's office will be assumed to cover all the 121sq.km.

The model is constructed using the ESRI GIS software ArcMap. Data for the model includes:

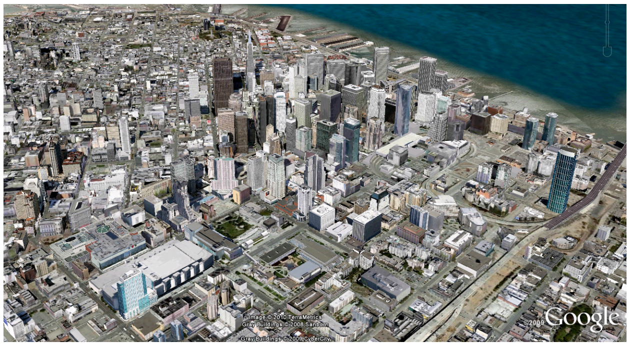

Figure 2: 3D rendering showing building coverage in validation area.

- Building heights as reported by the SFParcel GIS system controlled by the County of San Francisco
- Street width as measured using the ArcMap GIS software

Thus building heights and street width at a Point Of Interest (POI) are "known" variables in the model and could be obtained from the GIS system of most city assessor's office. The "unknown" variable is the satellite count. To calibrate the model we measure satellite count on the ground in the proposed area. This is done by driving around with a GPS equipped Smartphone. We developed an application on the Windows Mobile 6.5 operating system and deployed it on two HTC phones, namely, the HTC Diamond and HTC Touch Pro 2. The application logs the following values:

- GPS Longitude and Latitude
- Number of Satellites Visible
- Number of Satellites Connected
- Vertical Dilution of Position (VDOP)
- Horizontal Dilution of Position (HDOP)

The number of satellites at a POI can be taken as an indicator of the GPS accuracy. However the Horizontal Dilution of Precision (HDOP) is a better indicator of the localization accuracy of the GPS. For example, given a fixed number of satellites, the accuracy is better at a POI where the satellites are seen well spread out as compared to a place where the satellites are more clustered together. The HDOP captures this.

The preliminary data collected is visualized in Figure 3. Given the total number of operational satellites (N = 30 [12]) and the predicted number of satellites (s) at a POI, the HDOP can be theoretically calculated as follows assuming the satellites to be uniformly spread in the space,

$$\mathrm{HDOP} = \frac{4}{s\left(1 - \frac{\sin(2\cos^{-1}(1-\frac{2s}{N}))}{2\cos^{-1}(1-\frac{2s}{N})}\right)}$$

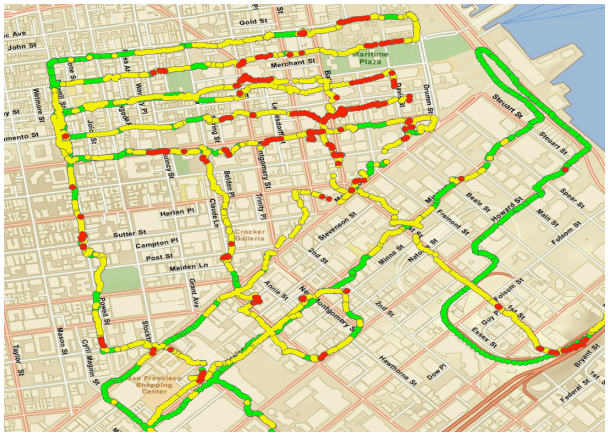

Figure 3: GPS data collection depicting satellite counts: >6 are shown in green, between 4 and 6 are shown in yellow, <4 are shown in red.

Figure 4 shows the theoretical and the empirical HDOP values obtained from the data set along with the 95% confidence intervals for the empirical HDOP. The empirical HDOP values were obtained from the data set collected in the city of San Francisco. The theoretical HDOP is obtained from the equation above.

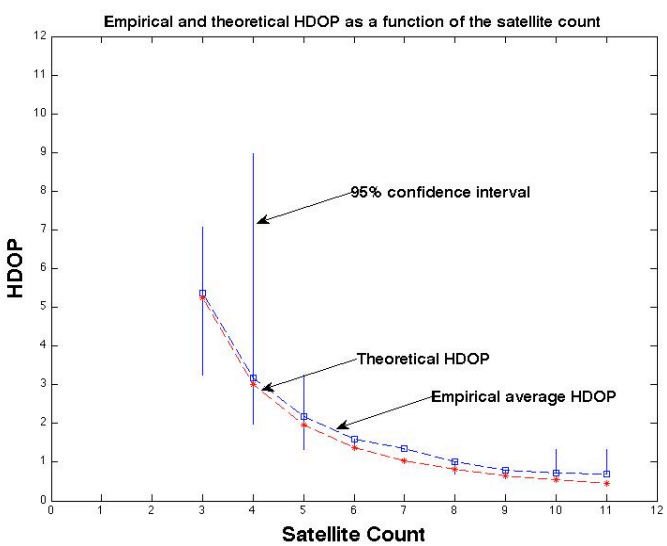

Figure 4: Empirical and predicted HDOP values with 95% confidence intervals.

The question of what values of HDOP are good for high-accuracy localization would depend on the receiver, ionospheric conditions etc. Typically, under normal conditions, HDOP values below 4 are considered to be good [14]. However, for high accuracy applications, we would require the HDOP values to be lesser than 2 or 1. Using this as a rule of thumb, based on Figure 4, we roughly categorized the satellite counts as < 4, 4 to 6 and > 6 and the model predictions were carried out for these three categories. The use of 6 as a threshold for good GPS coverage is empirically supported by a 100 miles of driving data [10]. In the next section we will describe the models used to predict the satellite counts and evaluate the performance based on the collected data.

HIDDEN MARKOV MODEL TO PREDICT SATELLITE COUNTS

This section describes the method used to predict the number of satellites at a POI given the GIS data i.e. building heights and street widths. The estimate of the satellite count at the point of interest is obtained as follows. We think of the satellites as being placed on the surface of a hemisphere with a radius R centered at the POI. We assume the POI is occluded from the satellites only by buildings on the sides of the street and there is visibility in the forward and backward directions as in Figure 5. The mask angle alpha (shown

Figure 5: Mask-Angle Representation using street width and building heights.

In Figure 5) is calculated based on the heights of the buildings and street width as follows.

$$\begin{aligned} \alpha &= \tan^{-1}\left(\frac{h_1}{w}\right) + \tan^{-1}\left(\frac{h_2}{w}\right) \\ \alpha &= \text{mask angle} \\ h_1, h_2 &= \text{height of building} \\ w &= \text{street width} \end{aligned}$$

The satellites visible at this point, are essentially the ones lying on a strip of the hemisphere with angular width alpha. The fraction of these satellites is given by,

$$\text{Satellite Count} = \text{N} \times \alpha$$

where N is the total number of satellites in orbit. The satellite count data collected in downtown San Francisco

is compared against the predicted count computed as described above. The data set consisted of 1657 data points. These were a subset of the 13822 data points collected overall. The subset was chosen by excluding data points that were not part of downtown San Francisco as in Figure 3 and data points that did not have corresponding meaningful building heights in the SFParcel GIS system. Out of these 1657 data points, 34 data points had satellite counts < 4, 568 points had satellite counts 4 to 6 and the rest had satellite counts > 6.

Table 2 shows the prediction accuracy of the model. Each column in this table is the prediction accuracy for the corresponding category of satellites. For example column 1 shows the percentage of data points having < 4 satellites being predicted as < 4, 4 to 6 and > 6 number of satellites.

Satellites	True < 4	True 4 to 6	True >6
Predicted < 4	0.74	0.16	0.02
Predicted 4 to 6	0.24	0.25	0.07
Predicted > 6	0.02	0.59	0.91

Table 2: Model Prediction Accuracy.

The overall prediction accuracy is around 69%. The prediction accuracy is calculated by adding the fraction of data points in each of the categories multiplied with the corresponding diagonal entry. We next use a Hidden Markov Model (HMM) [2] to improve the prediction accuracies. The HMM captures the statistical dependence of the satellite count at a POI based on the building heights from neighboring points as well.

The idea of the HMM modeling is as follows. Depending on the time of day, climatic conditions, or scatter in the environment, the number of satellites visible at a point could vary significantly. These random parameters lead to a stochastic dependence between the building heights and the number of satellites at a given point. Furthermore, the number of satellites at a particular point would be dependent on the number of satellites in nearby points. These dependencies could be captured by a HMM as shown in Figure 6.

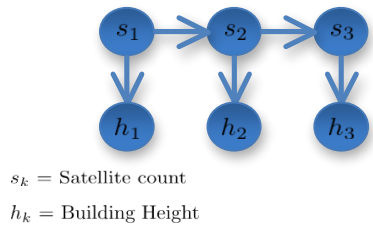

Figure 6: HMM for predicting satellite count.

The nodes corresponding to the heights are values that are known. The nodes corresponding to the satellite count are the hidden nodes that need to be estimated. The hidden nodes are connected to their neighbors to model the dependency between satellite counts in adjacent regions. The transition probabilities between the satellite count variables are modeled using a sticky Markov chain [2]. This is validated using the empirical data. The distribution of the building height given the satellite count is modeled as a Gaussian random variable with mean and variance empirically determined from the collected data. The mean height and variance for the Gaussian model and the transition probabilities between the states were obtained empirically from the collected data. 95% confidence intervals were calculated for the estimated parameters of the model. The satellite counts were predicted by taking the mean of the estimated parameters and the corresponding prediction accuracies for this model are as shown in Table 3.

Satellites	True < 4	True 4 to 6	True >6
Predicted < 4	0.70	0.01	0
Predicted 4 to 6	0.08	0.41	0.11
Predicted > 6	0.22	0.58	0.89

Table 3: HMM Satellite Count Accuracies.

The results of this model are 87% accurate. Figure 7 shows the results of the model for the city of San Francisco. The prediction is made for all the streets of San Francisco where we have the building height data from the assessor's office. The figure is drawn by aggregating and averaging the values of the model in 10m x 10m grids that lie on roads. Each grid is given a color based on the average satellite count in that grid: red if <4, yellow if between 4 and 6 and green if >6. The 14% of San Francisco for which we do not have height data is not modeled. It appears as grey areas in the figure. We would also like to note that the effects of multipath are not taken into account in this modeling. Thus the green areas do not necessarily reflect regions of high accuracy. Even though there is a satellite visibility of >6, multipath can cause

significant errors. However, we can say with high confidence that the dark areas are regions of bad location accuracies. Experimental results [10] have shown that with less than 7 satellites, GPS estimates have errors on the order of 1 meter. Our model predicts that 0.3 to 4 % of the streets of San Francisco have a satellite coverage of less than 7 satellites with 95% confidence.

Table 3 can be compared to existing work in the literature. The methods in the literature use precise LiDAR data to yield good prediction accuracies [7]. The authors of [7] evaluated their method with two test cases. With a 5m RADAR digital surface map, the mean error in the predicted number of satellites using their model is 2.7 and 4.85 satellites with the corresponding error percentages being 46% and 82% in the two test cases. This improves with a more precise digital surface map. However it would be very expensive to obtain precise LiDAR data for an entire city and all the cities in the US. Under the same error metric, the mean error in the predicted number of satellites using our model is 1.85 satellites and the error percentage is 29.14%.

Figure 7: GPS data estimating satellite counts: >6 or more are shown in green, between 4 and 6 are shown in yellow, <4 are shown in red.

We use only building height data that can be obtained from the city planning department or a similar agency. However the comparison of the models is to be taken with caution given that the test data under consideration is vastly different for the two approaches. Our accuracies could be improved by having more data points. Extensive data collection during different times of the day and in different regions can help build and evaluate better models.

USE OF THE MODEL

The approach we adopted to obtain the satellite count in San Francisco can be extended to other cities in the US. The method is based on a GIS model constructed from the building heights and street widths in different parts of the city – information that is easily accessible for all urban centers in the US via the local assessor's office. This would quantify the area where decimeter level accuracy cannot be achieved today with the current GPS and DGPS technology in many cities nationwide.

Our method produces a figure such as Figure 7. Such a figure can guide the phased deployment of C-HALO infrastructure and provide insight into the full extent of new infrastructure required. The red areas are candidates for a first phase C-HALO deployment. Fortunately, they are also few in number, suggesting one might reap considerable improvements in location accuracies for moderate initial investment.

The benefits of deploying in red or yellow areas can be better understood by using GIS tools to overlay other statistics such as the distribution of accidents on Figure 7. We have done this for San Francisco. Figure 8 shows a quadrant of San Francisco which includes 1000m x 1000m grids of 2008 accident data as reported by NHTSA where red are areas of high accident counts, yellow those of medium counts and green of low or no accidents. This type of plot could be repeated by overlaying emissions data, or congestion data or other type of data relevant to the benefit measures guiding C-HALO pilot deployments or initial infrastructure investments.

Figure 8: Overlay of accident data on satellite count projection data.

CONCLUSION

We presented a novel GIS based HMM predictive model to estimate the areas of low satellite visibility and validated the model based on experiments in San Francisco. This opens up many areas of research to improve the analysis and predictive capacity of the models. We can continue to update the model with more empirical data and expand it to other cities and determine the nationwide 'dark area.' Different models than the HMM could be investigated for better accuracy. Obtaining real data that spans across weather conditions

and time would enable the model to provide time-based projection of satellite coverage, yielding a statement such as satellite count less than 6 for x % of the time. The use of variations on the satellite count, such as the number of satellites used to calculate position could offer a better understanding of multi-path errors in the region.

ACKNOWLEDGEMENT

The authors gratefully acknowledge a gift from ATLIS Wireless, LLC that has, in part, made this work possible. The contents of this paper are solely determined by the authors.

REFERENCES

[1] Adam Goodliss, Christian Manasseh, Venkatesan Ekambaram, Raja Sengupta, Kannan Ramchandran, "Cooperative High-Accuracy Location (C-HALO) service for Intelligent Transportation Systems: A Cost Benefit Study", ION GNSS, 2011.

[2] M.I.Jordan, "Learning in Graphical Models", Kluwer Academic Publishers, 1998.

[3] Volker Janssen, "New Ways to Network RTK: How do VRS and MAC measure up?" http://eprints.utas.edu.au/9327/

[4]http://www.fhwa.dot.gov/publications/research/operations/02110/index.cfm.

[5] Alcantarilla, I., Porras, D., Tajdine, A., Zarraoa, N., Lévy, J.C., "The Benefits of Multi-constellation GNSS Augmentations," Proceedings of the 19th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2006), Fort Worth, TX, September 2006, pp. 930-938.

[6] M. Zabic and O.A. Nielsen, "An analysis of stand-alone GPS quality and simulated GNSS quality for Road Pricing."

[7] George Taylor, Jing Li, David Kidner and Mark Ware "Surface Modelling for GPS Satellite Visibility", LNCS 2005, Vol 3833/2005, Springer.

[8] LeLievre, Jaqueline. "GPSnet: NRTK Corrections and Dynamic Accuracy Trials". GNSS Technical Support Newsletter. Issue 29, June 2008. pp. 6.

[9] Hofmann-Wellenhof, B. and Lichtenegger, H. and Collins, J.,"GPS: Theory and practice", Springer Vienna, 1994.

[10] S. Rezaei and R. Sengupta, "Kalman Filter-Based on Integration of DGPS and Vehicle Sensors for Localization," *IEEE Transactions on Control Systems Technology*, November 2007, Vol. 15, No. 6, pp. 1080-1088.

[11] Sengupta R, Rezaei S, Shladover S. E, Cody D, Dickey S, Krishnan H, Cooperative collision warning systems: Concept definition and experimental implementation, Journal of Intelligent Transportation Systems.

[12]http://gispub02.sfgov.org/website/sfparcel/INDEX.htm

[13]http://en.wikipedia.org/wiki/Global_Positioning_System

[14]http://www.developerfusion.com/article/4652/writing-your-own-gps-applications-part-2/2/