Combining Image Processing with Signal Processing to Improve Radio Position Estimation

Amy M. Abraham, Richard K. Martin, and Kirk A. Mathews
The Air Force Institute of Technology
Department of Electrical and Computer Engineering
WPAFB, OH 45433
amy.abraham.4@us.af.mil, richard.martin@afit.edu, kirk.mathews@afit.edu

Abstract—This paper uses aerial, orthorectified imagery to model the radio multipath environment. This model is used to improve time difference of arrival (TDOA) based radio positioning. The imagery is processed using a building extraction algorithm compiled from current techniques. The building model is used to develop a shortest-path routing protocol from each candidate transmitter position to each receiver, creating more accurate models of the TDOA information than are available when assuming all paths are line of sight (LOS). The final positioning algorithm is a maximum likelihood algorithm that compares the observed and theoretical TDOA values at each grid point. When compared to the Chan & Ho method which assumes line of sight (LOS), the method in this paper improves transmitter geolocation error by an average of 44 m (53%) in non-LOS environments. However, in cases where all receivers actually have a LOS, the current method is faster and slightly more accurate. Therefore, the method in this paper is most applicable to scenarios requiring position estimation of vehicles in an urban environment using stationary receivers and significant computing power.

I. INTRODUCTION

There are many situations in which it is necessary or useful to locate someone or something by using the signal emissions from a radio transmitter on the ground. Geolocation can be applied to military, commercial, law enforcement, and search & rescue scenarios. Due to the prevalent need for this capability, many methods have been developed for accomplishing geolocation. Most of these methods rely on signal timing information, typically time difference of arrival (TDOA) data. TDOA methods solve for the intersection of a set of hyperbolic curves which are defined by the TDOA data. However, this assumes that the signal travels along a direct path from the transmitter to each of the receivers. This is rarely the case in an urban environment. Therefore, the method developed in this paper seeks to improve geolocation when there are obstructions between the receivers and the transmitter, using an aerial, orthorectified, RGB image of the area in which the transmitter and the receivers are located.

One of the common methods of solving for the intersection of the hyperbolic curves created by the TDOA data is the Taylor series method [1], [2]. It is an iterative method which improves upon an initial location estimate by repeatedly linearizing and re-solving the local least-squares (LS) cost function. If the initial location guess is not close enough to the actual location, then local minima may be mistaken for the actual location. Also, the solution may not converge. Aside from these issues, the Taylor series method provides very accurate results when there are no obstructions between the receivers and the transmitter. Therefore, this method, as presented by Chan and Ho [1], is used as the basis for our comparisons.

Little research is available that accounts for obstructions between the transmitter and the receivers. The method discussed by [2] attempts to correct the bias to the TDOA values caused by the longer path taken by the signal in the obstructed case, but this method relies on assumptions about the variance and distribution of the biased data, rather than on knowledge about the actual obstructions [2]. The only localization methods found to incorporate predictability using the actual environment are based on small, easily controlled, indoor environments and do not involve obstructions [3]. Therefore, this paper proposes a new method that uses building extraction from an overhead image to predict the actual paths taken by the signal.

There are many existing methods for processing an image for building extraction. Current image preprocessing techniques involve different color models, normalization, thresholding, contrast enhancement, filtering, etc. [4], [5]. Extracting objects of interest from an image generally requires some form of image segmentation to break down an image into relevant regions. Image segmentation methods include edge detection, seeded region growing, multithresholding, and clustering [4], [6], [7], [8], [9]. Finally, there are various methods of determining which of the segmented regions are likely to represent buildings. Image post-processing techniques include using spectral patterns, morphological operations, shadow markers, and feature extraction [8], [10].

The contributions of this paper are as follows. First, we develop a model for the non line of sight (NLOS) path length given a set of endpoints and a model of an urban multipath environment. Next, we combine existing image analysis techniques for building extraction with our new path length model to develop a maximum likelihood (ML) geolocation algorithm. Finally, we simulate the new algorithm in a variety of scenes and transmitter/receiver geometries to characterize its performance compared to LOS-based methods.

---

Funded in part by the Air Force Office of Scientific Research and the Air Force Research Laboratory. The views expressed in this paper are those of the authors, and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government. This document has been approved for public release; distribution unlimited.
II. PROPOSED ALGORITHM

The proposed approach has three components: (i) building extraction from overhead imagery, (ii) shortest path length modeling, and (iii) ML estimation. We address each of these in turn.

The inputs to Step (i) are an orthorectified, aerial, RGB image; and rough minimum and maximum expected building perimeter estimates. All image processing techniques used to extract building locations existed prior to this research [7], [8], [10]. Fig. 1 shows an example of the main steps in the building extraction process for one of the images used. First, in Fig. 1(a), the RGB values in the image are converted to grayscale intensity values and a contrast enhancement technique is applied. In Fig. 1(b), pixels with an intensity below a threshold are classified as shadows, and the remaining pixels in the image are clustered based on their intensity values. The connected pixels in each cluster are potential building objects. The shadow objects are processed separately and evaluated based on their shape and size. The direction of the sun is calculated, and this direction indicates which side of each shadow is likely to overlap the building that casts the shadow. In Fig. 1(c), those building objects which abut the same shadow in the direction of the sun are combined into one object. Each combined object is then classified based on size and shape as probable buildings in the image, and in Fig. 1(d) a rectangle of best fit is found for each object. It is assumed for simplicity that the buildings are all rectangular with four walls. Since building extraction is only a preprocessing step, we will not expand on it here due to space limitations; see [7], [8], [10] for further details.

Step (ii) is the creation of a subroutine that can determine the shortest path from a transmitter location to a receiver location, such that the path does not pass through any buildings but is otherwise a series of connected line segments whose endpoints lie on these buildings. This assumes that the building map has already been used to determine that the receiver in question does not have a LOS from the transmitter.

The approach we take here is essentially a version of Dijkstra’s shortest path routing algorithm, with modifications to account for the fact that building walls are not points but rather are extended objects [12]. To find all possible paths, all of the potential reflection points must first be found. Each building is looked at individually to find which of its four walls, if any, faces the transmitter with an unobstructed view. The angle of the normal vector $\Omega_{wall}$ for each wall is computed; this normal vector is oriented to point out of the building. The wall is determined to be facing the transmitter if and only if the transmitter lies in the region bounded by the line running through the points in the wall and on the same side of the wall as its normal vector.

$$\langle x_{tx} - x_{wall} \rangle \cdot \Omega_{wall} > 0,$$  \hspace{1cm} (1)

where $x_{tx}$ is the transmitter location and $x_{wall}$ is the midpoint of the wall. If the wall is facing the transmitter, the LOS between the transmitter and the wall is checked for obstructions. This is computationally intensive, because every point along the wall must be checked; heren, we used a discrete grid of points along the wall with spacing equal to the pixel size, since the building extraction routine already limits the accuracy to the quantization of the pixel size. All unobstructed points in the wall are retained.

Next, a second list is compiled of walls which can serve as a second reflector. Each wall in the first list is compared to every other wall which does not belong to the same building. If two walls face one another (determined via the same process as in 1), then every point in the potential second reflection wall is checked against every point in the candidate first-reflector wall for obstructions. All points with an unobstructed view of at least one point on the opposing wall are retained. A third list could be created of triple-reflection paths. However, we assume that three reflections would make the signal too weak for detection by the receivers because the signal loses power with each reflection.

Finally, we complete the paths to the receiver. The receiver must be checked against every wall in both lists of single and double reflection points to see if the wall faces the receiver with an unobstructed view. Once the point of reflection is found for each wall piece in regards to the path it describes, the length of that path is calculated as the sum of the distances from one point to the next along the path, beginning with the transmitter and ending with the receivers. Note that throughout this process, we restrict the problem to two dimensions and assume that the wall surfaces of the buildings are rough enough to exhibit significant reflectance in all directions.

Step (iii) of the proposed approach is ML estimation of the transmitter’s location using the observed TDOAs and the model from step (ii). We assume that the error $\epsilon_{i,j}$ in the TDOA between receivers $i$ and $j$ is Gaussian [1], with zero mean and variance $\sigma^2_{\epsilon}$. Thus, the model for the TDOA between receivers $i$ and $j$ is

$$\delta_{i,j} = c^{-1} (f(x_{tx}, x_{rx,i}) - f(x_{tx}, x_{rx,j})) + \epsilon_{i,j}$$  \hspace{1cm} (2)

Fig. 1. Building extraction from imagery. The major steps are ((a), top left) the grayscale intensity image after contrast enhancement; ((b), top right) the shadow objects and the computed direction of the sun; ((c), bottom left) the building objects found to overlap the shadow objects; and ((d), bottom right) the result of fitting rectangles to the building objects. Original image data available from the U.S. Geological Survey [11].
TABLE I. STATISTICS FOR THE BUILDING EXTRACTION PROCESS, WHERE GSD IS THE GROUND SAMPLE DISTANCE.

<table>
<thead>
<tr>
<th>Original Size (pixels)</th>
<th>GSD</th>
<th>Scale Factor</th>
<th>Scaled Size (pixels)</th>
<th>GSD</th>
<th>Extraction Runtime</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>977 × 1149</td>
<td>0.61 m</td>
<td>1/70</td>
<td>98 × 115</td>
<td>6.1 m</td>
</tr>
<tr>
<td>2</td>
<td>1009 × 1177</td>
<td>0.30 m</td>
<td>1/15</td>
<td>101 × 119</td>
<td>4.5 m</td>
</tr>
<tr>
<td>3</td>
<td>1019 × 2201</td>
<td>0.30 m</td>
<td>1/15</td>
<td>112 × 137</td>
<td>3.5 m</td>
</tr>
</tbody>
</table>

where \( f(x_{tx}, x_{rx,i}) \) is the evaluation of the path length function from step (ii) for transmitter position \( x_{tx} \) and receiver position \( x_{rx,i} \), and \( c \) is the speed of light. For Gaussian noise, the ML solution reduces to a nonlinear least squares (NLS) problem,

\[
x_{tx} = \arg \min_{x_{tx}} \sum_{i=1}^{S} \left( \delta_{i,1} - \frac{f(x_{tx}, x_{rx,i}) - f(x_{tx}, x_{rx,1})}{c} \right)^2
\]

where \( S \) is the total number of receivers. Note that \( \delta_{i,1} = \delta_{1,1} \), so it suffices to only use TDOAs relative to receiver 1 rather than include all pairs of receivers. Herein, we solve (3) via a simple grid search to ensure an optimal solution (to within the resolution of the grid cells), though greedy or gradient descent approaches could be devised as alternatives. In the grid search approach, a coarse grid is used for an initial estimate (with grid cells 10 pixels wide), then a refined search is performed to search within the initial grid cell.

III. SIMULATION RESULTS

Since the overhead images were available online [11], a TDOA experiment could not be set up concurrent with the image collection, hence the TDOA data is simulated here. The true TDOA is created in the same manner as used in the estimation procedure, i.e. (2) is used, which may or may not be indicative of the actual multipath length in a real-world experiment. However, the results shown here are realistic enough to motivate further experimental validation. Three images (scenes) were used here; we will creatively refer to them as scenes 1, 2, and 3. Table I provides the relevant information about the 3 images used in the geolocation algorithm. The column “Original Size” refers to the image size within [11], but we scaled down the sizes to reduce computation time. The scaled images had pixel widths of either 6.1 m or 4.5 m, so the coarse grid search described in Step (iii) above had initial grid sizes of 61 m or 45 m. The column “Extraction Runtime” gives the runtime for the image processing portion, though as the building extraction algorithm developed in this paper is a compilation of techniques already in use, these statistics are only included here for reference.

Four experiment were conducted, and these will be discussed momentarily. For each setup, 50 realizations were generated, and the average error and runtime are recorded both for the proposed method, denoted (P), and for the current Taylor series method of [1], denoted (C). The Taylor series method requires an initial position guess, but the placement of this guess does not affect the result as long as there are no local minima. Therefore, the same initial position guess is used for the Taylor series method throughout the simulations. Position locations are described by pixels in relation to an origin located at the top left corner of the image.

![Image](image323to564to738)

Fig. 2. Sample results for scenes 1, 2, and 3. In each image, the true transmitter location is marked “o”, and the shortest paths from the transmitter to the receivers are drawn as dashed blue lines. The result of the proposed algorithm is plotted as a red “*”, and the paths from this location to the receivers are drawn as green lines. Each iterative result of the current method [1] is plotted as a yellow circle and connected to the next iteration by a yellow line, with the final estimate marked “x”. The subplots show (a) scene 1 in the top left, (b) scene 2 in the top right, and (c) scene 3 at the bottom; c.f. Table II. Original image data available at [11].

TABLE II. COMPARISON OF GEOLOCATION FOR 3 DIFFERENT IMAGES AVERAGED OVER 50 REALIZATIONS FOR THE PROPOSED (P) AND CURRENT TAYLOR SERIES (C) METHODS.

<table>
<thead>
<tr>
<th></th>
<th>P Error</th>
<th>C Error</th>
<th>P Offline</th>
<th>P Online</th>
<th>C Runtime</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>23.7 m</td>
<td>48.9 m</td>
<td>17.4 min</td>
<td>68.2 s</td>
<td>0.0065 s</td>
</tr>
<tr>
<td>2</td>
<td>30.4 m</td>
<td>66.4 m</td>
<td>30.9 min</td>
<td>134.4 s</td>
<td>0.022 s</td>
</tr>
<tr>
<td>3</td>
<td>94.2 m</td>
<td>143.1 m</td>
<td>7.9 min</td>
<td>41.3 s</td>
<td>0.063 s</td>
</tr>
</tbody>
</table>

The y direction increases positively when traveling vertically downward from the origin. Unless specified otherwise, the default parameter values are: scene 1, TDOA standard deviation \( \sigma_{e} = 35 \) m, transmitter location \( x_{tx} = (75, 35) \), and the receiver configuration shown in Fig. 2(a).

In the reported results, the runtime for the proposed method is divided into an offline runtime and an online runtime. The offline runtime includes the time it takes for the building extraction algorithm and the time it takes to create partial paths beginning at the receivers, which will be completed by each potential transmitter location. The online runtime is the time it takes to search for the transmitter location in the image grid of possible locations.

The first experiment varies the scene (image) that was used. Fig. 2 illustrates one of the 50 realizations of the random data for each of the images. The thin line indicates the progression of the iterations of the Chan & Ho method [1], culminating in the estimate marked with an “x”. The true transmitter location and the estimate from the proposed method are marked with “o” and “*”, respectively; and the thick lines emanating from these two markers indicate the true and estimated radio paths. The numbers indicate the positions of the five receivers. Table II compares the mean absolute position error (MAPE) of the position estimates and the runtime, averaged over 50 realizations.
The second experiment varies the true transmitter location within scene 1. One realization for each transmitter location is illustrated in Fig. 3. The MAPE and online runtimes are presented in Table III. The offline runtime is 17.4 minutes, the same value reported in Table II. The Taylor series method does not converge to a result for transmitter location (a). The same nonconvergence issue arises in some of the following simulations as well. This lack of convergence is a downfall of the particular method of solving the hyperbolic equations, especially when the multipath makes the presumed model invalid; but the Taylor series method is one of the more common methods, since it is simple to implement with a very fast runtime. It is adequate for comparison purposes to paint an overall picture of the effectiveness of the proposed method.

The third experiment varies the receiver configuration. One realization for each receiver configuration is illustrated in Fig. 4, and the MAPE and online runtimes are presented in Table IV. The results show that receiver configuration (a) in Fig. 4 is the most ideal configuration out of the 5 that were simulated, which agrees with well-known results for LOS-only situations. Configurations such as (b) and (c) cause ambiguity in the results due to a loss in dimensionality. In a case like configuration (d), where the transmitter location is not within the convex hull of the receivers, both methods perform poorly. For the Taylor series method, however, this is not a fair conclusion, since the algorithm terminated iterations at a local minimum value, and a better initialization (e.g. in a tracking scenario) would yield better performance. Also note that receiver configuration (e) shows a situation where each receivers has a LOS to the transmitter. The fact that it is so rare to find an unobstructed transmitter location proves how necessary it is to account for obstructions. Unfortunately, the proposed geolocation algorithm takes much longer than the Taylor series method to calculate a result.
The final experiment varies the TDOA error via the $\sigma_N$ parameter. Fig. 5 plots the average error in the result of both methods across 50 realizations for each $\sigma_N$ value. Note that the buildings used in the scenes here are typically smaller than 50 m in either dimension, so the proposed method yields large MAPE improvements so long as the error $\sigma_N$ is smaller than the buildings. Thus, for a reasonable noise level, the proposed method performs more accurately than the current Taylor series method when obstructions are present.

IV. CONCLUSIONS AND FUTURE WORK

When compared to the Taylor series method, the proposed method improves the geolocation error by an average of 44 m, or 53%, in cases when the receivers do not all have LOSs to the transmitter. This improvement is based on the case of $\sigma_N = 35$ m simulations, and for fairness this does not include the cases in which the Taylor series method does not converge or converges to a local minimum. The runtime of the proposed method is divided into an average online runtime ranging from around 20 seconds to around 2 minutes and an offline runtime ranging from around 7 minutes to around 30 minutes. The runtime of the Taylor series method averages around 0.02 seconds.

Depending on the application, the improved result may be worth the slower computation time, and the offline portions can be performed in advance. However, the Taylor series method performs more accurately in situations where all of the receivers have unobstructed lines of sight. There may, however, be a way to classify the TDOA as either reflected or direct based on measured and expected signal strength and to use this information to choose the appropriate method. Once could also compute the Taylor series method first and then check if the estimated transmitter location does not have a LOS to each receiver, and if it does not, then apply the proposed method.

It is important to assess the algorithm’s performance with experimental data, rather than simulated data, though development of a full radio TDOA testbed is beyond the scope of this conference paper. It would also be more accurate to treat the problem as three-dimensional, which would require more advanced image processing to determine the height of the buildings. The biggest drawback to this algorithm is the amount of required computation time. Unfortunately, any type of grid search will be time consuming. It is therefore desirable to eliminate the need for a grid search. There may be other existing geolocation techniques that can be adapted to include image intelligence.

Finally, there are other potential uses of the image data in vehicle tracking problems. In particular, if a segmentation algorithm is used to extract the roads, then the position search could be restricted to locations on the roads, both reducing the computation time and improving the accuracy of the final estimate.

Additional details about this work are available in [13].

REFERENCES