

Article

## A Photogrammetric Approach for Assessing Positional Accuracy of OpenStreetMap<sup>®</sup> Roads

Roberto Canavosio-Zuzelski <sup>1,\*</sup>, Peggy Agouris <sup>1</sup> and Peter Doucette <sup>2</sup>

<sup>1</sup> Department of Geography and Geoinformation Science, George Mason University, 4400 University, MS6C3, Fairfax, VA 22030, USA; E-Mail: pagouris@gmu.edu

<sup>2</sup> Integrity Application Incorporated, 15020 Conference Center, Suite 100, Chantilly, VA 20151, USA; E-Mail: pdoucette@integrity-apps.com

\* Author to whom correspondence should be addressed; E-Mail: rczuzelski@hotmail.com; rcanavos@gmu.edu; Tel.: +1-703-378-8672 (ext. 3498).

*Received: 16 January 2013; in revised form: 28 February 2013 / Accepted: 18 March 2013 /*

*Published: 2 April 2013*

---

**Abstract:** As open source volunteered geographic information continues to gain popularity, the user community and data contributions are expected to grow, e.g., CloudMade, Apple, and Ushahidi now provide OpenStreetMap<sup>®</sup> (OSM) as a base layer for some of their mapping applications. This, coupled with the lack of cartographic standards and the expectation to one day be able to use this vector data for more geopositionally sensitive applications, like GPS navigation, leaves potential users and researchers to question the accuracy of the database. This research takes a photogrammetric approach to determining the positional accuracy of OSM road features using stereo imagery and a vector adjustment model. The method applies rigorous analytical measurement principles to compute accurate real world geolocations of OSM road vectors. The proposed approach was tested on several urban gridded city streets from the OSM database with the results showing that the post adjusted shape points improved positionally by 86%. Furthermore, the vector adjustment was able to recover 95% of the actual positional displacement present in the database. To demonstrate a practical application, a head-to-head positional accuracy assessment between OSM, the USGS National Map (TNM), and United States Census Bureau's Topologically Integrated Geographic Encoding Referencing (TIGER) 2007 roads was conducted.

**Keywords:** OpenStreetMap (OSM); positional accuracy; vector; roads; adjustment

---

## 1. Introduction

### 1.1. Motivation and Problem Statement

OpenStreetMap<sup>®</sup> (OSM) is an open source geographical mapping project that provides the public with a free digital map of the world. It is unique because its users are the main contributors and responsible for building and maintaining the map. OSM contributors can edit (update) the map in several ways including using Global Positioning System (GPS) waypoints and tracks to identify features in the field or measuring satellite imagery to identify roads and other geographical features of interest [1].

This type of open contributing by volunteer contributors is known as crowdsourcing and refers to many volunteers providing information into the database, where each individual volunteer contributes a small portion that pertains to their local knowledgebase. Goodchild [2] has coined the term “Volunteered Geographic Information” (VGI) to describe this type of crowdsourcing activity.

The VGI phenomenon raises some important questions regarding the expertise of its contributors, especially when it is applied to mapping features on the earth’s surface. For example, in the United States a Land Surveyor is required to hold a bachelor’s degree in Surveying Engineering and receive several years of practical experience to be eligible to sit for state exams to become licensed. The training covers complex measurement theory such as GPS positioning, photogrammetry, remote sensing, least squares statistical analysis, error propagation, and state boundary law [3]. Moreover, once a surveyor becomes licensed, most projects they work on require the adherence to strict mapping standards and positioning guidelines [4–6].

However, OSM contributors are not required to have any formal education or training in the mapping profession, nor required to follow any positioning standards to derive the vector data that is contained in the database. As OSM continues to gain popularity the user community and data contributions are expected to grow, e.g., CloudMade, Apple, and Ushahidi now provide OSM as a base layer for some of their mapping applications [7–9]. This, coupled with the lack of cartographic and data quality standards and the expectation to one day be able to use this vector data for more geospatially sensitive applications, like GPS navigation, leaves potential users and researchers to question the accuracy of the database.

### 1.2. Proposed Solution

To address this issue, a rigorous photogrammetric approach is taken to determine the positional accuracy of OSM roads by using aerial stereo imagery, ground control points (GCP’s), and a vector adjustment model to determine real world geolocations for OSM shape points. The vector adjustment is based on a traditional photogrammetric bundle adjustment model and results in: (1) adjusted OSM shape point locations, (2) shape point residuals (which describe the positional accuracy), and (3) confidence regions about the adjusted shape point locations, Circular Error and Linear Error at the 90th percentile (CE90/LE90).

The benefit of this approach is that it is based on recreating the sensor geometry present when the image was taken and using proven photogrammetric principles to estimate a ground position. The proposed approach takes into consideration distortions present in the imaging system and applies corrections (interior orientation) to more accurately represent the ideal image position, free of systematic error. GCP's are used for absolute orientation of the imagery and to form an accurate image-to-ground relationship via an analytical stereo model. The vector adjustment allows for a simultaneous adjustment of all the vector shape points at one time, thereby offering more redundancy to the least squares model and improving the overall confidence in the results. The approach is also location and vector dataset independent, allowing for different geographic databases in different areas of the world to be analyzed with this approach.

To demonstrate the proof-of-concept the proposed vector adjustment was used to compute the positional accuracy of several urban gridded city streets in the OSM database. Additionally, to validate the performance, the Root Mean Square Error (RMSE) metric for the adjustment residuals will be compared to the RMSE value determined from the differences between the OSM database and the corresponding location determined by high-accuracy GPS surveying techniques (ground truth). This type of comparison provides a quantitative way to describe the performance of the adjustment and demonstrates the positional accuracy computed is consistent with what is actually present in the database. Once this positional accuracy information is known it can be carried along as an attribute to the vector shape points and polylines at the feature level, thereby providing valuable metadata and improving the overall usefulness of the database.

To demonstrate a practical application of the vector adjustment, a head-to-head accuracy assessment between OSM [10], the United States Geological Survey (USGS) National Map (TNM) [11], and the United States Census Bureau's Topologically Integrated Geographic Encoding Referencing (TIGER) 2007 [12] road vectors was conducted to determine which database is the most positionally accurate over the test area. Having an approach to understand the spatial accuracy of geographic data can be useful for many applications including route planning, GPS navigation, and "*smart conflation*" where reference/target vectors are chosen based on their positional accuracy.

## 2. Background and Previous Work

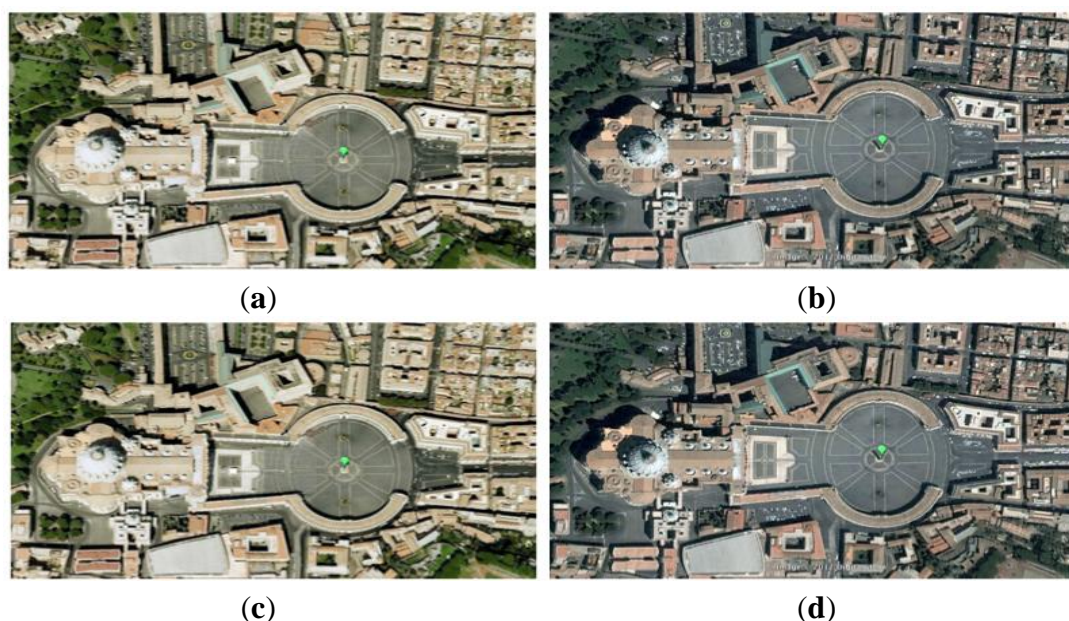
A review of the existing literature related to determining the positional accuracy of vector data has provided insight into key research performed on this topic. Goodchild and Hunter [13] suggest that the positional accuracy of spatial objects can be defined through measures of the differences between the apparent location of the feature recorded in the database and the feature's true location. Different metrics describing positional accuracy are presented depending on the geometry of the feature being tested, e.g., the accuracy of points are usually expressed as RMSE error estimates, while linear features could be compared using the buffer method. Kiiveri [14] suggests the need for objective methods for assessing, representing, and transmitting uncertainty in vector data through calculations so decision makers have some idea of the reliability of the information. Positional accuracy can be separated into two classes, absolute and relative [15], where relative positional accuracy describes the consistency of any position on a map with respect to any other and absolute accuracy is a measure of deviation of an estimate from the true value [16].

Spatial data quality research begins with the acknowledgement that the data stored in a geographic database are rarely, if ever, truly free of error and the database contains an approximation of the real world [17]. Doucette *et al.* [18] suggest that uncertainty handling is relevant to any scientific activity that involves making measurements of real world phenomena. Studies to describe the positional accuracy of geographical vector databases have been conducted where a test dataset is compared with a truth dataset of higher positional accuracy [19–23]. While other studies compared a test vector dataset with GPS locations [24,25] and positions determined from georeferenced orthoimagery [26].

Haklay [21,27] has presented landmark papers on the OSM project and determining the positional accuracy of a portion of the database in London, England. Road centerlines from OSM are compared to their corresponding location in the UK's Ordnance Survey (OS) using the buffering method. The results show positional accuracies ranging from 3.17 m to 8.33 m with the average difference at 5.8 m.

One of the most common ways to build vector datasets is by digitizing, tracing, or vectorizing a raster aerial image (both satellite and airborne based platforms). Since the resulting vector features will take on the coordinate reference frame of the aerial image, it is important to ensure the imagery is georeferenced properly. While image-processing methods can routinely provide georeferenced products with an accuracy of 2 m (CE90) for satellite based products [28], and 0.30 m (horizontal RMSE) for aerial based products [29], it is important to keep in mind that different processing techniques yield products with varying degrees of accuracies. This is apparent in Figure 1, which depicts a location in a 2007 Google Earth<sup>®</sup> image [30] as compared to the same location in older archived imagery. The results show positional differences ranging from 5 to 24 m.

**Figure 1.** Archived Google Earth<sup>®</sup> imagery of St. Peters Square, Vatican City. (a) 2007 reference image; (b) 2005 image 5-m different; (c) 2004 image 24-m different; (d) 2002 image 19-m different.



Geostatistics can be used to make predictions of sampled attributes at unsampled locations from sparse, often expensive data [31]. Burrough [31] defines two issues that geostatistical research can help address as: (1) interpolation errors and (2) error propagation in spatial models. Ever since spatial

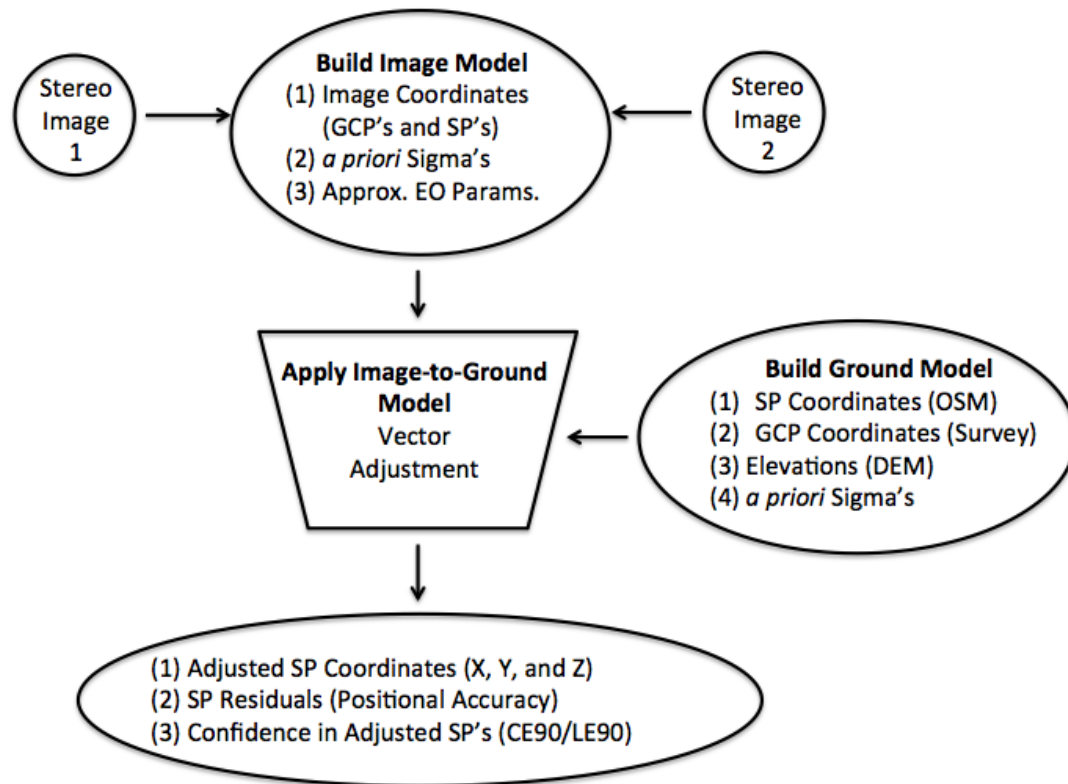
database accuracy was identified as a research initiative at the US National Center for Geographic Information and Analysis [32] an effort has been made to investigate various aspects of accuracy in spatial databases [33–37]. Shi *et al.* [38] contributes to the topic by presenting an S-Band model for describing the characteristics of positional uncertainty of geometric features, such as line segments, line features, boundary lines and area features. However, a key assumption is that the measurement errors are independent and uncorrelated. To address this, Shi and Liu [39] develop a G-Band model that handles correlated measurements. Love *et al.* [40] extend the G-Band concept with a Bayesian model to incorporate expert and historical knowledge that reduces the number of observations needed to make an accurate error analysis of vector data.

Even though work has been done on the subject of determining spatial accuracy of vector datasets, the current literature falls short when it comes to analyzing the positional accuracy using analytical photogrammetric techniques, such as the bundle adjustment. Photogrammetry has traditionally been used to: (1) determine three dimensional real world object space coordinates from stereo imagery; (2) propagate the error present in the sensor imaging system to a ground location; (3) derive three dimensional terrain surfaces, such as digital elevation models; or (4) extend highly accurate GCP's to adjacent image strips or blocks to facilitate the accurate mapping of the terrain and infrastructure from aerial imagery. Therefore, it is conceivable that photogrammetric techniques could be used to address the problem of determining the positional accuracy of geographical vector data by developing an analytical stereo model based on accurate GCP's and image coordinates to estimate the true location of the vector shape points by enforcing the collinearity equations.

### 3. Methodology

#### 3.1. Vector Adjustment Concept

The vector adjustment model used to determine the positional accuracy of road vectors is built upon the photogrammetric bundle adjustment model and extended to include vector shape points as the object points. Since the object points are expressed in three dimensions, the elevations for the shape points are estimated using a 1/3 arc second raster Digital Elevation Model (DEM) obtained over the area of interest from the USGS National Elevation Dataset (NED) [41]. For this research, road features from the OSM database and surveyed GCP's were used as the object points in the adjustment to determine the positional accuracy of the vector shape points. The general idea is to use heavily weighted (well-known) GCP's and image coordinates to establish the sensor's exposure station and enforce the collinearity concept to solve for the true location of the OSM shape points. The adjustment residuals then describe the difference between the adjusted ("true") location and the OSM database position. In addition, since this is a rigorous application, where the uncertainties in GCP's, image coordinates, and object points are known, error propagation is used to compute confidence regions about the adjusted shape point locations (CE90 and LE90). The overall concept is depicted in Figure 2.

**Figure 2.** Schematic diagram of the vector adjustment concept.

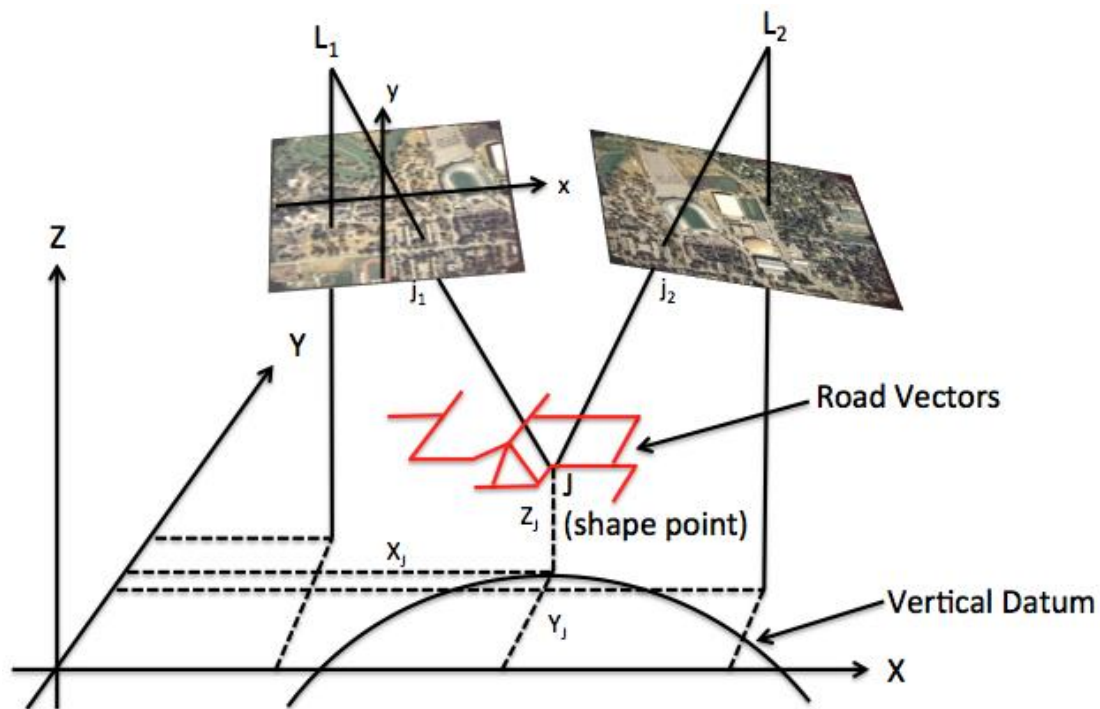
In Figure 2 the *SP* represents the OSM shape points from the geographical database that make up a particular road vector being tested. Image coordinates of the OSM shape points are measured on each stereo image and used as input to the adjustment model. The image coordinates are considered to be well known (heavily weighted) and will ultimately control the adjusted shape point location in ground space. The assumption here is that the imagery is considered to be “truth”, either more updated or of higher importance than the OSM database location. In addition, initial approximations for the exterior orientation (EO) parameters are provided to facilitate the adjustment process, note the EO parameters are solved for in the adjustment so an initial estimate is all that is needed.

Well-known (heavily weighted) GCP's are considered the absolute control for the adjustment. GCP's are measured in the imagery and used with their ground space coordinates in the adjustment. Since the GCP's are used to formulate the analytical stereo model and to create the image-to-ground relationship, it is important that they be known to a high degree of accuracy. Furthermore, it is essential for the GCP's and OSM shape points to be referenced to the same horizontal/vertical datum and map projection to minimize any misalignment that could otherwise be removed.

The outputs from the vector adjustment are: (1) adjusted OSM shape point locations, (2) shape point adjustment residuals (which describe the positional accuracy), and (3) confidence regions about the adjusted shape point locations (CE90/LE90).

The relationship between the stereo images and an OSM shape point is an extension to the space resection problem [42,43]; the geometric model can be visualized in Figure 3.



**Figure 3.** Geometric diagram of the vector adjustment concept.

where,  $L_1$  and  $L_2$  are the exposure station coordinates for stereo image one and two and  $j_1$  and  $j_2$  are measured image coordinates of the OSM shape point,  $J$ , on images one and two. For this work, the measuring of image coordinates was done manually, *i.e.*, a user measures each shape point in the stereo imagery. However, it is anticipated that this process could be automated by incorporating an image-to-vector registration process. Once the image and object space coordinates are estimated they can be used as initial approximation in the bundle adjustment model. The collinearity condition, which states that the exposure station ( $X_{L_i}, Y_{L_i}, Z_{L_i}$ ), an object points image derived location ( $x_j, y_j$ ), and the object point in ground space ( $X_J, Y_J, Z_J$ ) all lie on a single line is used to form the observation equations in the bundle adjustment.

### 3.2. Photogrammetric Bundle Adjustment

The basic geometric unit in photogrammetry is the image ray, an image can be thought of as a bundle of rays converging at the perspective center with an unknown position and orientation in space [42]. The bundles from all photos are adjusted simultaneously so that corresponding light rays from the measured image coordinates intersect at positions of the object points on the ground. The unknown quantities to be obtained from the bundle adjustment consist of the adjusted object coordinates of the vector shape points and GCP's ( $X_J, Y_J, Z_J$ ) and the EO parameters ( $\omega, \phi, \kappa, X_{L_i}, Y_{L_i}, Z_{L_i}$ ) of all the images. The EO parameters describe the exposure station coordinates ( $X_{L_i}, Y_{L_i}, Z_{L_i}$ ) and the image orientation parameters ( $\omega, \phi, \kappa$ ) that are used in the collinearity equations to formulate the image-to-ground relationship.

According to Wolf and Dewitt [43] the bundle block adjustment can be formulated using the collinearity equations, which are the foundation of the observation equations and used to form the mathematical model. The collinearity equations are documented in Equations (1) and (2) as:

$$x_j = x_o - f \left[ \frac{m_{11i}(X_J - X_{Li}) + m_{12i}(Y_J - Y_{Li}) + m_{13i}(Z_J - Z_{Li})}{m_{31i}(X_J - X_{Li}) + m_{32i}(Y_J - Y_{Li}) + m_{33i}(Z_J - Z_{Li})} \right] \quad (1)$$

$$y_j = y_o - f \left[ \frac{m_{21i}(X_J - X_{Li}) + m_{22i}(Y_J - Y_{Li}) + m_{23i}(Z_J - Z_{Li})}{m_{31i}(X_J - X_{Li}) + m_{32i}(Y_J - Y_{Li}) + m_{33i}(Z_J - Z_{Li})} \right] \quad (2)$$

where,  $x_j$  and  $y_j$  are the measured image coordinates of  $J$ ;  $X_J$ ,  $Y_J$  and  $Z_J$  are the coordinates of object point  $J$ ,  $X_{Li}$ ,  $Y_{Li}$  and  $Z_{Li}$  are coordinates of the image exposure station;  $x_o$  and  $y_o$  are the coordinates of the principle point known from the camera calibration report;  $f$  is the focal length of the camera (also known from the camera calibration report); and  $m_{11i}$ ,  $m_{12i}$ , ...,  $m_{33i}$  are the rotation matrix terms formulated in Equation (4).

The rotation matrix terms are a result of individual rotations of  $\omega$ ,  $\phi$ , and  $\kappa$  being applied to rotations about the  $x$ ,  $y$ , and  $z$ -axes, respectively. The individual rotation matrices are structured as follows:

$$\begin{aligned} M_\omega &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \omega & \sin \omega \\ 0 & -\sin \omega & \cos \omega \end{bmatrix} \\ M_\phi &= \begin{bmatrix} \cos \phi & 0 & -\sin \phi \\ 0 & 1 & 0 \\ \sin \phi & 0 & \cos \phi \end{bmatrix} \\ M_\kappa &= \begin{bmatrix} \cos \kappa & \sin \kappa & 0 \\ -\sin \kappa & \cos \kappa & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned} \quad (3)$$

With the total rotation matrix is a result of combining the individual rotations as follows:

$$M = M_\kappa M_\phi M_\omega \quad (4)$$

Since the collinearity equations are nonlinear, they are linearized by applying the first-order terms of the Taylor's series at a set of initial approximations, a detailed derivation is provided in [43].

The equations used to setup and solve the bundle adjustment have been well documented in several textbooks [42,43]. However, it is worth reviewing the basic mathematical model and matrices used to solve the system of linear equations. The general form is based on the unified least squares model and expressed as:

$$\dot{B}_{ij} \dot{\Delta}_i + \ddot{B}_{ij} \ddot{\Delta}_j = \varepsilon_{ij} + V_{ij} \quad (5)$$

where,  $\dot{B}_{ij}$  is a matrix of partial derivatives of the observation equations with respect to the exterior image orientation parameters ( $\omega$ ,  $\phi$ ,  $\kappa$ ,  $X_{Li}$ ,  $Y_{Li}$ ,  $Z_{Li}$ );  $\dot{\Delta}_i$  is a vector of corrections to the exterior orientation parameters;  $\ddot{B}_{ij}$  is a matrix of partial derivatives of the observation equations with respect to the object point coordinates ( $X_J$ ,  $Y_J$ ,  $Z_J$ );  $\ddot{\Delta}_j$  is a vector of corrections to the object point coordinates (OSM shape points and GCP's);  $\varepsilon_{ij}$  is the misclosure vector which is used to minimize the sum of the squared residuals;  $V_{ij}$  is the vector of residuals for the measured image coordinates. Adjustment residuals are important because they describe the difference between the measured and the adjusted values, which is a good indicator of how much the adjustment moved the input measurements as a result of the adjustment process.



The matrices used in the least squares model are structured as follows, for simplification the matrices are shown for a single image  $i$  and object point  $j$ . The  $\dot{B}_{ij}$  and  $\ddot{B}_{ij}$  matrices are made up of terms resulting from linearizing the collinearity equations in [43].

$$\dot{B}_{ij} = \begin{bmatrix} b_{11ij} & b_{12ij} & b_{13ij} & -b_{14ij} & -b_{15ij} & -b_{16ij} \\ b_{21ij} & b_{22ij} & b_{23ij} & -b_{24ij} & -b_{25ij} & -b_{26ij} \end{bmatrix} \quad (6)$$

The  $\dot{B}_{ij}$  matrix as:

$$\ddot{B}_{ij} = \begin{bmatrix} b_{14ij} & b_{15ij} & b_{16ij} \\ b_{24ij} & b_{25ij} & b_{26ij} \end{bmatrix} \quad (7)$$

The  $\dot{\Delta}_i$  vector as:

$$\dot{\Delta}_i = \begin{bmatrix} d\omega_i \\ d\phi_i \\ d\kappa_i \\ dX_{L_i} \\ dY_{L_i} \\ dZ_{L_i} \end{bmatrix} \quad (8)$$

The  $\ddot{\Delta}_j$  vector as:

$$\ddot{\Delta}_j = \begin{bmatrix} dX_j \\ dY_j \\ dZ_j \end{bmatrix} \quad (9)$$

The  $\varepsilon_{ij}$  vector is formulated with Taylor Series expansion terms in [43] as:

$$\varepsilon_{ij} = \begin{bmatrix} J_{ij} \\ K_{ij} \end{bmatrix} \quad (10)$$

And the  $V_{ij}$  vector as:

$$V_{ij} = \begin{bmatrix} v_{x_{ij}} \\ v_{y_{ij}} \end{bmatrix} \quad (11)$$

The adjusted object points are expressed as a function of the initial measurement and the adjustment residual through the following relationship:

$$\begin{aligned} X_j &= X_j^{oo} + v_{X_j} \\ Y_j &= Y_j^{oo} + v_{Y_j} \\ Z_j &= Z_j^{oo} + v_{Z_j} \end{aligned} \quad (12)$$

where,  $X_j$ ,  $Y_j$  and  $Z_j$  are the unknown coordinates of point  $j$ ;  $X_j^{oo}$ ,  $Y_j^{oo}$  and  $Z_j^{oo}$  are the measured coordinate value for point  $j$ ; and  $v_{X_j}$ ,  $v_{Y_j}$  and  $v_{Z_j}$  are the coordinate residuals for point  $j$ . To be consistent with the collinearity equations the object point observation equations will need to be evaluated at initial approximations as follows:

$$\begin{aligned} X_j^o + dX_j &= X_j^{oo} + v_{X_j} \\ Y_j^o + dY_j &= Y_j^{oo} + v_{Y_j} \\ Z_j^o + dZ_j &= Z_j^{oo} + v_{Z_j} \end{aligned} \quad (13)$$

With  $X_j^o$ ,  $Y_j^o$  and  $Z_j^o$  being the initial approximations for the coordinates of point  $j$  and  $dX_j$ ,  $dY_j$ ,  $dZ_j$  being the corrections to the approximations for coordinate point  $j$ , as solved for in the  $\ddot{\Delta}_j$  matrix. Since the collinearity equations are nonlinear an iterative approach is used where corrections to the parameters are solved for and applied each time through the loop until the difference in the correction is very small or essentially unchanged. This condition is referred to as convergence. Simplifying and expressing in matrix form yields:

$$\ddot{\Delta}_j = \ddot{C}_j + \ddot{V}_j \quad (14)$$

With  $\ddot{C}_j$  being the difference between the current approximation and the initial approximation.

$$\ddot{C}_j = \begin{bmatrix} X_j^{oo} - X_j^o \\ Y_j^{oo} - Y_j^o \\ Z_j^{oo} - Z_j^o \end{bmatrix} \quad (15)$$

And,  $\ddot{V}_j$  being the residual for the object point  $j$  expressed in each component.

$$\ddot{V}_j = \begin{bmatrix} v_{X_j} \\ v_{Y_j} \\ v_{Z_j} \end{bmatrix} \quad (16)$$

Weights for the object point coordinates are also used in the adjustment and based on the accuracy of the GCP's and vector shape points. The weights for  $X_j$ ,  $Y_j$ , and  $Z_j$  object point  $j$  are expressed in matrix form as

$$\ddot{W}_j = \sigma_o^2 \begin{bmatrix} \sigma_{X_j}^2 & \sigma_{X_j Y_j} & \sigma_{X_j Z_j} \\ \sigma_{Y_j X_j} & \sigma_{Y_j}^2 & \sigma_{Y_j Z_j} \\ \sigma_{Z_j X_j} & \sigma_{Z_j Y_j} & \sigma_{Z_j}^2 \end{bmatrix}^{-1} \quad (17)$$

where  $\sigma_o^2$  is the *a priori* reference variance;  $\sigma_{X_j}^2$ ,  $\sigma_{Y_j}^2$ , and  $\sigma_{Z_j}^2$  are the variances in  $X_j^{oo}$ ,  $Y_j^{oo}$ , and  $Z_j^{oo}$ , respectively; and the off diagonal terms are correlation coefficients. The final types of observations used in the bundle adjustment are the EO parameters. Their observation equations take on a form similar to the object points and are given as:

$$\begin{aligned} \omega_i &= \omega_i^{oo} + v_{\omega_i} \\ \phi_i &= \phi_i^{oo} + v_{\phi_i} \\ \kappa_i &= \kappa_i^{oo} + v_{\kappa_i} \\ X_{L_i} &= X_{L_i}^{oo} + v_{X_{L_i}} \\ Y_{L_i} &= Y_{L_i}^{oo} + v_{Y_{L_i}} \\ Z_{L_i} &= Z_{L_i}^{oo} + v_{Z_{L_i}} \end{aligned} \quad (18)$$

And the weight matrix for the EO parameters for a single image  $i$  is structured as:

$$\dot{W}_i = \begin{bmatrix} \sigma_{\omega_i}^2 & \sigma_{\omega_i\phi_i} & \sigma_{\omega_i\kappa_i} & \sigma_{\omega_iX_{L_i}} & \sigma_{\omega_iY_{L_i}} & \sigma_{\omega_iZ_{L_i}} \\ \sigma_{\phi_i\omega_i} & \sigma_{\phi_i}^2 & \sigma_{\phi_i\kappa_i} & \sigma_{\phi_iX_{L_i}} & \sigma_{\phi_iY_{L_i}} & \sigma_{\phi_iZ_{L_i}} \\ \sigma_{\kappa_i\omega_i} & \sigma_{\kappa_i\phi_i} & \sigma_{\kappa_i}^2 & \sigma_{\kappa_iX_{L_i}} & \sigma_{\kappa_iY_{L_i}} & \sigma_{\kappa_iZ_{L_i}} \\ \sigma_{X_{L_i}\omega_i} & \sigma_{X_{L_i}\phi_i} & \sigma_{X_{L_i}\kappa_i} & \sigma_{X_{L_i}}^2 & \sigma_{X_{L_i}Y_{L_i}} & \sigma_{X_{L_i}Z_{L_i}} \\ \sigma_{Y_{L_i}\omega_i} & \sigma_{Y_{L_i}\phi_i} & \sigma_{Y_{L_i}\kappa_i} & \sigma_{Y_{L_i}X_{L_i}} & \sigma_{Y_{L_i}}^2 & \sigma_{Y_{L_i}Z_{L_i}} \\ \sigma_{Z_{L_i}\omega_i} & \sigma_{Z_{L_i}\phi_i} & \sigma_{Z_{L_i}\kappa_i} & \sigma_{Z_{L_i}X_{L_i}} & \sigma_{Z_{L_i}Y_{L_i}} & \sigma_{Z_{L_i}}^2 \end{bmatrix}^{-1} \quad (19)$$

The weights for the  $x$  and  $y$  image coordinates are expressed in the matrix form for an object point  $j$  on photo  $i$  as:

$$W_{ij} = \sigma_o^2 \begin{bmatrix} \sigma_{x_{ij}}^2 & \sigma_{x_{ij}y_{ij}} \\ \sigma_{y_{ij}x_{ij}} & \sigma_{y_{ij}}^2 \end{bmatrix}^{-1} \quad (20)$$

where,  $\sigma_o^2$  is the *a priori* reference variance;  $\sigma_{x_{ij}}^2$  and  $\sigma_{y_{ij}}^2$  are the variances in,  $x_{ij}$  and  $y_{ij}$ , respectively; and the covariance  $\sigma_{x_{ij}y_{ij}} = \sigma_{y_{ij}x_{ij}}$  are assumed to be uncorrelated. After the individual matrices for the observation equations are formed the full set of normal equations can be structured in matrix form as:

$$N\Delta = K \quad (21)$$

where the matrices are formatted as:

$$N = \begin{bmatrix} \dot{N}_1 + \dot{W}_1 & 0_6^6 & 0_6^6 & \dots & \bar{N}_{11} & \bar{N}_{12} \dots & \bar{N}_{1n} \\ 0_6^6 & \dot{N}_2 + \dot{W}_2 & 0_6^6 & \dots & \bar{N}_{21} & \bar{N}_{22} \dots & \bar{N}_{2n} \\ 0_6^6 & 0_6^6 & \dot{N}_m + \dot{W}_m & \dots & \bar{N}_{m1} & \bar{N}_{m2} \dots & \bar{N}_{m3} \\ \bar{N}_{11}^T & \bar{N}_{21}^T & \bar{N}_{31}^T & \dots & \ddot{N}_1 + \ddot{W}_1 & 0_3^3 & 0_3^3 \\ \bar{N}_{12}^T & \bar{N}_{22}^T & \bar{N}_{32}^T & \dots & 0_3^3 & \ddot{N}_2 + \ddot{W}_2 & 0_3^3 \\ \bar{N}_{1n}^T & \bar{N}_{2n}^T & \bar{N}_{3n}^T & \dots & 0_3^3 & 0_3^3 & \ddot{N}_n + \ddot{W}_n \end{bmatrix} \quad (22)$$

And the  $\Delta$  matrix being a combination of the corrections to the exterior orientation parameters and the corrections to the object point coordinates, with the size being determined by how many images and object points are in the adjustment.

$$\Delta = \begin{bmatrix} \dot{\Delta}_1 \\ \dot{\Delta}_2 \\ \dot{\Delta}_3 \\ \vdots \\ \dot{\Delta}_m \\ \ddot{\Delta}_1 \\ \ddot{\Delta}_2 \\ \ddot{\Delta}_3 \\ \vdots \\ \ddot{\Delta}_n \end{bmatrix} \quad (23)$$

Similarly, the  $K$  matrix is structured as:

$$K = \begin{bmatrix} \dot{K}_1 + \dot{W}_1 \dot{C}_1 \\ \dot{K}_2 + \dot{W}_1 \dot{C}_2 \\ \dot{K}_3 + \dot{W}_1 \dot{C}_3 \\ \vdots \\ \dot{K}_m + \dot{W}_m \dot{C}_m \\ \ddot{K}_1 + \ddot{W}_1 \ddot{C}_1 \\ \ddot{K}_2 + \ddot{W}_2 \ddot{C}_2 \\ \ddot{K}_3 + \ddot{W}_3 \ddot{C}_3 \\ \vdots \\ \ddot{K}_n + \ddot{W}_n \ddot{C}_n \end{bmatrix} \quad (24)$$

The submatrices used above are defined as:

$$\dot{N}_i = \sum_{j=1}^n \dot{B}_{ij}^T W_{ij} \dot{B}_{ij} \quad (25)$$

$$\bar{N}_{ij} = \dot{B}_{ij}^T W_{ij} \ddot{B}_{ij} \quad (26)$$

$$\dot{N}_j = \sum_{i=1}^m \ddot{B}_{ij}^T W_{ij} \ddot{B}_{ij} \quad (27)$$

$$\dot{K}_i = \sum_{j=1}^n \dot{B}_{ij}^T W_{ij} \varepsilon_{ij} \quad (28)$$

$$\ddot{K}_j = \sum_{i=1}^m \ddot{B}_{ij}^T W_{ij} \varepsilon_{ij} \quad (29)$$

With  $m$  being the number of images,  $n$  is the number of object points,  $i$  is the image subscript, and  $j$  is the object point subscript. If a point  $j$  does not appear on image  $i$  than a zero submatrix is used. When the normal equations are being formed it is recommended to compute the *a posteriori* reference variance, which is a unit less scalar quantity that describes the uncertainty found in the observations post adjustment. It is a function of the various weight matrices (image coordinate accuracies, EO parameter accuracies, and the object point accuracies) propagated into the misclosure of the collinearity and object point observation equations. The *a posteriori* standard error of unit weight can be computed as:

$$S_o = \sqrt{\frac{\sum_{j=1}^n \sum_{i=1}^m \varepsilon_{ij}^T W_{ij} \varepsilon_{ij} + \sum_{i=1}^m \dot{C}_i^T \dot{W}_i \dot{C}_i + \sum_{j=1}^n \ddot{C}_j^T \ddot{W}_j \ddot{C}_j}{n.o. - n.u.}} \quad (30)$$

where  $n.o.$  is the total number of observations and  $n.u.$  is the total number of unknowns in the adjustment model. Once the solution converges the normal equations are then scaled by the *a posteriori* reference variance to compute the final variance co-variance matrix for the adjustable parameters as:

$$\Sigma_{\Delta\Delta} = S_o^2 N^{-1} \quad (31)$$

where the resulting matrix is block diagonal with variances of the exterior orientation parameters and the object point coordinates consistent with how the  $N$  matrix was formed. The standard deviations of the adjustment parameters are computed by taking the square root of the diagonal variances.

### 3.3. Adjustment Weighting

The concept of weighting in the vector adjustment is very important and directly impacts the adjusted values, as well as the estimated uncertainty of the adjusted points. Adjustment weighting consists of assigning a numerical value to the adjustment observations and parameters based on how well the specific quantity is known. Weights are determined for both: (1) the image points, based on an *a priori* estimate of how well the points can be mensurated in the image, and (2) the object point coordinates, which consist of the GCP's and vector shape point accuracies.

The general idea is the higher confidence a user has in a measurement the smaller the standard deviation will be assigned to that measurement. For example, the GCP's used in this study were established by GPS survey, therefore the accuracies assigned to these coordinates are very small, on the order of 10-cm, or less three dimensionally. One can reason that a small standard deviation corresponds to a high user confidence in the numerical value of the measurement. On the other hand, the vector shape point coordinates were determined with less stringent methods; such as measuring a single satellite image or with a hand held GPS receiver. In this case the coordinates are much less well known, *i.e.*, a user has less confidence in the accuracy of the point, so a standard deviation on the order of a few meters or more could be realistic.

In most cases the accuracy of the OSM shape points will not be known *a priori*, therefore the following approach can be implemented to estimate them. First, assign an arbitrary large standard deviation to the OSM shape points. Second, run an adjustment to determine the shape point residuals and compute the RMSE of the shape point residuals to estimate the positional displacement. Thirdly, use the RMSE value and residuals to estimate *a priori* standard deviations for the OSM shape points. Lastly, rerun the adjustment to compute accurate adjusted information.

### 3.4. Absolute Accuracy

Accuracy is a term that refers to the closeness between measurements and their true values [44]. The closer a measurement is to the true value the better accuracy the measurement has. In reality, the true value is not known and can only be estimated by making measurements and analyzing those measurements. Therefore, the true value is often referred to as the expected value and can be measured by computing the average or mean of a subset of measurements. Additionally, the measurements themselves are not perfect quantities and are subject to errors resulting from the people making the measurements, the equipment used to make the measurement, and random noise in the measurements themselves.

Product accuracy statistics summarize the dispersion of the individual errors of a large set of checkpoints. The military community requires an accuracy statistic that combines biases and random errors and then estimates them at the 90% probability level. Vertical accuracy is reported as a linear error (LE) because it is an uncertainty along the single vertical accuracy. Horizontal accuracy is a function of the two horizontal dimensions in the x and y directions. It can be considered a circular error

(CE), which refers to the radius of the circle, centered about the derived location, within which the true or expected location of points lies [45]. Ager derives the CE90 and LE90 in detail in [45], so the reader is referred here for more information.

#### Root Mean Square Error (RMSE)

The Root Mean Square Error (RMSE) is defined as the square root of the average squared discrepancies in coordinate values [6]. The RMSE is made up of the mean square error (MSE) in the various  $x$ ,  $y$ , and  $z$  components. The MSE is a value measuring the sum of the squared differences between a measured value and its truth-value and is defined in [44] as:

$$MSE = \frac{\sum (x - t)^2}{n} \quad (32)$$

where,  $x$  is the measured value,  $t$  is the truth-value, and  $n$  is the number of measurements. In reality, the truth-value is rarely known, so this work seeks to replace truth with the best estimate of the measured value. For example, the GPS established ground control points used in the bundle adjustment are considered truth, while their values were actually determined by measurement with inherent uncertainty. However, the positions are the best estimate of truth and will be used as such. Root mean square positional error can then be defined by taking the square root of the MSE in the  $x$  and  $y$  directions.

$$RMSE = \sqrt{MSE_x + MSE_y} \quad (33)$$

The RMSE statistic can then be used to describe the uncertainty present in a set of like measurements.

#### 4. Experimental Results and Analysis

The experimental results presented here are meant to test the accuracy of the proposed vector adjustment model and how well it recovers the actual positional displacement present in the input vectors. Shape points from the OSM database will be used, along with aerial stereo imagery, a USGS 1/3 arc second DEM, and GCP's as input to the vector adjustment model. The positional accuracy will be measured by comparing shape point coordinate differences of the OSM database and an adjusted shape point location to the real world ground surveyed location (ground truth). For example, the difference between the OSM database and the surveyed location represents the actual positional displacement present in the shape point; similarly, measuring the difference between the adjusted shape point and the surveyed location describes the accuracy of the adjustment itself. In addition, comparing the adjustment residuals to those determined from ground truth will provide a measure of performance to how well the vector adjustment model recovers the true positional displacement. The bundle adjustment routine in the SOCET GXP<sup>®</sup> (version 3.2) photogrammetry and geospatial exploitation software by BAE Systems Corporation was used to facilitate the adjustment testing, while the RMSE metric was used to quantify the results.



#### 4.1. Project Datasets

The aerial imagery used for this project is aerial frame stereo imagery taken with a Leica (Wild) RC10 aerial photography camera (sensor) over the Purdue University campus in West Lafayette, Indiana on 5 October 1999. The imagery has been scanned into digital format and has a scale of 1:4,000 or 1" = 100 m with a 12 cm ground sample distance (GSD). The altitude of the aircraft was 610 m above the terrain or 798 m above mean seal level and was collected to include 80% forward overlap and 60% side overlap. The sensor was calibrated by the USGS and includes a camera calibration report, which was used for interior orientation and contains information such as calibrated focal length (152.4 mm), lens distortion parameters, and calibrated fiducial marks. No Inertial Navigation System (INS) data was collected during the imagery acquisition.

The DEM is used to estimate initial approximations for the shape point elevations, and consists of a 1/3 arc second (about 10-m) raster DEM in GeoTiff format [41]. The NED is the primary elevation data product of the USGS and is a seamless dataset with the best available raster elevation data of the conterminous United States, Alaska, Hawaii, and territorial islands. The NED is derived from diverse source data that are processed to a common coordinate system and unit of vertical measure, being geographic coordinates in units of decimal degrees, and in conformance with the North American Datum of 1983 (NAD 83). All elevation values are in meters and, over the conterminous United States, are referenced to the North American Vertical Datum of 1988 (NAVD 88). NED data is available nationally (except for Alaska) at resolutions of 1 arc-second (about 30 m) and 1/3 arc-second (about 10 m), and in limited areas at 1/9 arc-second (about 3 m) [41]. Care was taken to ensure the appropriate conversion between the NAD 83 and WGS 84 datum were applied when working with the dataset.

The truth dataset was determined by locating the subject test vectors (road centerlines) with high order GPS surveying techniques. The field survey campaign was conducted in July of 2012 on the campus of Purdue University. The purpose of the survey was to locate the actual real world positions of the road centerlines being tested, as well as establish GCP's to facilitate the testing. The GPS equipment used was a Leica Viva RTK system and consisted of a GPS receiver (rover) connected to the Indiana Department of Transportation (INDOT) INWL Continuously Operating Reference Station (CORS- base station) via a digital data modem. This type of setup allows the user to connect directly to a very high accuracy "zero order" control point and receive adjusted survey grade coordinates at the rover in real time.

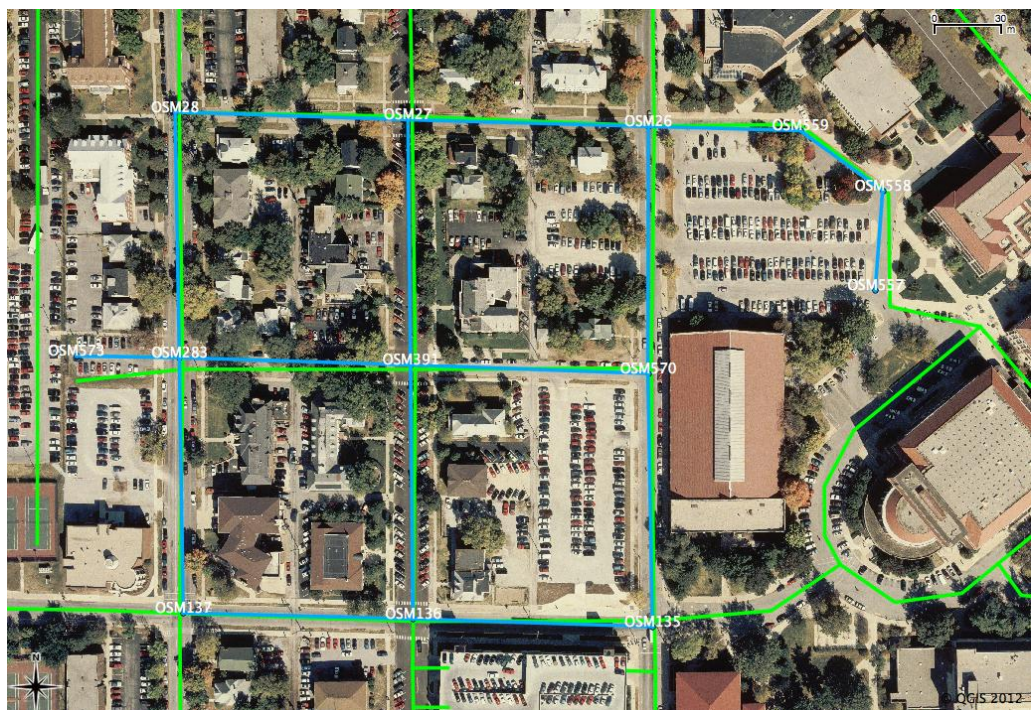
#### 4.2. OSM Urban City Streets Scenario

The city streets scenario is meant to test urban roads in the OSM database for positional accuracy by measuring the intersections of gridded street vectors. This scenario is located near the center of the campus and covers two-square blocks (1 block =  $\pm 100$  m) along 3rd, 4th, 5th, Russell, Waldron, and University streets. The adjusted vectors were derived with the proposed approach and can be visualized in Figure 4, where the original OSM database vectors are in green and the adjusted vectors in blue.

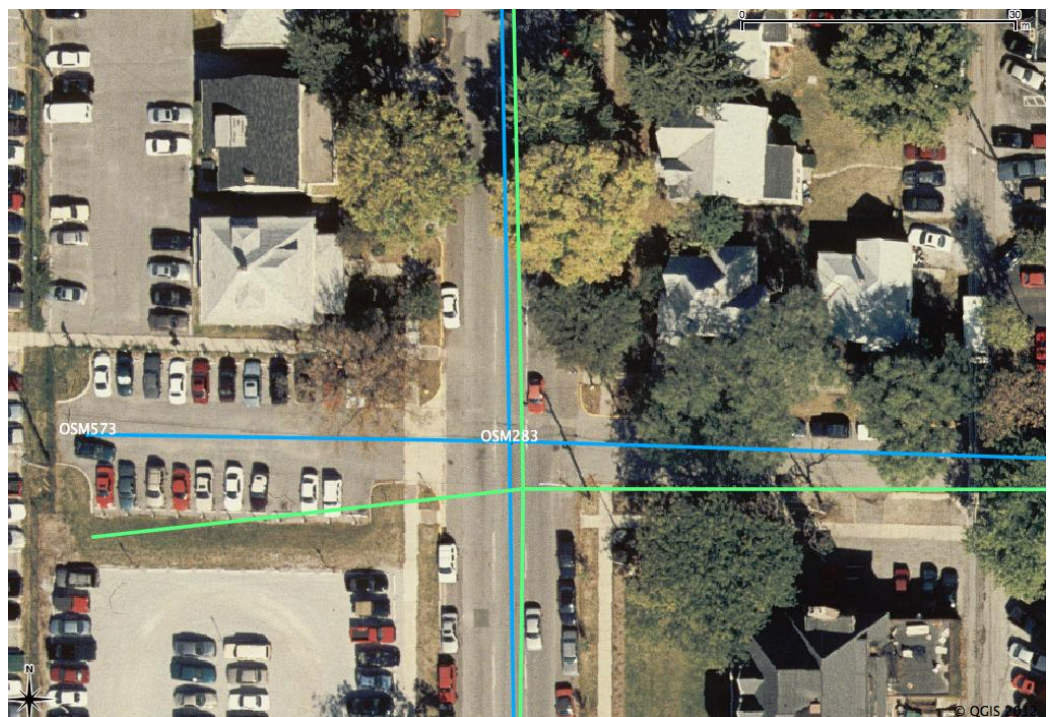
The base imagery has been georeferenced with surveyed ground control points so an accurate spatial relationship can be drawn between the vector data and the real world surroundings depicted in the imagery. A close up view of two areas is provided in Figures 5 and 6.

Visual inspection shows the adjusted vectors more closely match the centerline of the road in the imagery, as compared to the original OSM database vectors. The adjusted shape points coordinates and statistical confidences are shown in Table 1.

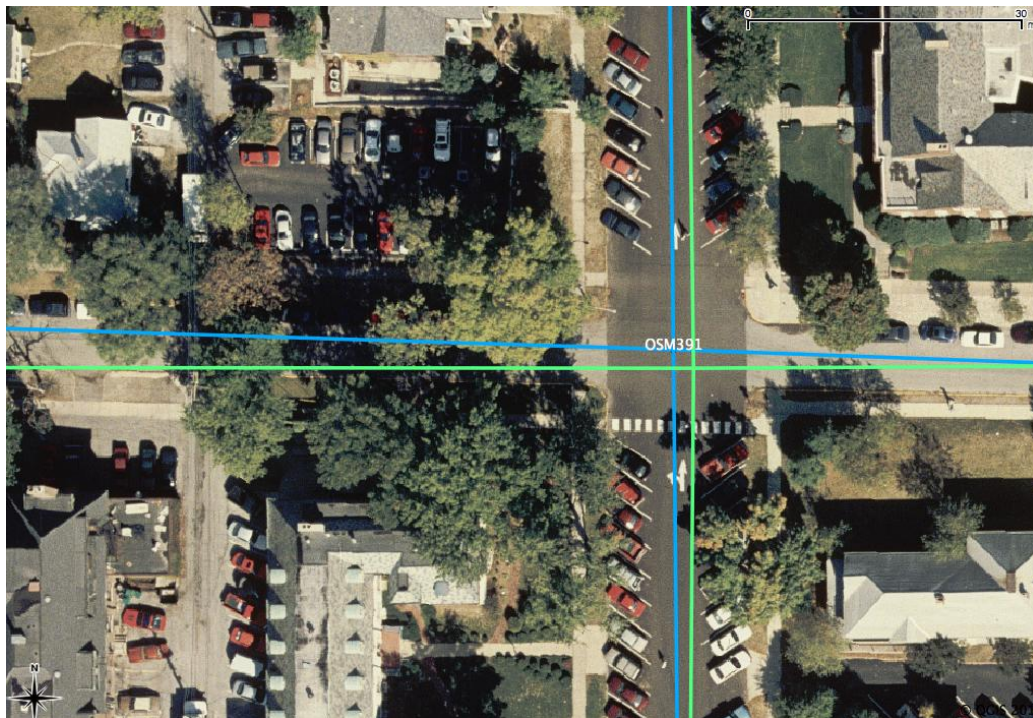
**Figure 4.** OpenStreetMap<sup>®</sup> (OSM) (green) and adjusted (blue) road vectors.



**Figure 5.** Close up area 1.





**Figure 6.** Close up area 2.**Table 1.** Adjusted shape point information.

| PT. ID. | CE90 (m) | LE90 (m) | X (m)      | Y (m)        | Z (m)  |
|---------|----------|----------|------------|--------------|--------|
| OSM135  | 0.12     | 0.20     | 507,069.35 | 4,475,181.34 | 188.66 |
| OSM136  | 0.13     | 0.20     | 506,961.99 | 4,475,184.75 | 188.76 |
| OSM137  | 0.13     | 0.20     | 506,859.00 | 4,475,187.43 | 190.16 |
| OSM26   | 0.15     | 0.20     | 507,067.46 | 4,475,405.79 | 189.65 |
| OSM27   | 0.15     | 0.20     | 506,959.67 | 4,475,408.83 | 189.56 |
| OSM28   | 0.16     | 0.20     | 506,855.31 | 4,475,412.24 | 189.86 |
| OSM283  | 0.14     | 0.20     | 506,857.23 | 4,475,301.99 | 190.06 |
| OSM391  | 0.13     | 0.20     | 506,960.84 | 4,475,298.73 | 189.16 |
| OSM557  | 0.14     | 0.20     | 507,169.15 | 4,475,332.21 | 189.77 |
| OSM559  | 0.15     | 0.20     | 507,135.45 | 4,475,403.94 | 189.87 |
| OSM570  | 0.13     | 0.20     | 507,067.27 | 4,475,294.90 | 189.16 |
| OSM573  | 0.15     | 0.20     | 506,810.98 | 4,475,302.72 | 190.36 |

Where, three GCP's were used to adjust the vector shape points and had *a priori* standard deviations of 2 cm in the  $x$  direction, 2 cm in the  $y$  direction, and 1 cm in the  $z$  direction. The reader is referred to [46] for additional information on the minimum number and configuration of GCP's needed to produce reliable adjustment results. The *a priori* standard deviations for the OSM shape points were significantly higher at 2, 3, and 0.20 m in the  $x$ ,  $y$ , and  $z$  directions, respectively. The *a posteriori* reference variance for this adjustment was 0.89, signifying the *a priori* weights going in are consistent with the uncertainty found in the observations post adjustment. The adjustment residuals are summarized in Table 2.

**Table 2.** Adjustment residuals.

| PT. ID. | Vx (m) | Vy (m) | Vz (m) |
|---------|--------|--------|--------|
| OSM135  | −0.76  | −2.61  | 0.00   |
| OSM136  | −0.17  | 1.08   | 0.01   |
| OSM137  | 1.55   | 0.21   | 0.00   |
| OSM26   | −2.74  | −0.20  | 0.00   |
| OSM27   | −1.96  | −1.46  | 0.00   |
| OSM28   | −1.90  | 0.18   | 0.01   |
| OSM283  | −1.55  | 5.20   | 0.01   |
| OSM391  | −2.22  | 2.03   | 0.00   |
| OSM557  | −6.67  | 7.26   | 0.00   |
| OSM559  | 0.56   | −2.71  | 0.00   |
| OSM570  | −1.52  | −2.42  | 0.00   |
| OSM573  | −0.65  | 11.19  | 0.01   |
|         | RMSE   | 5.04   |        |

The residuals indicate which way the OSM shape point had to move to become more positionally accurate, with the positive value denoting a point moving in the East (Vx) and North directions (Vy). It is interesting the vertical residuals are essentially zero, signifying no change from between the initial approximation (DEM) and the adjusted shape point elevations. In general, it is conceivable because the DEM was generated by photogrammetric methods and tied to the NAVD88 datum. Similarly, the GCP's used as the reference frame for this project were also tied to the NAVD88 datum. In addition, the intersections are fairly flat and wide paved areas, thereby increasing the accuracy of the DEM in these areas. However, keep in mind the LE90 from Table 1 is 0.20 m for these points. To investigate further the results are compared to surveyed ground truth in Table 3.

**Table 3.** Truth comparison.

| PT. ID. | Un-Adjusted    |                |                | Adjusted       |                |                |
|---------|----------------|----------------|----------------|----------------|----------------|----------------|
|         | $\Delta X$ (m) | $\Delta Y$ (m) | $\Delta Z$ (m) | $\Delta X$ (m) | $\Delta Y$ (m) | $\Delta Z$ (m) |
| OSM135  | 1.26           | 2.45           | 0.19           | 0.51           | −0.16          | 0.19           |
| OSM136  | 0.70           | −0.74          | 0.03           | 0.53           | 0.34           | 0.04           |
| OSM137  | −0.46          | −0.31          | 0.01           | 1.09           | −0.10          | 0.01           |
| OSM26   | 3.05           | 0.14           | 0.12           | 0.31           | −0.06          | 0.12           |
| OSM27   | 1.90           | 1.65           | 0.08           | −0.06          | 0.19           | 0.08           |
| OSM28   | 0.69           | 0.06           | 0.05           | −1.21          | 0.24           | 0.05           |
| OSM283  | 1.53           | −4.79          | 0.09           | 0.02           | 0.41           | 0.09           |
| OSM391  | 1.94           | −1.93          | 0.07           | −0.28          | 0.10           | 0.07           |
| OSM557  | 6.31           | −6.34          | 0.23           | −0.36          | 0.92           | 0.22           |
| OSM559  | −1.18          | 3.07           | 0.29           | −0.61          | 0.36           | 0.28           |
| OSM570  | 1.03           | 2.08           | 0.30           | −0.49          | −0.33          | 0.30           |
| OSM573  | 0.35           | −11.09         | 0.27           | −0.30          | 0.09           | 0.28           |
|         | RMSE           | 4.81           |                | RMSE           | 0.69           |                |

The Un-adjusted deltas refer to the OSM shape point minus the truth coordinate value, while the adjusted information refers to the adjusted shape point coordinate minus the ground truth. This type of comparison shows the roads actual positional displacement (un-adjusted) and what is left over after adjustment. The results show an 86%  $(100 - 0.69/4.81)$  improvement in positional accuracy (RMSE)

post adjustment. Furthermore, the vector adjustment was able to recover 95% (4.81/5.04) of the actual displacement present in the road shape points. The results also show that two of the points (OSM 137 and 28) actually moved slightly farther away from truth as a result of the adjustment, however one should remember the truth values were determined by estimating the center of a wide non-descript road intersection with traffic, making it difficult to measure the exact center of the roadways. So, it is possible the adjusted positions are closer to truth than the deltas initially suggest, especially when considering the CE90 of the two shape points at 13 and 16 cm, respectively.

The surveyed ground truth was determined in July of 2012, while the imagery was flown earlier in 1999. The challenge here was accounting for the construction that took place in the 12+ years in between. For example, it was evident that the roads associated with OSM shape points 557–573 have been rebuilt by the newer looking pavement, and it is possible there was a grade change associated with the reconstruction project that could account for the larger elevation differences seen in these points.

An additional check to ensure the adjustment produces reliable results was implemented. Several GCP's were included in the adjustment as “check points” only. Check points do not influence the outcome of the adjustment, rather their carried along and positioned using the computed model-to-ground relationship. The objective is to have known points that can be carried through the adjustment process to verify the performance of the vector adjustment. Table 4 contains the differences between the check points adjusted coordinates and the truth coordinates measured with GPS.

**Table 4.** Check point comparison.

| PT. ID. | $\Delta X$ (m) | $\Delta Y$ (m) | $\Delta Z$ (m) |
|---------|----------------|----------------|----------------|
| CKPT1   | −0.01          | 0              | 0              |
| CKPT2   | −0.12          | 0.08           | −0.12          |
| CKPT3   | 0.06           | −0.02          | 0.03           |
| CKPT4   | 0.15           | −0.27          | 0.08           |
| CKPT5   | −0.08          | 0.03           | −0.08          |
| CKPT6   | −0.09          | 0.02           | 0.06           |
|         | RMSE           | 0.15           |                |

The results in Table 4 show the RMSE of the check points to be 15 cm, which is consistent with the average CE90 of the vector shape points at 14 cm. This is another metric that provides confidence in adjusted shape point positions in Table 1.

#### 4.3. Feature Attributes

Implementing the proposed approach provides an adjusted road vector that is close to truth, as indicated from the scenario analysis above. The truth vector represents a location that is spatially accurate, in an absolute sense, in relationship to the actual features on the ground because it was derived by high-order geodetic surveying techniques (RTK GPS from a CORS) by a licensed professional surveyor. Therefore, since the adjusted position of the road vector is close to the truth vector, it can be considered to be “positionally accurate”. Following, having computed this valuable information it makes sense to store it at the feature level going forward so a user can easily

access and exploit the information to help them make more informed decisions about the data they are working with.

A feature level attribute refers to saving information about an individual geometric feature in a geographic database, e.g., recording the shape point adjustment residuals or a polylines RMSE value, both of which describes their positional accuracy. Example attributes for a shape point are seen in Table 5 with the adjusted information in italics.

**Table 5.** Shape point attributes.

| Attribute      | Value (m)           |
|----------------|---------------------|
| Type           | Shape Point         |
| ID             | OSM391              |
| X              | 506,963.06          |
| Y              | 4,475,296.70        |
| V <sub>x</sub> | −2.20               |
| V <sub>y</sub> | 2.04                |
| <i>Adj. X</i>  | <i>506,960.86</i>   |
| <i>Adj. Y</i>  | <i>4,475,298.74</i> |
| <i>Adj. Z</i>  | <i>189.17</i>       |
| <i>CE90</i>    | <i>0.13</i>         |
| <i>LE90</i>    | <i>0.20</i>         |

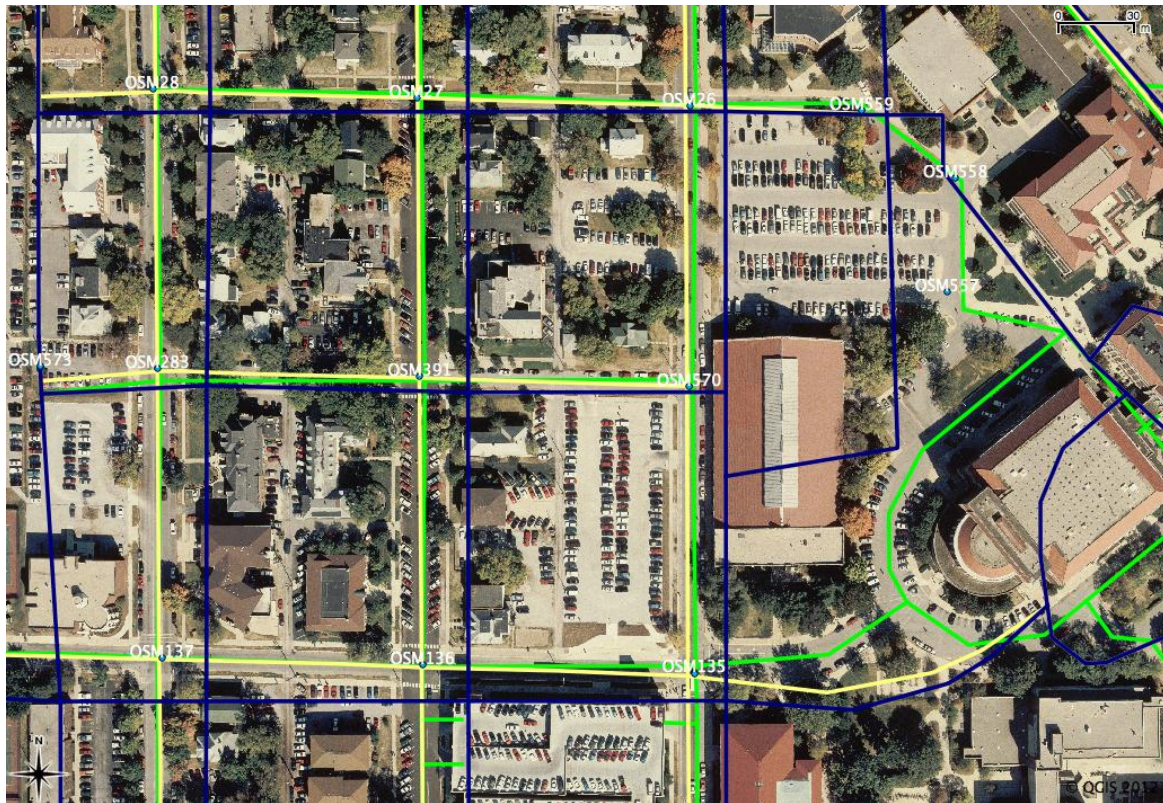
#### 4.4. Head-to-Head Accuracy Assessment of OSM, TNM, and TIGER 07 Roads

The head-to-head accuracy assessment of OSM, the USGS National Map (TNM), and TIGER 07 roads is meant to demonstrate a practical application of the proposed vector adjustment model. There are many applications that could benefit from knowing the positional accuracy of shape points and vectors in a geographical database, e.g., GPS navigation, vector-to-vector conflation problems, operational planning, and an improved decision making capability just to name a few. In this example, vector road centerlines and shape points were extracted from each database for the same location used in the urban city streets scenario above. Care was taken to ensure the datasets were referenced to the same horizontal and vertical datum, WGS84 and NAVD88 respectively. Each dataset was adjusted independently using the same GCP's with the same *a priori* standard deviations (5-cm in x/y and 8-cm in elevation) and same image coordinates for the shape points. The *a priori* standard deviations for the ground location of the shape points in each dataset were varied according to an estimate of the positional accuracy. Figure 7 depicts the datasets referenced to an accurately georeferenced base image.

In Figure 7 the OSM vectors are in green, TNM vectors in yellow, and the TIGER 07 vectors in blue. From visual inspection, the TIGER 07 lines appear to be shifted more than the OSM or TNM. However, keep in mind that a general user may not have base imagery or be comfortable with the accuracy of the georeferencing to make this judgment without the vector adjustment. The horizontal adjustment residuals for each dataset are summarized in Table 6.



**Figure 7.** OSM (green), USGS National Map (TNM) (yellow) and United States Census Bureau's Topologically Integrated Geographic Encoding Referencing (TIGER) 2007 (blue) roads.



**Table 6.** Adjustment residuals for OSM, TNM, and TIGER 07 roads.

| PT. ID. | OSM         |             | TNM    |             | TIGER 07 |              |
|---------|-------------|-------------|--------|-------------|----------|--------------|
|         | Vx (m)      | Vy (m)      | Vx (m) | Vy (m)      | Vx (m)   | Vy (m)       |
| OSM135  | −0.75       | −2.63       | 2.34   | 1.22        | −12.74   | 11.15        |
| OSM136  | −0.17       | 1.08        | 0.18   | 1.39        | −18.31   | 14.71        |
| OSM137  | 1.55        | 0.23        | 0.98   | 0.87        | −17.78   | 17.51        |
| OSM26   | −2.70       | −0.20       | 1.34   | −0.07       | −14.39   | 2.54         |
| OSM27   | −1.91       | −1.45       | −0.52  | 0.15        | −20.37   | 5.68         |
| OSM28   | −1.86       | 0.20        | −1.22  | 1.22        | −22.94   | 9.24         |
| OSM283  | −1.52       | 5.22        | −0.16  | 1.57        | −19.61   | 7.21         |
| OSM391  | −2.20       | 2.04        | −0.03  | 0.77        | −19.32   | 6.61         |
| OSM559  | 0.61        | −2.72       | 5.81   | 1.29        | −10.06   | 2.11         |
| OSM570  | −1.50       | −2.42       | 0.67   | 0.20        | −14.69   | 2.65         |
| OSM573  | −0.63       | 11.22       | −0.61  | 6.09        | −0.63    | 11.20        |
|         | <b>RMSE</b> | <b>4.35</b> |        | <b>2.89</b> |          | <b>19.17</b> |

A close up view of the center intersection showing the spatial relationship of the road centerlines is provided in Figure 8.

**Figure 8.** Close up view showing OSM, TNM, and TIGER 07 roads.

Analyzing the numbers suggests that TNM roads have the best positional accuracy, as indicated by the smallest RMSE estimate of 2.89 m, compared to OSM at 4.35 m and TIGER 07 at 19.17 m. This is also confirmed by comparing the shape point adjustment residuals, which shows that the TNM points moved less than the OSM and TIGER 07 roads. Note the RMSE value for OSM is slightly different than in the analysis above because two points were removed so an accurate comparison could be made between the three datasets. The TIGER 07 roads have the largest displacement, as suggested by visual inspection, with two of the shape points moving over 20 m in the west direction. Overall, the TIGER 07 roads appear to have a bias in the East and South directions, as indicated by the negative x-residuals and positive y-residuals.

## 5. Conclusions and Future Work

### 5.1. Conclusions

The proposed approach and vector adjustment model was developed to assess the positional accuracy of geographical vector data, such as road centerlines. The OSM database provides a unique dynamic environment to use as the test subject for this research because its underlining purpose of providing open source mapping by the people, to the people, suggests the importance of knowing how good the data is. For example, OSM contributors are mostly voluntary, non-professionals, who have an interest in mapping a local area they are familiar with. In addition, there is no cartographic or data quality standards in place to ensure contributors “map” in a similar fashion or adhere to any specific equipment requirements (GPS), field collection procedures, image mensuration standards, or map accuracy standards.



Most of the current methods of determining positional accuracy are based on comparing test vectors to a reference/truth dataset that is known to be of higher quality. However, these methods are not rigorous in nature (thoroughly modeling and propagating error in the system). To address this issue, the vector adjustment model presented here is based on applying rigorous photogrammetric positioning principles to vector data to determine the positional accuracy of shape points. Aerial imagery was used to build an analytical stereo model, with GCP's and vector shape points, to enforce the collinearity equations with a bundle adjustment. Post adjustment the vector shape points are transformed closer toward their "true" real world ground locations and the adjustment residuals describe the positional accuracy of the shape point. Furthermore, accuracy estimates (CE90 and LE90) are computed for the adjusted shape points, which afford a user confidence in the information coming out of the adjustment.

The proposed approach was tested on several urban gridded city streets from the OSM database. The results show that the post adjusted shape points improved positionally by 86%. Furthermore, the vector adjustment was able to recover 95% of the actual positional displacement present in the shape points. Once this valuable information is computed for the vector data it can be recorded as an attribute at the feature level, thereby improving the overall usefulness of the database and allowing a user to make more informed decisions based on the data. To demonstrate a practical application of the vector adjustment, it was used to characterize the positional accuracy of OSM, TNM, and TIGER 07 road vectors by comparing the RMSE values of the adjustment residuals.

## 5.2. Future Work and Recommendations

An application of this research would be to use the adjusted shape point residuals output from the vector adjustment model as input for a "smart conflation" procedure. For example, the scenario might include two road networks with varying attributes and geometries that could benefit from being integrated into one product. Traditionally, the user specifies one of the datasets to be held as the reference to control the geometry (for matched features) that prevails in the conflated product. This decision is usually based on the users overall knowledge and experience with the data and is made at the database or layer level. However, implementing the proposed vector adjustment model could provide positional accuracy information at the feature level, *i.e.*, individual road centerlines. This information could then be used in a conflation solution where the matched features are compared and the one with the smallest positional displacement prevails as the reference and transferred to the conflated product.

This research project utilized aerial frame imagery that covered an area of less than 1 km<sup>2</sup> of area on the ground. Although, it was useful for demonstrating this proof-of-concept, it would be beneficial to extend this work to satellite imaging technology. For example, commercial satellites such as Quickbird are known to capture imagery having 10 km<sup>2</sup>, or more, of ground coverage. This would allow a much larger piece of a vector database to be tested. The positioning model for the imagery would be a bit different because satellites use scanning systems to acquire lines of imagery over time, which requires the use of sensor modeling to determine accurate image and ground coordinates. Nevertheless, the sensor model could be obtained from the satellite image provider or approximated using a replacement sensor model [47].

The proposed approach is somewhat manual at this point, requiring an individual to mensurate the imagery to identify image coordinates for the shape point locations. However, this process could be

automated by performing a vector-to-image registration between the vector road network and the corresponding roads in the imagery. Road intersections serve as an ideal candidate to establish correspondence between each dataset for two reasons: (1) the linear road vectors usually intersect at a node whose coordinate represents the center of the intersection, and (2) intersections usually perform well for Automated Feature Extraction (AFE) algorithms, which are needed to identify the intersection in the imagery. The vector adjustment could then be automated based on input files for the image coordinates, object point coordinates (vector database), GCP's, initial approximations for the EO parameters, their associated positional uncertainties, *etc.*

Finally, open source geographic data offers a unique opportunity to exploit the geolocation attributes associated with it. Future research should investigate how OSM roads could be used (as pseudo-GCP's) to determine/improve the location of a sensor (position and attitude) that collects imagery. This is essentially the reverse of what is being done here and could prove useful for geopositioning applications of non-traditional sensing systems, such as Unmanned Aerial Vehicle (UAV) platforms.

## References

1. *OpenStreetMap® Online Documentation Wiki*. Accessed online: <http://wiki.openstreetmap.org/wiki/About> (accessed on 7 August 2012).
2. Goodchild, M.F. Geographic information science. *Int. J. Geogr. Inf. Syst.* **1992**, *6*, 31–45.
3. *Ferris State University Surveying/Engineering Course Work*. Available online: <http://www.ferris.edu/HTMLS/colleges/cot/surveying/CheckSheets/SureCheckSheetBS.htm> (accessed on 1 December 2012).
4. Thompson, M. *Maps for America: Cartographic Products of the U.S. Geological Survey and Others*; US Dept. of Interior, Geological Survey: Washington, DC, USA, 1988.
5. American Congress on Surveying and Mapping (ACSM). The proposed standard for digital cartographic data. *Am. Cartogr.* **1988**, *15*, 9–140.
6. *ASPRS Accuracy Standards for Large-Scale Maps*; American Society of Photogrammetry and Remote Sensing (ASPRS): Bethesda, MD, USA, 1990; p. 1069.
7. *CloudMade Online Documentation*. Available online: <http://cloudmade.com/about> (accessed on 7 November 2012).
8. *Tech Crunch Online News Article*. Available online: <http://techcrunch.com/2012/05/04/apple-finally-gives-proper-credit-to-openstreetmap-in-iphoto-for-ios/> (accessed on 7 November 2012).
9. *Ushahidi Online Wiki Documentation*. Available online: <http://en.wikipedia.org/wiki/Ushahidi> (accessed on 7 November 2012).
10. *OpenStreetMap® Online Mapping Project Database*. Available online: <http://www.openstreetmap.org/> (accessed on 11 December 2011).
11. *US Geological Survey National Map Database*. Available online: <http://viewer.nationalmap.gov/viewer/> (accessed on 2 July 2012).
12. *US Census Bureau Topologically Integrated Geographic Encoding Referencing (TIGER) 2007 Shapefiles Database*. Available online: <http://www.census.gov/cgi-bin/geo/shapefiles/national-files> (accessed on 2 July 2012).
13. Goodchild, M.F.; Hunter, G. A simple positional accuracy measure for linear features. *Int. J. Geogr. Inf. Sci.* **1997**, *11*, 299–306.

14. Kiiveri, H. Assessing, representing, and transmitting positional uncertainty in maps. *Int. J. Geogr. Inf. Sci.* **1997**, *11*, 33–52.
15. US Dept. of Interior. *A Study of Land Information*; US Department of Interior: Washington, DC, USA, 1990; pp. 100–409.
16. Stanislawski, L.; Dewitt, B.; Shrestha, L. Estimating positional accuracy of data layers within a GIS through error propagation. *Photogramm. Eng. Remote Sensing* **1996**, *62*, 429–433.
17. Heuvelink, G.; Burrough, P. Developments in statistical approaches to spatial uncertainty and its propagation. *Int. J. Geogr. Inf. Sci.* **2002**, *16*, 111–113.
18. Doucette, P.; Motsko, D.; Sorenson, M.; White, D. Uncertainty Handling in Geospatial Data. In Proceedings of SPIE 2012 Defense, Security, and Sensing Symposium, Baltimore, MD, USA, 23–27 April 2012.
19. Lee, B. Spatial pattern of uncertainties: An accuracy assessment of the TIGER files. *J. Geogr. Geol.* **2009**, *1*, 2–12.
20. Zent, C.J. TIGER vs. DLG Road Feature Data. M.Sc. Thesis, The State University of New York at Buffalo, New York, NY, USA, 1996.
21. Haklay, M. How many volunteers does it take to map an area well? The validity of Linus' Law to volunteered geographic information. *Cartogr. J.* **2008**, *47*, 315–322.
22. Zielstra, D.; Zipf, A. A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany. In Proceedings of 13th Agile International Conference on Geographical Information Science 2010, Guimaraes, Portugal, 11–14 May 2010.
23. Cipeluch, B.; Jacobs, R.; Winstanley, A. Comparison of the Accuracy of OpenStreetMap for Ireland with Google Maps and Bing Maps. In Proceedings of the Spatial Accuracy Symposium, Leicester, UK, 20–23 July 2010.
24. O'Grady, K. *A DOQ Test Project: Collecting Data to Improve TIGER*. Available online: [http://www.census.gov/geo/mod/esri\\_paper.pdf](http://www.census.gov/geo/mod/esri_paper.pdf) (accessed on 11 February 2012).
25. Liadis, J. *GPS TIGER Accuracy Analysis Tools (GTAAT) Evaluation and Test Results*. Available online: <http://www.census.gov/geo/www/tiger/gtaat2000.pdf> (accessed on 11 February 2012).
26. Zandbergen, P.; Lenzer, K.; Ignizio, D. Positional accuracy of TIGER 2000 and 2009 road networks. *Trans. GIS* **2011**, *15*, 495–519.
27. Haklay, M. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environ. Plan. B* **2010**, *37*, 682–703.
28. GeoEye, Inc. *GeoEye Product Specifications*. Available online: <http://www.geoeye.com/CorpSite/products-and-services/imagery-collection/satellite-imagery-products/product-specifications.aspx> (accessed on 23 February 2012).
29. Talaya, J.; Kornus, W.; Alamus R.; Soler, E.; Pla, M.; Ruiz, A. Analyzing DMC Performance in a Production Environment. In Proceedings of the 37th ISPRS Congress, Beijing, China, 3–11 July 2008.
30. *Google Earth Geo Viewer (v.6.1)-Imagery Archive 2012*; Google Inc.: Mountain View, CA, USA, 2007.
31. Burrough, P.A. GIS and geostatistics: Essential partners for spatial analysis, *Environ. Ecol. Stat.* **2001**, *8*, 361–377.

32. Goodchild, M.F. Towards Models of Error for Categorical Maps. In Proceedings of First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographical Information and Analysis, Montecito, CA, USA, 13–16 December 1988.
33. Goodchild, M.F.; Gopal, S. *Modelling Error in Objects and Fields. Accuracy of Spatial Databases*; Taylor & Francis: New York, NY, USA, 1989; pp. 107–114.
34. Heuvelink, G.; Burrough, P.; Stein, A. Propagation of errors in spatial modelling with GIS. *Int. J. Geogr. Inf. Syst.* **1989**, *3*, 303–322.
35. Veregin, H. *Accuracy of Spatial Databases: Annotated Bibliography*; Technical Paper 89-9; NCGIA: Santa Barbara, CA, USA, 1989.
36. Veregin, H. *A Taxonomy of Error in Spatial Databases*; Technical Paper 89-12; NCGIA: Santa Barbara, CA, USA, 1989.
37. Goodchild, M.F. Keynote Address. In Proceedings of Symposium on Spatial Database Accuracy, Melbourne, Australia, 19–20 June 1991; pp. 1–6.
38. Shi, W.; Tempfli, K.; Ehlers, M. Analytical modelling of positional and thematic uncertainties in the integration of remote sensing and geographical information systems. *Trans. GIS* **1999**, *3*, 119–136.
39. Shi, W.; Liu, W. A stochastic process-based model for the positional error of line segments in GIS. *Geogr. Inf. Sci.* **2000**, *14*, 51–66.
40. Love, K.; Ye, K.; Smith, E.; Prisley, S. *Error Models in Geographic Information Systems Vector Data using Bayesian Methods*; Technical Report for the National Geospatial-Intelligence Agency (NGA); NGA-US Government: Springfield, VA, USA, 2007.
41. US Geological Survey. *National Elevation Dataset Online Database*. Available online: <http://ned.usgs.gov/> (accessed on 8 December 2011).
42. Mikhail, E.; Bethel, J.; McGlone, J.C. *Introduction to Modern Photogrammetry*; John Wiley & Sons: New York, NY, USA, 2001.
43. Wolf, P.; Dewitt, B. *Elements of Photogrammetry with Applications in GIS*, 3rd ed.; McGraw-Hill Publishing Company: Madison, WI, USA, 2000.
44. Andersen, J.; Mikhail, E. *Surveying Theory and Practice*, 7th ed.; McGraw Hill Publishers: New York, NY, USA, 1998; p. 31.
45. Ager, T.P. *An Analysis of Metric Accuracy Definitions and Methods of Computation*; InnoVision Whitepaper; National Imaging and Mapping Agency (NIMA): Springfield, VA, USA, 2002.
46. Canavosio-Zuzelski, R. A Photogrammetric Approach for Geopositioning OpenStreetMap Roads. Ph.D. Thesis, George Mason University, Fairfax, VA, USA, 2013.
47. Iiyama, M.; Dolloff, J.; Taylor, C. *The ABC's of RSM: An Introduction to the Replacement Sensor Model*; Technical Report; BAE Systems, Inc.: Aberdeen, SD, USA, 2006.