Computers & Geosciences 52 (2013) 85–94

Contents lists available at SciVerse ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Lattice-based clustering and genetic programming for coordinate transformation in GPS applications

Chih-Hung Wu [a,*], Wei-Han Su [b]

[a] Department of Electrical Engineering, National University of Kaohsiung, 700, Kaohsiung University Road, Nan-Tzu District, Kaohsiung 811, Taiwan
[b] POLSTAR Technologies Inc., Zhubei City, Hsinchu County, Taiwan

A R T I C L E   I N F O

Article history:
Received 9 July 2012
Received in revised form
22 September 2012
Accepted 24 September 2012
Available online 2 October 2012

Keywords:
Clustering
Genetic programming
Symbolic regression
GPS
Lattices
Coordinate systems

A B S T R A C T

Coordinate transformation is essential in many georeferencing applications. Level-wised transformation can be considered as a regression problem and done by machine-learning approaches. However, inaccurate and biased results are usually derived when training data do not uniformly distribute. In this paper, the performance of regression-based coordinate transformation for GPS applications is discussed. A lattice-based clustering method is developed and integrated with genetic programming for building better regression models of coordinate transformation. The GPS application area is first partitioned into lattices with lattice sizes being determined by the geographic locations and distribution of the GPS reference points. Clustering is then performed on lattices, not on data points. Each cluster of lattices serves as a training data set for a genetic regression model of coordinate transformation. In this manner, the data points contained in the different lattices can be considered to be of the same importance. Biased regression results caused by the imbalanced distribution of data can also be eliminated. The experimental results show that the proposed method can further improve the positioning accuracy than previous methods.

1. Introduction

Coordinate transformation is an important task in georeferencing applications. With the popularity of the Global Positioning System (GPS), GPS has become a convenient tool for georeferencing applications (El-Rabbany, 2002; Barbeau et al., 2010). GPS is a satellite-based navigation system and it describes a position in the Earth-Centered, Earth-Fixed (ECEF) Cartesian coordinate frame. Information for calculating the coordinates of a position (ϕ_{84}(latitude), λ_{84}(longitude), and h_{84}(height) in WGS84) is formatted in NMEA-0183 and transmitted by GPS satellites (NGS, 2007; DMA, 1987). In practical GPS applications, it is frequently required to convert the coordinates of a position from one geographic coordinate system to another. Typical conversion methods perform level-wised transformation by referring to various mathematical models. Such models usually involve complicated, nonlinear, algebraic formulas and parameters associated with a geodetically specific datum. A geodetic *datum* is a reference system officially established for a specific area and is used as national or continental standard for positioning. For example, TM2 is a 2-D geodetic datum based on GRS67 and has being widely used in many existing GPS applications used in the local area where the authors are from. In these applications, the coordinates need to be transformed from WGS84 to TWD97/TWD67 in TM2 to become useful (Tseng and Chang, 1999). Fig. 1 depicts the conversion process.

Consider the level-wised transformation from WGS84 to TWD97/TWD67 in the local applications. By counting the operators involved in the transformation formulas, it takes more than 165 and 186 floating-point operations to obtain a position in E_{67}^{TM2} (easting) and N_{67}^{TM2} (northing) from ϕ_{84}, λ_{84}, h_{84}, respectively (Doong, 2008). Recently, the demands on faster computation and low energy consumption when using handheld GPS devices receive more and more attentions. For example, the embedded systems transmitting real-time GPS data for monitoring slope stability, debris flow, bridges or crustal deformation (Mayer et al., 2010; He et al., 2011; Moschas and Stiros, 2011; He et al., 2004) are usually implemented on wireless, portable devices working with batteries or sonar cells. Due to the limited resources associated with handheld GPS devices, low power consumption is an important consideration in implementing such systems. The level-wise transformation has been used for years and many software tools have been developed based on these formulas. Recent studies attempt to reduce the costs of GPS-based coordinate transformation. For example, Soler et al. (2012) introduced a least squares solution to determine a unique set of geodetic coordinates, with accompanying accuracy predictions. The method is based on the given sets of individual (x,y,z) GPS-obtained values and their variance–covariance matrices. Shu and Li (2010) developed an

* Corresponding author. Tel.: +886 7 5919446; fax: +886 7 5919049.
E-mail address: johnw@nuk.edu.tw (C.-H. Wu).

0098-3004/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cageo.2012.09.022

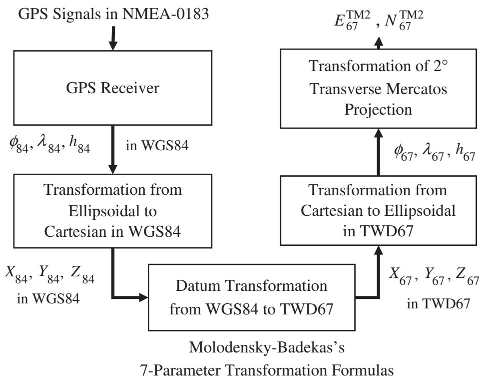

Fig. 1. GPS coordinate transformation: level-wised method.

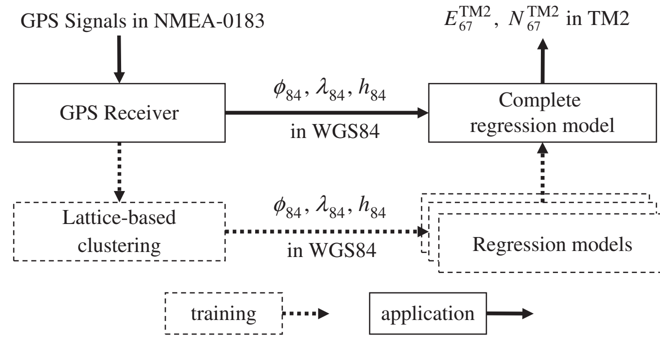

Fig. 2. The proposed method for GPS coordinate transformation.

iterative algorithm for the transformation from Cartesian to geodetic coordinates using the Newton–Raphson method to solve a quartic equation of the Lagrange parameter. Civicioglu (2012) presented the differential search algorithm to solve the problem of transforming the geocentric cartesian coordinates into geodetic coordinates. Recently, computational intelligence methods (Tierra et al., 2008; Wu et al., 2008b; Gullu, 2010) are also applied to estimate the transformation of coordinates.

As in most machine learning applications, a sufficient amount of high-quality training data is the key to success. The learning results may be trivial or meaningless if too many or too few training data are used. Coordinate transformation can be rephrased as a regression problem that builds simpler formulas of transformation from a training data set of GPS reference points (Wu et al., 2008a). However, the GPS reference points are not normally distributed in the application areas. Many reference points are established in the plain areas but few are in the mountain areas. For example, the application area under study is about 35,915 km^2 and contains a variety of land-forms, most are mountains. In the work of Wu et al. (2008a), 2700 reference points were used as training data, most of which are established on the west-side plain areas. The regression model is accurate and applicable in the areas where many (or enough) reference points are established. However, the regression model is not always accurate when being applied in the areas where few reference points are established.

This study intends to improve the performance of regression models of coordinate transformation so as to improve the positioning accuracy of GPS applications. Two learning techniques are integrated, clustering and genetic programming (GP) (Koza, 1992), for this purpose. To maximize the performance of both learning techniques, a lattice-based clustering method is developed. A lattice is a partition of the application area, containing a subset of GPS reference points extracted from the training data set. The lattice size is determined according to the geographic locations and distribution of the data points. Clustering is then performed on lattices, not data points. GP-based regression is then performed on the data contained in the lattices belonging to the same clusters. In this way, the data points contained in different lattices can be considered to have the same importance. Biased regression results caused by the distribution of data can be eliminated. The experimental results show that the proposed method can further improve the positioning accuracy than the previous methods. The basic idea of the proposed method is depicted in Fig. 2.

The remaining of this paper is organized as follows. Section 2 presents a brief review on related studies. In Section 3, the clustering problem and regression problem related to GPS coordinate transformation are defined. The techniques of lattice-based data clustering and the GP regression models are presented in Section 4. The results of experiments are given in Section 5. Finally, we conclude this study in Section 6.

2. Related work

In various machine learning applications, training on clustered or partitioned data usually can produce focused results. Below are several interesting studies that integrate clustering with machine learning methods. In the work of Ari and Güvenir (2002), clustered linear regression improved the accuracy of classical linear regression by partitioning training space into subspaces for linear approximations. Haeb-Umbach (2001) integrated bottom-up clustering with maximum likelihood linear regression for building speech models for speaker adaptation. Martinez-Estudillo et al. (2006) hybridized evolutionary algorithms and local search to solve nonlinear-regression problems. The product-unit-based neural network was employed for regression wherein associated parameters were determined by a hybrid evolutionary algorithm with local search and data clustering. Kung and Su (2007) developed a technique to establish affine Takagi–Sugeno fuzzy model for nonlinear systems. In their method, input–output data were first clustered by the fuzzy c-means clustering technique. A linear function was obtained from each cluster to adjust the parameters of each fuzzy rule. Angelov (2004) presented a similar idea for adaptation of fuzzy rule-based modeling using the techniques of on-line clustering. Clustering techniques are also applied to prevent regression engines from the pollution of noise data. For example, the study in Smyth et al. (2006) demonstrated the effectiveness of clustering noise data for tree-based regression. Multivariate regression trees with principal component scores and multivariate regression trees with factor scores were discussed. Liu et al. (2012) discussed the geometrical properties and attributes in spatial clustering and developed a new density-based spatial clustering algorithm that considers both spatial proximity and attribute similarity. Karakus (2011) employed the techniques of symbolic regression and genetic programming to analyze laboratory strength and elasticity modulus data for some granitic rocks. Similarly, a Bayesian model for clustered outliers in multiple regression was studied in Mohr (2007). Sato-Ilic (2009) evaluated the fuzzy regression models by considering the degree of belongingness and the degree of relativity in a cluster for a fixed object. In the work of Patil et al. (2012), a hybrid genetic algorithm which is tuned by support vector regression was developed to predict wave transmission of horizontally interlaced multilayer moored floating pipe breakwater. Tzima et al. (2012) suggested that clustering-based initialization can indeed improve the predictive accuracy as well as the interpretability of the induced knowledge of supervised LCS algorithms. From their experimental results, clustering results are satisfactory if training