STATISTICS-BASED FUSION OF TERRAIN DATA SETS AND CHANGE DETECTION

Jen-Jer JAW
Assistant Professor, Department of Civil Engineering
National Taiwan University
Taipei, 10617, Taiwan, ROC
Tel: (886)-2-23678645 Fax: (886)-2-23631558
E-mail: jejaw@ce.ntu.edu.tw

KEY WORDS: Fusion, Least Squares Collocation, Clustering, Prediction, Double Measurements, Error Propagation

ABSTRACT: This study aims at outlining a working experience where fusion and change detection of terrain data sets of dual-temporal but of the same area was performed via the technique of least squares adjustment. The proposed working flow first predicted the heights of one of the terrain data set onto the other one by means of least squares collocation, leading to point to point correspondence of the data sets on a horizontal component basis. The varied heights on the same horizontal positions in two data sets were considered the data type of double measurements of the very same terrain but with different variance-covariance matrices. A thresholding process followed by considering the randomness of the data sets, resulting in highlighting those points with change potential. The adjustment using weighting average then fused the heights that were reported unchanged. The areas of change were detected by the method of clustering. The outcomes in this study showed not only the higher accuracy of the fused terrain but also the areas of change in concern. Last but not least, the proposed workflow did not require any specific data format, thus avoiding degrading the original data quality due to unnecessary interpolation processes like other methods do.

1. INTRODUCTION

The handling of topographical data and the work of topographical change detection are frequent and important on account of economic development, engineering requirement and disaster evaluation, etc. It is a fairly practical subject to study how to utilize topographical data to perform the regional detection of topographical change for the government or civil class, providing helps to them for the engineering construction and the work of disaster reconstruction. Furthermore, for the purpose of gaining higher efficiency of data usage, it is quite natural to think of fusing terrain data sets once the areas of change have been detected. The realization of the above works, in this study, tempted to not only report the areas where the changes of topographic surface have taken place, but also render the higher accurate as well as reliable topographic data due to the process of fusing different data sets which actually represent the very same terrain.

To make use of the full feasibility of the terrain data sets, the proposed method does not rely on any specific data format. That is to catch directly the original data sources, irregular or regularly gridded, for further treatment, thus reducing the smoothing effect, as often seen, when pre-process performing interpolating data sets onto regular grid on the same horizontal component basis. Without the restriction of the data set format, the proposed method for change detection and fusion of terrain data sets is briefly outlined as follows: (1). The technique of least squares collocation was used to predict the heights of one of the terrain data set onto the other one, leading to point to point correspondence (or registration equivalently) of the data sets on a horizontal component basis. (2). A thresholding process followed by taking the randomness of the registered data sets into consideration resulting in highlighting those points with change potential. The areas of change then were detected by the method of clustering. (3). With the exclusion of surface in the change area, the varied heights on the same plane position in two data sets are considered the data type of double measurements of the very same terrain but with different variance-covariance matrices, in particular the predicted one showing non-diagonal structure due to collocation. The adjustment using weighting average then fused the heights and obtained a updated terrain data set that is more accurate as well as reliable than any one of the two original data sets.

The working flowchart can be seen as figure1. Note that the scenario in this study excluded the situation of having systematic error, such as datum shifting, between the two data sets.

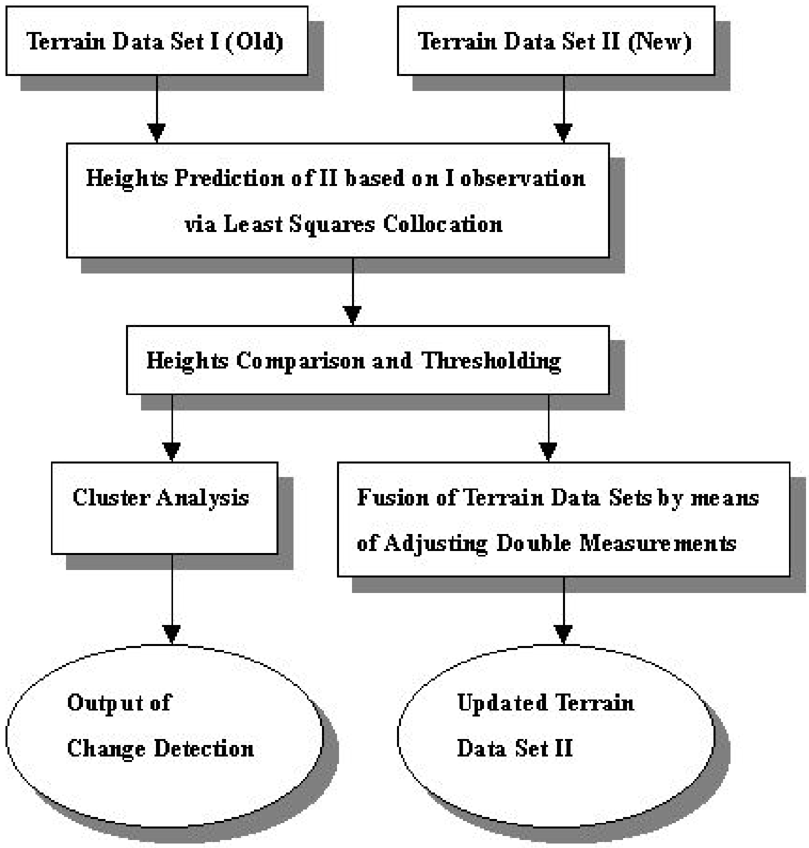


Figure 1 Flowchart of Change Detection and Fusion of Terrain Data Sets

In order to fulfill the requirement of making a compact paper, the methods involved in this study, including least squares collocation, cluster analysis and the way that the terrain data sets are fused, will only be briefly introduced in the next section. The interested readers could refer to the related literatures for more detail.

2. METHODS

2.1 Least Squares Collocation (Moritz, 1973)

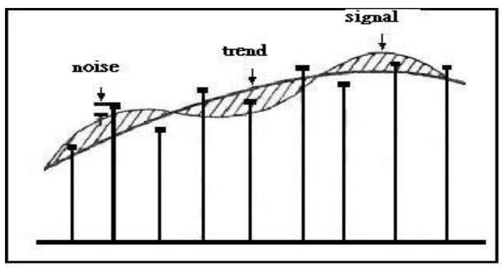
Applying least squares collocation technique in the task of surface interpolation, one would consider two major processes in need. One of which is to deal with the estimation of the unknown parameters and the other is to compute value of heights where the observations have not been truly taken, which is called "prediction" as well.

2.1.1 Estimation Of Unknown Parameters Of Terrain Surface

The observed heights, in this study, of terrain surface are considered as comprising three components: measuring noise, trend and signal, as shown in figure 2. Equation (1) expresses the relationship among these components.

$$\underset{m\times 1}{z} = \underset{m\times u}{A}\,\underset{u\times 1}{x} + s + r \qquad (1)$$

Where z: observed heights (m observations)
A: design matrix
x : unknown parameters (u unknowns)
s : signal
r : measuring noise



igure 2 Decomposition of Observation of Terrain

The first component (Ax, trend surface) in equation (1) represents the approximation, in a large-scale sense, of the terrain. It is determined by choosing an appropriate mathematical function, lower-order polynomials in general. The second and the third components, signal(s) and noise(r), respectively, actually constitute the core elements of the collocation engine.

It is the beauty of correlated components, namely signal, so that the spatial structure is formed accordingly and prediction follows. The spatial structure of the points is usually explained by a covariance function. It turns out that the determination of a proper covariance function plays a fatal role when evaluating the performance of collocation (Yang, 2001). Meanwhile, the measuring noise relating to the uncertainty of the observation is assumed purely random thus uncorrelated and with zero mean.

2.1.2 Prediction Of Unobserved Surface Points

Without further proof in this study, least squares collocation comes out with equation (2) for predicting unobserved surface points.

$$\hat{s}_u = \Sigma_{ls_u}{}^{T} \Sigma_{ll}{}^{-1} l \tag{2}$$

Where $\hat{s}_u$: Estimation of predicted points

Σ_{ls_u} : Covariance matrix constituted by predicted (unobserved) points and observed points

Σ_{ll} : Covariance matrix constituted by observed points

l : The detrended heights.

And the prediction error can be derived as in equation (3)

$$\Sigma_{\hat{s}_u} = \Sigma_{s_u} - \Sigma_{ls_u}{}^{T} \Sigma_{ll}{}^{-1} \Sigma_{ls_u} \tag{3}$$

Where Σ_{s_u} : The variance-covariance matrix of unobserved points

2.2 Thresholding For Change Detection

The operation of applying equation (2) in this study is to predict the heights of those points in data set II with known horizontal location taking observation of data set I . The predicted heights from data set I do not necessarily agree with the heights of data set II. The variation reveals two possibilities: the random nature of the measurement or mismatch of the data sets, partial or all, due to terrain change along the time. It follows that cutting a line for clarifying the reasonable justification of change detection appears somewhat subjective if there is no further assumption made. Thresholding with a lower bound tends to achieve a too exaggerated change report, which may be not the case. On the contrary, over higher the threshold may likely result in the loss that real surface change will not be detected. It is advisable to take both the randomness of the data sets and the error estimation through the prediction into consideration. To this end, the error propagation can be seen a useful tool for providing a suitable thresholding value with which change detection would be performed profoundly well. With the hypothesis that the terrain stays unchanged and adding the assumption of randomness of measurement, the threshold used to detect points with change potential is set by equation (4)

$$\text{Threshold} = 3 \cdot \sqrt{(\text{prediction error})^2 + s^2_{\text{data set II}}} \tag{4}$$

The quantity following the number "3" in the right-hand side of equation (4) accounts for the standard deviation of the height comparison of registered points between data set I and II. Three times of which already shows very little possibility of appearance under the assumption that height differences follow Gauss distribution. .

2.3 Cluster Analysis

Clustering is a discovery process that groups a set of data so that the intra-cluster similarity, homogeneity, is maximized while the inter-cluster similarity is minimized. The method used for measuring similarity includes (1). Distance Measures, (2). Association Measures and (3). Correlation Measures. The algorithm developed for clustering consists of hierarchical clustering method and non-hierarchical clustering method. For the purpose of aggregation, hierarchical clustering can be achieved either by way of linkage or via minimum variance method. On the other hand, K-Means method is often seen as far as the application of non-hierarchical clustering is concerned. The more thorough description of cluster analysis can be found in the literatures of (Fraley, etc, 1996)(Posse, 1996)(Hair, etc, 1998). For this study, the author followed the conclusion in (Yang, 2001) by employing single linkage method for achieving better outcome of clustering.

2.4 Fusion Of Terrain Data Sets Via Adjustment Computation

The other parallel process following the prediction is to fuse the two data sets. For the area staying unchanged status after change detection, the associated data sets can be interpreted as double measurements of surface points. Therefore, the least squares adjustment computation which integrates two data sets would bring about a updated data set with more reliable, in the sense of change detection, as well as accurate, in the sense of adjusting double measurements, terrain data set. The following equations mathematically demonstrate the process of fusing terrain data sets.

After the exclusion of change area, let vector l', size of m x 1, represent the predicted heights based on data set I, and the vector l'' stand for the associated heights in data set II. For outlining the adjustment of double measurements, equation (5) and (6) formulate the observation equations.

$$\underset{m\times1}{l'} = \underset{m\times m}{I} \cdot \underset{m\times1}{\bar{l}} + \underset{m\times1}{e'} \tag{5}$$

$$\underset{m\times1}{l''} = \underset{m\times m}{I} \cdot \underset{m\times1}{\bar{l}} + \underset{m\times1}{e''} \tag{6}$$

Where I: identity matrix

e', e'' : errors

$\bar{l}$: ture heights

Or equivalently

$$\underset{2m\times1}{y} = \underset{2m\times m}{\begin{bmatrix} l' \\ l'' \end{bmatrix}} = \underset{2m\times m}{\begin{bmatrix} I \\ I \end{bmatrix}} \cdot \underset{m\times1}{\bar{l}} + \underset{2m\times1}{\begin{bmatrix} e' \\ e'' \end{bmatrix}}, \qquad \text{With} \quad \underset{2m\times2m}{\Sigma_y} = \begin{bmatrix} \Sigma_{l'} & 0 \\ 0 & \Sigma_{l''} \end{bmatrix} = s_0^2 \cdot P^{-1} \tag{7}$$

where $\Sigma_{l'}$: variance and covariance matrix of l', usually a non-diagonal structure due to the collocation;

$\Sigma_{l''}$: variance and covariance matrix of l''; s_0^2 : variance component; P : weight matrix

The solution of least squares fusing the data sets can be obtained via equation (8)

$$\hat{l} = (A^T \cdot P \cdot A)^{-1} A^T \cdot P \cdot y \tag{8}$$

where $\hat{l}$: the estimation of $\bar{l}$; $A = [I \quad I]^T$

and the estimated dispersion matrix of $\hat{l}$ ($\Sigma_{\hat{l}}$) can be computed by

$$\Sigma_{\hat{l}} = \hat{s}_0^2 (A^T \cdot P \cdot A)^{-1} \tag{9}$$

With estimated variance component $\hat{s}_0^2 = \dfrac{\tilde{e}^T \cdot P \cdot \tilde{e}}{2m - m} = \dfrac{\tilde{e}^T \cdot P \cdot \tilde{e}}{m}$ (10)

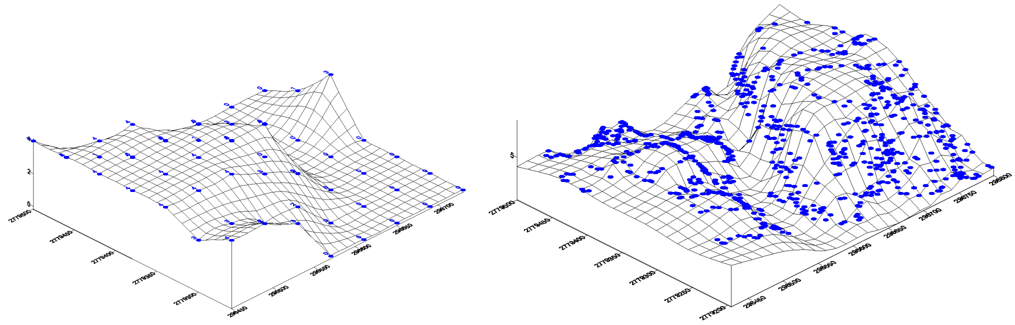
where $\tilde{e}$: residuals

3. EXPERIMENTS AND RESULT ANALYSIS

To illustrate the above idea, the author conducted a test in which the change detection and fusion of terrain data sets was tried by using two terrain data sets taken with about fifteen years time separation. The test site is located in the country of northern Taiwan, called "Guan-Du". The first data set (data set I), gridded DEM with 40 meters interval, was generated as of early 80s by Institute of Forestry, Aerial Survey, Bureau of Forest using photogrammetric method, while the second data set (data set II), points along contour-line extracted from one layer of 1/1000 topographic map in digital form, was prepared as of middle of 90s by Taipei municipality by ways of field surveying. Additionally, within the same data source as of data set II, the author also took spot heights from topographic map data layer used as checkpoints for experimental purpose. The aforementioned data sets can be summarized in table 1. The three-dimensional perception of the surface point distribution can be seen in figure 3a and 3b, as of data set I and II, respectively.

Table 1 The Description of Test Data

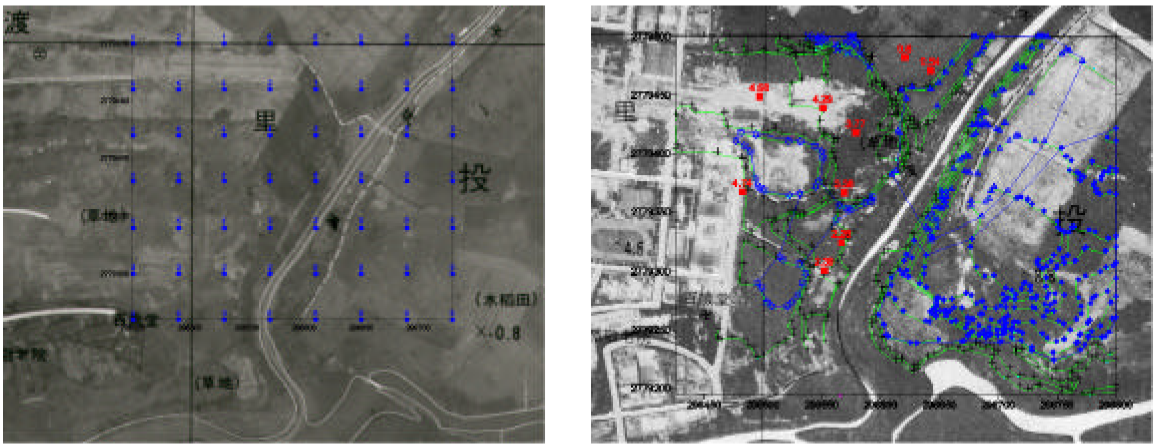
Data set	Year of generation	Data structure	Standard Deviation ($\pm m$)	No. of points
I	Early of 80s	Regular grid	0.66	56
II	Middle of 90s	Points along contour	0.27	807
Checkpoints	Middle of 90s	Spot heights		9



(a). Data Set I (b). Data Set II

Figure 3 Perspective View of Terrain (surface points are those marked)

The report of change detection and cluster analysis can be seen in figure 4b where four clusters segmented by polygons with blue lines represent four areas of change as the result of detection. To judge the performance of change detection, one could look at the base-map, as the background in figure 4a and 4b by realizing that the difference revealed from the radiometric information also reaches alike conclusion. In fact, some constructions have been done in this area during the last decades. A golf court showing up in upper-right part of figure 4b was not there when the first data set was collected. Besides, roads and man-made building in figure 4b are not seen in figure 4a. The constructions inevitably result in change of terrain.



(a) Data Set I with Base-map as Background

● : Gridded surface points

(b) Data Set II and the Indication of Change

:Contour-line

: Areas of change

+: Surface points along contours

◆, △, ✕, ✩ :Group 1,2,3,4 of clustering

■: Check points

Figure 4

The fusion of terrain data sets as specified in section 2.4 gave 0.265m as averaging standard deviation (theoretically) for updated data set II, rendered a slight increase of accuracy.

To further investigate the experimental performance of the proposed method, the author examined how the fusion process following least squares collocation would gain benefit by introducing 9 check points as specified in previous paragraph. One can find the distribution of which in figure 4b. After removing the surface points in the change area from the two data sets, the test went on predicting heights for the two data sets onto the checkpoints, the differences, treated as true errors, revealed the accuracy of prediction. Then, the fusion was performed on the two predicted data sets, the statistical indicators, mean or root mean square estimation (RMSE) of the heights difference deviated from checkpoints reported the experimental accuracy of fusion in overall. The result can be seen in table 2.

Table 2 Experimental Results for Prediction and Fusion

Data set	Prior error ($\pm m$)	Theoretical error ($\pm m$)	Experimental error ($\pm m$)	
			Mean	RMSE
I--→Prediction (Ip)	0.66	0.64	0.40	0.49
II--→Prediction(IIp)	0.27	0.31	0.26	0.30
Ip+IIp--→Fusion	0.31~0.64	0.24	0.19	0.26

From table 2, it is clearly shown that the fusion of predicted terrain data sets gives the better quality, in terms of both theoretical and experimental accuracy, than that of either individual data set would perform.

4. CONCLUSIONS

(1). The proposed method for change detection and fusion of terrain data sets comes without restricting data format which really differentiates itself from other methods or algorithms that are dependent of specified data formats.

(2). The spatial structure of surface told by covariance function from surface points paves the way for least squares collocation. The prediction of further unobserved points goes with locating the predicted signal while filtering out the noise, which is superior to any other methods that apply arbitrary interpolation methods without too much interpretation of spatial structure.

(3). The cluster analysis in change detection helps not only in the visual perception, such as fast and systematical localization of change area but also in the computation aspects, such as the change of the volume and the evaluation of damage.

(4). Fusion of the data sets proposed in this study demonstrates the benefit, both for robustness and accuracy of integrating multi-source of terrain data sets. Yet, without shortening the research scope, one may go by proposing other structures for way of fusion.

(5). The change detection and fusion of terrain data sets, proposed as a chain in this study was verified as a working method with theoretical as well as experimental justification.

ACKNOWLEDGEMENTS

The author would like to express the thankfulness for the support of Zhuo-Zhang Zong Educational Foundation on this study. A big favor from graduate student, Tsung-Jen Huang, for performing computation and preparation of nice figures is also gratefully acknowledged.

REFERENCES

Fraley, C., Raftery, A.E., 1996, How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis, Technical Report No. 329, Department of Statistics, University of Washington, USA.

Hair, J., Anderson, Jr., R., Tatham, R. and Black, W., 1998, Multivariate Data Analysis, Upper Saddle River, NJ, Prentice-Hall, 5th ed

Moritz, H., 1973, Least Squares Collocation. Theoretical Geodesy (German Geodetic Comm.), No.75, Munich.

Posse, C., 1996, Hierarchical Model-based Clustering For Large Datasets, Technical Report, Talaria Inc., Seattle, USA.

Yang, G.H., 2001, Topographic Change Detection, Thesis, Department of Civil Engineering, National Taiwan University, Taiwan.