Error propagation in the fusion of multi-source and multi-scale spatial information

Yee Leung
Department of Geography and Resource Management
The Chinese University of Hong Kong
Shatin, Hong Kong
yeeleung@cuhk.edu.hk

Jiang-Hong Ma
Department of Mathematics and Information Science
Chang'an University
Xi'an, P.R.China
jhma@chd.edu.cn

Tung Fung
Department of Geography and Resource Management
The Chinese University of Hong Kong
Shatin, Hong Kong
tungfung@cuhk.edu.

Abstract—**This paper proposes an optimization model for error propagation in the fusion of multi-source and multi-scale spatial information commonly encountered in the analysis of heterogeneous spatial data. The proposed method is mathematically simple and practical, and can remarkably improve the error variance of the fused data. Different cases of multi-source and multi-scale measurements are explored and the theoretical analysis of the method is established. The simulation study supports the theoretical arguments. The model paves the path for further analysis of uncertainty in the fusion of multi-source and multi-scale data.**

Keywords: information fusion, multi-source, multi-scale, error propagation

I. INTRODUCTION

With the development and advancement of technologies in the acquisition of spatial information, data measured at different scales may be obtained from different sources. An important and practical problem is how to turn multi-source and multi-scale data into more accurate and revealing information for problem solving. Multi-scale problems are usually associated with multi-source problems as data from different sources (e.g. sensors) are generally of different properties (scales or resolutions). It has been demonstrated that objects can be more accurately identified in images with higher spectral resolution resulting from the fusion of various sources of information obtained in different resolutions (Schistad Solberg et al., 1994). Therefore, multi-source and multi-scale (or resolution) data may coexist in a problem and need to be handled simultaneously and in accordance. Similar situation is also encountered in vector-based geographical information systems (GIS).

A successful methodology to solve these problems is to adopt information fusion techniques that have been widely applied to military study in the early 1970s. The approach has been extended to many related fields including artificial intelligence, automated control, robotics (Joshi and Sanderson, 1999), GIS, and remote sensing (RS). Through fusion (or integration), benefits of information coming from multiple sources and measured in different scales can be integrated, resulting in a new database with more information and higher quality.

There are a number of information fusion methods in the literature. Schistad Solberg et al. (1996) divides the methods of data fusion into four categories: statistical (Laferte et al.1995; Lee et al. 1987), fuzzy logic (Grégoire and Konieczny, 2006), Dempster-Shafer evidence theory (Rottensteiner et al., 2005), and neural network. Pohl and Van Genderen (1998) review methods for fusion of multi-sensor image data in remote sensing. In terms of image fusion techniques developed, they include the Intensity-Hue-Saturation (IHS) technique (Zhang and Hong, 2005), wavelet transform (Li et al., 2002; Ulfarsson et al., 2003), multi-scale Kalman filter (MKF) (Simone et al., 2000), multisensor Kalman (MSK) filter (Caron et al., 2006), and pyramid based algorithms (Sadjadi, 2005). However, little or no systematic attempt has been made to study the relative merits of various fusion techniques and their effectiveness in real multi-sensor imagery. While information fusion has been rather extensively studied, there is very little research on error analysis and propagation in the fusion of multi-source and multi-scale spatial data. In general, when the information fusion methods in the literature are implemented, it is essential to know how the accuracy (or reversely the error variances) of the final output is determined by the accuracies (or error variances) of the input data from various sources with different scales. In other words, an error propagation scheme is absolutely necessary for such investigation. Although effective integration of multi-source and multi-scale geo-referenced data has been found to reduce uncertainty, a formal analysis of uncertainty reduction on the theoretical and empirical basis is necessary to make it convincing and accountable.

The consistent treatment of uncertainty is fundamental to the correct fusion of different data sources. Without knowing the relative weightings that we need to give to each data source, we cannot know how to correctly combine them and how to determine the error contained in the fused output. In this paper, we propose a simple uncertainty model for the fusion of multi-source and multi-scale information which assigns each source of data a weight proportional to its variance. By constructing a very basic and reasonably simple mathematical model, we develop a method of error propagation for the general fusion of vector-based and rasterbased data, and attempt to characterize the statistical aspects of the method, including the impact of random distortions.

The basic uncertainty model and the associated optimization method for the fusion of multi-source and single scale spatial data is first proposed in Section II. The model is then extended to error analysis in the fusion of multi-source and multi-scale spatial data in Section III. A simulation experiment is implemented in Section IV to evaluate the model and the optimization method. The paper is then concluded with a summary and outlook in Section V.

II. Error Propagation for the Fusion of Multi-Source and Single-Scale Spatial Data

A. The Case in Which all Variances are Known

Assume that there are m independent sources, denoted by $\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_m$, of measurement data. The measurement X_i from $\mathbf{S}_i$ for the same true value x (at a location in a region or a pixel in an image) is unbiased with an additive random error (that is, the expectation $EX_i = x$) and its (error) variance is $\mathrm{Var}(X_i) = \sigma_i^2$, $i = 1, 2, \ldots, m$. The measurements may be coordinates, pixel grey levels, or other form of measurement. Based on the idea of kriging in geostatistics, we can construct through linear combination a new and more precise measurement estimator:

$$X = \lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_m X_m = \sum_{i=1}^{m} \lambda_i X_i, \qquad (1)$$

where λ_i can be viewed as the weight of the measurement X_i, and X is a weighted average satisfying $\sum_{i=1}^{m} \lambda_i = 1,\ \lambda_i \geq 0,\ i = 1, 2, \ldots, m$. Under such transformation, X is still unbiased, i.e.,

$$E(X) = \lambda_1 E(X_1) + \lambda_2 E(X_2) + \cdots + \lambda_m E(X_m) = x. \qquad (2)$$

And we have

$$\mathrm{Var}(X) = \lambda_1^2 \mathrm{Var}(X_1) + \lambda_2^2 \mathrm{Var}(X_2) + \cdots + \lambda_m^2 \mathrm{Var}(X_m) = \sum_{i=1}^{m} \lambda_i^2 \sigma_i^2 \qquad (3)$$

In order to choose suitable weights λ_i $(i = 1, 2, \ldots, m)$ so that (3) is minimized, we can solve the following optimization problem:

$$\begin{cases} \text{Min: } f(\lambda_1, \lambda_2, \ldots, \lambda_m) \equiv \sum_{i=1}^{m} \lambda_i^2 \sigma_i^2 \\ \text{s.t.: } \sum_{i=1}^{m} \lambda_i = 1,\ \lambda_i \geq 0,\ i = 1, 2, \ldots, m. \end{cases} \qquad (4)$$

It can easily be derived that the solution is

$$\tilde{\lambda}_i = \left(\sum_{k=1}^{m} \frac{1}{\sigma_k^2} \right)^{-1} \frac{1}{\sigma_i^2},\ i = 1, 2, \ldots, m, \qquad (5)$$

where $(\sigma_i^2)^{-1}$ can be interpreted as a precision measure of the measurement from the *i*th data source. In other words, the larger is the variance, the smaller the weight. Thus in (5) the weight $\tilde{\lambda}_i$ of a measurement is proportional to the corresponding precision of the measurement. This result is natural and intuitive.

From (5) and (3) we can obtain the fused estimator

$$\tilde{X} = \sum_{i=1}^{m} \tilde{\lambda}_i X_i, \qquad (6)$$

and its variance

$$\mathrm{Var}(\tilde{X}) = \left(\sum_{k=1}^{m} \frac{1}{\sigma_k^2} \right)^{-1}, \qquad (7)$$

which is smaller than any σ_i^2 $(i = 1, 2, \ldots, m)$. In fact, for arbitrary *i*,

$$\mathrm{Var}(\tilde{X}) = \left(\sum_{k=1}^{m} \frac{1}{\sigma_k^2} \right)^{-1} < \left(\frac{1}{\sigma_i^2} \right)^{-1} = \sigma_i^2,\ \text{i=1, 2,..., m} \qquad (8)$$

B. The Case in Which all Variances are Unknown

If all variances are unknown, the "plug-in" method can be employed to obtain the corresponding estimator. That is, all variances σ_i^2 $(i = 1, 2, \ldots, m)$ in (5) are replaced by their sample estimates $\hat{\sigma}_i^2$. Thus, we can get a "plug-in" fused estimator

$$\hat{X} = \sum_{i=1}^{m} \hat{\lambda}_i X_i, \qquad (9)$$

where

$$\hat{\lambda}_i = \left(\sum_{k=1}^{m} \frac{1}{\hat{\sigma}_k^2} \right)^{-1} \frac{1}{\hat{\sigma}_i^2},\ i = 1, 2, \ldots, m. \qquad (10)$$

Nevertheless, the exact variance of $\hat{X}$ is complicated. An approximate expression can however be given by

$$\mathrm{Var}(\hat{X}) \approx \hat{\sigma}_*^2 \equiv \left(\sum_{k=1}^{m} \frac{1}{\hat{\sigma}_k^2} \right)^{-1}. \qquad (11)$$

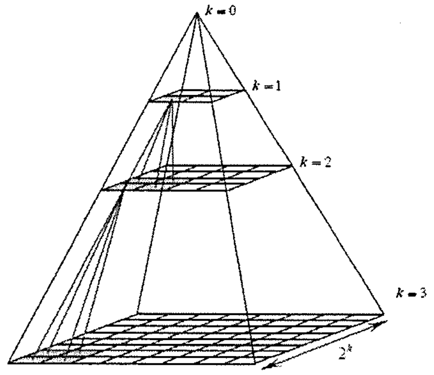
III. Error Propagation for the Fusion of Multi-Source and Multi-Scale Spatial Data

A. The Case in Which all Variances are Known

Assume that there are m independent data sources, denoted by $\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_m$, and the measurements X_{ij}

($j = 1, 2, \ldots, n_i$) from $\mathbf{S}_i$ for the true value x_{ij} in the quadtree are independently random with the expectation $EX_{ij} = x_{ij}$ and common variance $\mathrm{Var}(X_{ij}) = \sigma_i^2$, where n_i is the number of measurements from $\mathbf{S}_i$, $i = 1, 2, \ldots, m$. In addition, the measurements from different data sources have the corresponding scales $s_1, s_2, \ldots, s_m$ ($s_1 < s_2 < \ldots < s_m$) respectively. They form a pyramid structure as a quadtree (*Fig. 1*) in which the root node (the first level) corresponds to the smallest scale measurements and the leaf nodes (the m-th level) corresponds to the largest scale measurements.

Small Scale



Large Scale

Figure 1. Quadtree Pyramid Data Structure (adopted from Slatton et al., 2001)

Without loss of generality, we adopt quadtree as a basis for further analysis although the proposed method is not limited to quadtree. We assume that the measurements in the i-th level are X_{ij} (j=1, 2, ..., n). Then the i-th level measurement in the center location, which is assumed to have the form $X_i \equiv \sum_{j=1}^{n_i} \eta_{ij} X_{ij}$, can be computed by the weighted average of measurements of its neighboring units

$$X_i \equiv \sum_{j=1}^{n_i} \eta_{ij} X_{ij}, \tag{12}$$

where the weights η_{ij} are known and can usually be taken as identical values, i.e., $1/n_i$, $j=1,2,\ldots,n_i$. If so, it means that X_{ij} in (12) is the center of gravity of the i-th level measurements. In this case, we may still consider the fused estimator similar to that in (12). By the same token, we can obtain the optimal weights as follows

$$\tilde{\lambda}_i = \left(\sum_{k=1}^{m} \frac{1}{\tilde{\sigma}_k^2} \right)^{-1} \frac{1}{\tilde{\sigma}_i^2} \quad i = 1, 2, \ldots, m \tag{13}$$

where

$$\tilde{\sigma}_i^2 = \mathrm{var}(X_i) = \sigma_i^2 \sum_{j=1}^{n_i} \eta_{ij}^2 \tag{14}$$

Accordingly, the corresponding fused estimator and its variance can be derived as that in (6) and

$$\mathrm{Var}(\tilde{X}) = \left(\sum_{k=1}^{m} \frac{1}{\tilde{\sigma}_k^2} \right)^{-1}. \tag{15}$$

B. The Case in Which all Variances are Unknown

If all variances are unknown, the "plug-in" method can be employed to obtain the corresponding estimator, that is, all variances σ_i^2 ($i = 1, 2, \ldots, m$) in (5) are replaced by their sample estimates $\hat{\sigma}_i^2$. Thus, we can get a "plug-in" fused estimator (9), where

$$\hat{\lambda}_i = \left(\sum_{k=1}^{m} \frac{1}{\hat{\tilde{\sigma}}_k^2} \right)^{-1} \frac{1}{\hat{\tilde{\sigma}}_i^2}, i = 1, 2, \ldots, m, \tag{16}$$

where

$$\hat{\tilde{\sigma}}_i^2 \equiv \hat{\sigma}_i^2 \sum_{j=1}^{n_i} \eta_{ij}^2. \tag{17}$$

However, the exact variance of $\hat{X}$ is complicated. An approximate expression can be given by

$$\mathrm{Var}(\hat{X}) \approx \hat{\sigma}_*^2 \equiv \left(\sum_{k=1}^{m} \frac{1}{\hat{\tilde{\sigma}}_k^2} \right)^{-1}. \tag{18}$$

IV. SIMULATION

To show the effectiveness of the proposed method, a simulation experiment is carried out. The example gives a simple result of optimal fusion of measurements from two sources with the same scale.

Example Consider the case that $m = 2$, $\sigma_1^2 = 2$, $\sigma_2^2 = 6$. According to (4), we have $\tilde{\lambda}_1 = \frac{3}{4}$ and $\tilde{\lambda}_2 = \frac{1}{4}$, thus $\mathrm{Var}(\tilde{X}) = \mathrm{Var}(\frac{3}{4}X_1 + \frac{1}{4}X_2) = \frac{3}{2}$. In addition, $\sigma_{\min}^2 = 2$, $\sigma_{\max}^2 = 6$. Thus,

$$\tilde{\sigma}^2 \equiv \mathrm{Var}(\tilde{X}) = \tfrac{3}{2} < 2 = \sigma_{\min}^2,$$

$$\frac{\sigma_{\min}^2}{m} = 1 < \mathrm{Var}(\tilde{X}) = \tfrac{3}{2} < 3 = \frac{\sigma_{\max}^2}{m}.$$

The theoretically optimal uncertainty propagation scheme is depicted in *Fig. 2* and the simulation results are shown in Table 1 for different sizes of sample.

TABLE 1 SIMULATION RESULTS FOR DIFFERENT ESTIMATORS (M=2)

Sample size	100		500	
Source	S1	S2	S1	S2
x_center	0	0	0	0
True var.	2	6	2	6
	known var.	unknown var.	known var.	unknown var.
Sample var.	1.222365	1.224085	1.349582	1.346844
Approx.var.		1.361647		1.404552

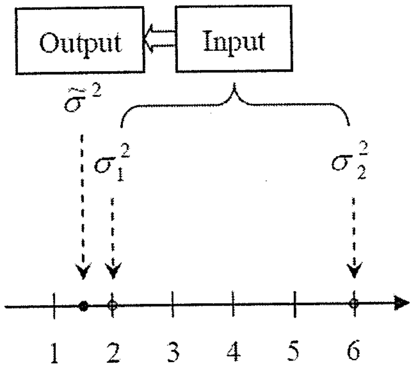


Figure 2. Uncertainty propagation

V. CONCLUSIONS

We have proposed in this paper an approach to error analysis in multi-source and multi-scale spatial data. The novelty of the approach is the provision of a simple and practical optimization scheme for error propagation in the fusion of multi-source and multi-scale measurement information. It has been demonstrated that the choice of an optimal error-propagation scheme is of paramount importance in handling errors in the fusion of spatial information coming from various sources and measured in different scales. Without such scheme, it is essentially impossible to analyze and track errors throughout the fusion process. Consequently, we will not be able to nail down the error contained in the final product on which decisions are made. The proposed method can remarkably improve the error variance of the information after fusion, a desirable result in data fusion. The simulation experiments have shown that the method can provide an optimal estimate for the fusion of multi-source and multi-scale measurements. The study paves the path for in depth analysis of error propagation under various schemes of information fusion. As a direction for further study, the method can be applied to the fusion of real-life data sets in GIS and/or remote sensing images. Another direction is to develop the error propagation schemes for major existing fusion algorithms, e.g., pyramid based algorithms, so that error analysis can be incorporated into such frameworks for the fusion of spatial data. This line of research is, however, more difficult and complex. It first requires the determination of uncertainty in each measurement source, the input-output propagation scheme of each fusion algorithm then needs to be investigated and characterized mathematically. Such research is nevertheless very valuable for the fusion of multi-source and multi-scale spatial information.

ACKNOWLEDGEMENTS

This project was supported by the earmarked grant CUHK 446907 of the Hong Kong Research Grants Council.

REFERENCES

Caron, F., E. Duflos, D. Pomorski, and P. Vanheeghe. (2006). GPS/IMU data fusion using multisensor Kalman filtering: introduction of contextual aspects. *Information Fusion*. 7 (2), 221-230.

Grégoire, E. and S. Konieczny. (2006). Logic-based approaches to information fusion. *Information Fusion*. 7 (1), 4-18.

Joshi, R. and A.C. Sanderson. (1999). *Multisensor fusion: A minimal representation framework*. Singapore: World Scientific Publishing.

Laferte, J.-M., Heitz, F., Perez, P. and Fabre, E. (1995). Hierarchical statistical models for the fusion of multiresolution image data, *Proceedings of Fifth International Conference on Computer Vision*, 20-23 June 1995 pp. 908 – 913

Lee, T., J. A. Richards, and P. H. Swain. (1987). Probabilistic and evidential approaches for multisource data analysis. *IEEE Transactions on Geoscience and Remote Sensing*. 25 (3), 283-293.

Li, S., J.T. Kwok, and Y. Wang. (2002). Using the discrete wavelet frame transform to merge Landsat TM and SPOT panchromatic images, *Information Fusion*. 3, 17–23.

Pohl, C., and J.L. van Genderen. (1998). Multisensor image fusion in remote sensing: concepts, methods, and applications, International Journal of Remote Sensing. 19, 823–854.

Rottensteiner, F., J. Trinder, S. Clode, and K. Kubik. (2005). Using the Dempster–Shafer method for the fusion of LIDAR data and multi-spectral images for building detection. *Information Fusion*. 6 (4), 283-300.

Sadjadi, F. (2005). Comparative Image Fusion Analysais, 2nd Joint IEEE International Workshop on Object Tracking and Classification in and Beyond the Visible Spectrum (OTCBVS) Program. San Diego, CA, USA, June 20, 2005 (http://www.cse.ohio-state.edu/OTCBVS/05/OTCBVS-05-FINALPAPERS/W01_13.pdf)

Schistad Solberg, A. H., A. K. Jain, and T. Taxt (1994). Multisource classification of remotely sensed data: fusion of Landsat TM and SARimages. *IEEE Transactions on Geoscience and Remote Sensing*. 32 (4), 768-778.

Schistad Solberg, A.H., T. Taxt, and A.K. Jain (1996). A Markov random field model for classification of multisource satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*. 34, 100–112.

Simone, G. Morabito, F.C. Farina, A. (2000). Radar image fusion by multiscale Kalman filtering. *FUSION 2000. Proceedings of the Third International Conference on Information Fusion*. Date: 10-13 July 2000, Vol. 2, pp: WED3/10 - WED3/17

Slatton, K.C., Crawford, M.M., and Evans, B.L. (2001). Fusing interferometric radar and laser altimeter data to estimate surface topography and vegetation heights. *IEEE Transactions on Geoscience and Remote Sensing*. 39, 2470-2482

Ulfarsson, M.O., J.A. Benediktsson, and J.R. Sveinsson (2003). Data fusion and feature extraction in the wavelet domain. *International Journal of Remote Sensing*. 24, 3933–3945.

Zhang, Y. and G. Hong. (2005). An IHS and wavelet integrated approach to improve pan-sharpening visual quality of natural colour IKONOS and QuickBird images. *Information Fusion*. 6 (3), 225-234.