OVERALL UNCERTAINTY OF GEOREFERENCING AND CLASSIFICATION

Lubomír Soukup

Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod vodárenskou věží 4, 182 08, Praha 8, Czech Republic
soukup@utia.cas.cz, http://www.utia.cas.cz

Research Institute of Geodesy, Topography and Cartography
250 66, Zdiby, Czech Republic, http://www.vugtk.cz

KEY WORDS: Bayesian classification, Georeferencing, Geometric transformation, Bayesian estimation, Fuzzy region

ABSTRACT:

Positional accuracy of geographic objects is treated in this contribution. Two main sources of uncertainty that influences the positional accuracy are considered:

- uncertainty of classification,
- uncertainty of georeferencing (image registration).

These two sources are embraced in the overall spatial precision of geographic objects. It is assumed that several geographic objects are given as a result of Bayesian classification of a given digital (raster) image. It means that these objects are not determined precisely as crisp objects. Georeferencing of the given image is supposed to be determined by a linear transformation, coefficients of which were estimated by means of Bayesian approach. This transformation is therefore imprecise as well. It is designed here how to properly combine these two partial imprecise results to get the overall spatial precision of each geographic object in the given image. The resulting individual geographic object is expressed by its boundary which is approximated by a polygon with imprecise vertices. Precision of the vertices is computed from the initial data by means of Bayesian estimation so that any border point between adjacent objects is unique though imprecise. Special probability distribution (close to Gaussian) was designed to model the positional accuracy of the vertices.

1 INTRODUCTION

There are two crucial problems in recent geoinformation technology:

1. representation and processing of uncertain spatial data,
2. recognition of objects in raster images and description of their shapes in vector format.

Each of them is highly complex task from the theoretical and computational point of view. There are plenty approaches and methods that tries to solve each of the both problems separately on various levels of proximity, efficiency, and applicability. Despite the multifarious character of the two problems, they should be solved in a unified, comprehensive way, since they usually occur simultaneously. The reasons are apparent.

1. Any observation of the earth is uncertain.
2. Raster data can be easily captured, but vector data are more suitable for spatial analysis.

The main task of this contribution is to express uncertain boundary of a geographic object in form of closed polygon with imprecise vertices. To correctly estimate the imprecision of the vertices probability distribution of them is inferred by means of Bayesian approach (see e.g. (Koch, 1990)).

It is assumed that the geographic object was recognized in a given digital image by Bayesian classification and then transformed to a cartographic coordinate system. Coefficients of the transformation are assumed to be estimated by means of Bayesian inference as well. Bayesian approach serves, actually, as a glue of the two different tasks that are partial problems of the main task mentioned above.

Another important approach that is fundamental for this contribution is based on combination of fuzzy set theory an probability theory. Fuzzy set theory has been successfully applied to model uncertain regions and their mutual relation during the last decade. Review of these attempts is presented in (Schmitz and Morris, 2006). Unfortunately, there is little effort to combine fuzzy theory with probability theory in the field of spatial data quality and experimental data processing. Some challenging hints can be found in book (Viertl, 1996).

2 FORMULATION OF THE PROBLEM

2.1 Given:

- result of Bayesian classification of a digital image, i.e. for each class $U \in \mathcal{U}$ probability p_U as a discrete function of pixels of the image is defined.

$$p_U : \mathcal{X} \to \langle 0,\, 1\rangle : [m, n] \mapsto p_U(m, n)\ ,$$

where
$\mathcal{X}$ … index set of pixels of the given image,
m, n … coordinates of a specified pixel

- random field (Markov or Gaussian) that describes mutual dependence of position-ally close pixels.

- transformation $\mathbf{T}_{\boldsymbol{\theta}}$ that represents georeferencing of the image.

$$\mathbf{T}_{\boldsymbol{\theta}} : \mathcal{X} \to \widetilde{\mathcal{X}} : \mathbf{x} \mapsto \mathbf{T}_{\boldsymbol{\theta}}(\mathbf{x}) \ ,$$

  shortly expressed:

$$\widetilde{\mathbf{x}} = \mathbf{T}_{\boldsymbol{\theta}}(\mathbf{x}) \tag{1}$$

  $\widetilde{\mathcal{X}}$ ... small part of Earth surface, $\widetilde{\mathcal{X}} \subset \mathbb{E}^2$
  $\mathbb{E}^2$ ... Euclidean plane,
  $\mathbf{x}$ ... position of a pixel in the image, $\mathbf{x} = [m, n]$
  $\widetilde{\mathbf{x}}$ ... cartographic coordinates of a point on earth,
  $\boldsymbol{\theta}$ ... transformation coefficients, $\boldsymbol{\theta} \in \mathcal{T} \subset \mathbb{R}^m$, $m \in \mathbb{N}$

- joint probability distribution of transformation coefficients $\boldsymbol{\theta}$. Probability density function of $\boldsymbol{\theta}$ is denoted by $\varphi(\boldsymbol{\theta})$.

2.2 Required:

- closed polygon that approximates border of the given region;

- joint probability distribution of the all vertices of the polygon.

3 SOLUTION

3.1 Approximation of the region boundary by a polygon

After Bayesian classification, each pixel in the image is assigned by a probability that the pixel belongs to a specified class, say U. These probabilities are given as values of the given function p_U. To describe the main idea clearly, let confine us to only one region on some background. The region of class U that was recognized by the Bayesian classification can be treated as fuzzy region whose membership function is p_U. Uncertain boundary of this fuzzy region will be approximated by a closed polygon with imprecise vertices. The main task of this section is to estimate probability density function of the vertices. Let us denote the density function by g_U.

$$g_U : \mathcal{X}^n \to \langle 0,\, 1\rangle : [\mathbf{x}_1\,,\, \ldots\,, \mathbf{x}_n] \mapsto g_U(\mathbf{x}_1\,,\, \ldots\,, \mathbf{x}_n) \ .$$

For concise notation, let us abbreviate the vertices by $\widehat{\mathbf{x}}$, i.e.:

$$[\mathbf{x}_1\,,\, \ldots\,, \mathbf{x}_n] \stackrel{\triangle}{=} \widehat{\mathbf{x}} \,.$$

Let us imagine another probability density function of vertices $\widehat{\mathbf{x}}$ that is parameterized by vector $\mathbf{q}$, say $f_{U,\mathbf{q}}$ It is possible, with aid of function $f_{U,\mathbf{q}}$, to construct continuous membership function of any closed polygon. Let us denote the function by $\mu_{U,\mathbf{q}}$. Value of this membership function for an arbitrary point $\mathbf{x}$ stands for probability that this point belongs to region r_U incident to class U.

$$\mu_{U,\mathbf{q}}(\mathbf{x}) = P(\mathbf{x} \in r_U) \ .$$

This approach that combines fuzzy set theory an probability theory represents the fundamental idea of this contribution.

When the membership function $\mu_{U,\mathbf{q}}$ becomes successfully constructed, it will be compared to given function p_U. As a result of the comparison, probability distribution of parameters $\mathbf{q}$ is inferred by Bayesian approach. Let us denote probability density function of $\mathbf{q}$ by h_U. Function h_U allows us to easily express function g_U that has been searched for.

$$g_U(\widehat{\mathbf{x}}) = \int_{\mathcal{Q}} f_{U,\mathbf{q}}(\widehat{\mathbf{x}})\, h_U(\mathbf{q})\, d\mathbf{q} \,. \tag{2}$$

Set $\mathcal{Q}$ stands for admissible values of parameter vector $\mathbf{q}$.

Now the procedure of determination g_U just described will be explained in more detail.

3.1.1 Setting up auxiliary probability distribution of vertices

The auxiliary probability distribution is the one with pdf (probability density function) $f_{U,\mathbf{q}}$. It is joint probability density of vertices $\widehat{\mathbf{x}}$. For simplicity reasons, let us require the vertices to be mutually statistically independent.

$$f_{U,\mathbf{q}}(\mathbf{x}_1\,,\, \ldots\,, \mathbf{x}_n) = \prod_{i=1}^{n} f_i(\mathbf{x}_1) \tag{3}$$

f_i is probability density function of individual vertex. It is supposed that for each $i \in \{1, \ldots, m\}$ f_i is normal (Gaussian) density with parameters $[\mu_i\,,\, \sigma_i] \stackrel{\triangle}{=} \mathbf{q}_i$, $\mathbf{q} \stackrel{\triangle}{=} [\,\mathbf{q}_1\,,\, \ldots\,,\, \mathbf{q}_n\,]$.

3.1.2 Construction of polygonal fuzzy region

This step is the crucial one of the whole procedure. It is so complicated that its entire mathematical formulation exceeds the scope of this contribution. Nevertheless, the main idea is quite simple.

For each point $\mathbf{x} \in \mathcal{X}$ we integrate function $f_{U,\mathbf{q}}$ over all possible polygons of length n that cover point $\mathbf{x}$. Probability density of each polygon is given by probability density of its vertices, i.e. by function $f_{U,\mathbf{q}}$. Since the number of vertices n is unknown, we start with rough approximation by triangle and then increase number n until average positional precision of vertices stabilizes. By this iterative process, membership function $\mu_{U,\mathbf{q}}$ and number n are determined.

3.1.3 Bayesian estimation of auxiliary parameters qb

Estimation of parameters qb is possible through comparison of functions $\mu_{U,\mathbf{q}}$ and p_U. Values of function p_U are in fact random since they originate from frequencies counted in training set. Statistical behaviour of these values is supposed to be known and defined by the given random field. Therefore probability distribution of matrix $p_U(\mathcal{X})$ is defined as well. Let us denote its pdf by ψ. Then matrix $p_U(\mathcal{X})$ can be statistically compared with matrix $\mu_{U,\mathbf{q}}(\mathcal{X})$ and thanks to Bayes theorem pdf of vector $\mathbf{q}$ can be inferred.

$$h_U(\mathbf{q}) = \frac{\psi(p_U(\mathcal{X}) - \mu_{U,\mathbf{q}}(\mathcal{X}))\, p(\mathbf{q})}{\displaystyle\int_{\mathcal{Q}} \psi(p_U(\mathcal{X}) - \mu_{U,\mathbf{u}}(\mathcal{X}))\, p(\mathbf{u})} \tag{4}$$

Function p stands for a prior pdf that has to be suitably designed with respect to parameters contained in vector $\mathbf{q}$. Result of this step is pdf h_U.

3.1.4 Determination of probability distribution of vertices

Result of this final step is pdf g_U that was mentioned in advance by equation (2).

3.2 Transformation to cartographic coordinate system

Transformation equations (1) can be rewritten to

$$[\widetilde{\mathbf{x}}_1\,,\, \ldots\,, \widetilde{\mathbf{x}}_n] = \mathbf{T}_n(\mathbf{x}_1\,,\, \ldots\,, \mathbf{x}_n\,,\, \boldsymbol{\theta}) \ , \tag{5}$$

where

$$\mathbf{T}_n : \mathcal{X}^n \times \mathcal{T} \to \widetilde{\mathcal{X}}^n : [\mathbf{x}_1\,,\,\ldots\,,\mathbf{x}_n\,,\,\boldsymbol{\theta}] \mapsto \mathbf{T}_n(\mathbf{x}_1\,,\,\ldots\,,\mathbf{x}_n\,,\,\boldsymbol{\theta})$$

$$\mathbf{T}_n(\mathbf{x}_1\,,\,\ldots\,,\mathbf{x}_n\,,\,\boldsymbol{\theta}) = [\mathbf{T}_{\boldsymbol{\theta}}(\mathbf{x}_1)\,,\,\ldots\,,\mathbf{T}_{\boldsymbol{\theta}}(\mathbf{x}_n)]\,, \quad \mathcal{T} \subset \mathbb{R}^m$$

Pdf of coordinates $\widehat{\mathbf{x}}$ can be easily expressed as transformation of random vectors $\widehat{\mathbf{x}}$, $\boldsymbol{\theta}$ with the aid of fundamental theorems of probability theory (see e.g. (Trivedi, 1982)). The resulting pdf $g_{\mathbf{T}_n}$ is as follows.

$$g_{\mathbf{T}_n}(\widehat{\mathbf{x}}) = \int_{\mathcal{T}} \overline{g}(\overline{\mathbf{T}}_n^{-1}(\widehat{\mathbf{x}}\,,\,\boldsymbol{\theta}))\, \frac{1}{\left|\det \dfrac{\partial \overline{\mathbf{T}}_n}{\partial \mathbf{z}}(\overline{\mathbf{T}}_n^{-1}(\widehat{\mathbf{x}}\,,\,\boldsymbol{\theta}))\right|}\, d\boldsymbol{\theta}\;, \tag{6}$$

where

$$\overline{\mathbf{T}}_n : \mathcal{X}^n \times \mathcal{T} \to \widetilde{\mathcal{X}}^n \times \mathcal{T} : [\widehat{\mathbf{y}}\,,\,\boldsymbol{\vartheta}] \mapsto \overline{\mathbf{T}}_n(\widehat{\mathbf{y}}\,,\,\boldsymbol{\vartheta}) = [\mathbf{T}(\widehat{\mathbf{y}}\,,\,\boldsymbol{\vartheta})\,,\boldsymbol{\vartheta}],$$

$$\mathbf{z} \triangleq [\widehat{\mathbf{y}}\,,\,\boldsymbol{\vartheta}]\;.$$

Pdf $\overline{g}$ is joint pdf of random vectors $\widehat{\mathbf{y}}$, $\boldsymbol{\vartheta}$. Thus

$$\overline{g}(\widehat{\mathbf{y}}\,,\,\boldsymbol{\vartheta}) = g_U(\widehat{\mathbf{y}})\,\varphi(\boldsymbol{\vartheta})\;.$$

4 CONCLUSIONS AND FUTURE WORK

The proposed method presents general procedure for estimation of polygonal shapes with imprecise vertices in given cartographic coordinate system. The uncertainty of the polygon comprises uncertainty of classification and uncertainty of georeferencing. The resulting pdf of vertices of the required polygon is given by equation (6).

The unified methodology that was applied is based on Bayesian approach. The proposed method is therefore purely probabilistic and manipulates with the input data in consistent, highly correct way.

The algorithm described in this contribution is very general and its implementation is computationally demanding. Design of effective numerical solution of the integrals is still under preparation.

Investigating different kinds of transformation $\mathbf{T}$ introduced in (1), namely nonlinear, and different families of probability distributions, namely $f_{U,\mathbf{q}}$ mentioned in (3) are subject of further research.

ACKNOWLEDGMENTS

This work was partially supported by grant project InGeoCalc of the Ministry of Education of the Czech Republic (No. 2C06028).

REFERENCES

Koch, K.-R., 1990. Bayesian Inference with Geodetic Applications. Lecture Notes in Earth Sciences, Vol. 31, Springer-Verlag.

Schmitz, A. and Morris, A., 2006. Modelling and manipulating fuzzy regions: strategies to define the topological relations between fuzzy regions. Control and Cybernetics 35, pp. 245–261.

Trivedi, K. S., 1982. Probability & Statistics with Reliability, Queuing and Computer Science Applications. Prentice-Hall.

Viertl, R., 1996. Statistical Methods for Non-Precise Data. CRC Press.