

SPIATIAL DISTRIBUTION OF CONTROL POINTS

Ricardo Vergara
Alejandro C. Frery
Julio d'Alge

INPE - DPI
Av. dos Astronautas, 1758
12201 - Sao Jose dos Campos, SP, BRAZIL

ISPRS - COMMISSION III

ABSTRACT:

We present some Monte Carlo results about the spatial distribution of the Control Points used in the geocodification of satellite images in Brazilian Amazonas.

This information aims at establishing criteria for the choice of number, quality and spatial distribution of Control Points (CPs) to be used in image matching techniques.

Some theory about the statistical analysis of spatial point patterns is recalled, as an aid to a future formulation of this problem in terms of an experiments design problem.

KEY WORDS: Accuracy, Algorithm, Image Matching, Image Processing, Registration, Simulation, Theory.

1. INTRODUCTION

The geometric accuracy of remote sensing satellite images plays a central role in most applications (as in Cartography, to name one), so these images are geometrically corrected before being used.

Moreover, when digital images are used, their geometric accuracy is improved by the registration with respect to a map.

In this image processing technique, clearly identifiable points (the "Control Points", or CPs) are sought in both the image and the map. Taking the second as a reference, and using the CPs, it is possible to estimate the -mathematical- mapping that corrects the digital image. Usually, a polynomial transformation is used, and the degree of the polynomial determines the minimum number of CPs needed to estimate its coefficients (Mascarenhas & Velasco, 1989).

The registration accuracy between an image and a map depends on the location accuracy and on the number and the spatial distribution of the CPs used to perform it (Ford & Zanelli, 1985). The same happens with the attainable precision in the evaluation of the geometric accuracy of an image.

Suppose that is possible to have CPs with arbitrarily high location accuracy, either to evaluate the geometric precision of an image or to accomplish the registration between an image and a map. Precision will be highly dependent on the number and spatial distribution of the CPs. Concerning to the number, it is possible to say that the higher the quantity of CPs, the higher the achieved precision (Ford & Zanelli, 1985; Orti, 1981).

Relating to the spatial distribution, a specification of the adopted criterion has not been detailed reported in the literature, but in general is considered that a uniform distribution of -evenly spaced- CPs is the most appropriate choice (Ford & Zanelli, 1985). However, it usually is difficult to find CPs distributed in this way; sometimes the CPs are grouped in a small part of

the area under study or a very small number of dependable CPs is available. Obviously, all CPs must be used when the last situation occurs; in this case it is not possible to make a choice among them.

But suppose the case of having plenty of CPs of good location accuracy. In this situation the following question appears: Which should be selected in order to obtain a spatial distribution that makes possible the achievement of an accurate geometric correction? Or, if the registration has already been made with no good accuracy: How to know if that is due to a inadequate spatial distribution of the CPs used to perform it? And when this is the reason: Is there some manner to detect those CPs whose locations damage the quality of the geometric transformation? Note that the same questions are valid when the evaluation of the geometric quality of an image has to be assessed.

The aim of this paper is proposing some partial answers to these questions, providing a tool for either doing an adequate selection -concerning to the spatial distribution- of CPs to be used for performing an image-map registration or for evaluating the geometric accuracy of an image. This procedure allows the identification of CPs whose locations damage the quality of the spatial distribution of the set of points; so, their substitution for other CPs located in more appropriate places is made possible. The proposed methodology consists in selecting CPs whose plane coordinates are known with precision, and then in submitting these points to one or more spatial distribution tests. For the evaluation of this procedure, different registered images of Brazilian Amazonas were taken and their respective sets of CPs were submitted to the tests above. The Amazonian region was chosen for this study because it has special characteristics that make hard the finding of CPs spatially well distributed. To compare results, the same kind of analysis was carried out using a registered image of the Buritama region, Sao Paulo State, where the relative abundance of CPs allowed the selection of those whose spatial distribution is nearer to the optimal model.

2. DEFINITIONS AND NOTATION

Since we are interested in saying something (mathematically) about the spatial distribution of CPs, we shall introduce some basic ideas and notation.

Our main tool will be the comparison of the spatial distribution of the (say "n") given CPs with respect to the uniform distribution of (say "n", again) points in the unit square (that we will write "S"); whenever data do not come in this scale, the transformation is obvious.

The n CPs are said to be uniformly distributed in S iff their coordinates $(x_1, y_1), \dots, (x_n, y_n)$ are the outcomes of $2n$ independent identically distributed random variables with the uniform distribution in the $[0,1]$ interval (the reader should notice that these are not n evenly spaced points). We will test the (empirical) distribution of this set of n coordinates, against the hypothesis of uniform distribution over S, using simulation. The use of simulation is needed since there are no theoretical results about the tests we shall use.

Let d_j be the distance of the point j to its nearest neighbour; let $d' = (d_1, \dots, d_n)$, with $m = n(n-1)/2$ be the $n(n-1)/2$ dimensional vector of the nearest neighbour distances among the n points. Denote d' the vector d' with its elements sorted in ascending order, and d the vector formed by the first n elements of d' . We will compare the values of the n components of d with other values obtained by the sampling from the hypothesized distribution; this sampling is repeated a fixed number (M) of times ("iterations"), usually with $M=19$ or $M=99$.

3. THE ALGORITHM

The algorithm is as follows (Diggle, 1983):

- 1) For $j = 1$ to n do
 $\min(j) = 2$
 $\max(j) = 0$
 $\text{mean}(j) = 0$
- 2) For $k = 1$ to M do
 - 2.1) Observe $u_1, \dots, u_n; v_1, \dots, v_n$: $2n$ independent outcomes of a random variable uniformly distributed over $[0,1]$.
 - 2.2) Form the coordinates set for the current iteration $(u_1, v_1), \dots, (u_n, v_n)$.
 - 2.3) Calculate $e = (e_1, \dots, e_n)$, where e is a n -dimensional vector defined in the same manner as d above, for the current set of coordinates.
 - 2.4) For $j = 1$ to n do
 $\text{mean}(j) = \text{mean}(j) + e_j$
 if $e_j < \min(j)$ then
 let $\min(j) = e_j$
 if $e_j > \max(j)$ then
 let $\max(j) = e_j$
- 3) For $j = 1$ to n do
 $\text{mean}(j) = \text{mean}(j)/n$
- 4) Draw the plot
 $\text{mean}(j), \min(j), d_j, \max(j)$
 for every j between 1 and n .

5) End

4. INTERPRETATION OF RESULTS

For every j , $\min(j)$ and $\max(j)$ form an approximated confidence interval of level $M/(M+1)$ for d_j , obtained by simulation. $\text{mean}(j)$ gives a (rough, though useful) approximation of the mathematical expectation of the random variable E_j , whose outcomes e_j were observed. The plot j shows a confidence envelope for the input data: if the data lie in the region above the minima and below the maxima we accept the hypothesis of uniform distribution over the sets. If the data lie in the region below the minima, there is evidence favouring an "attraction" model for the location of CPs, i.e. they are more spatially clustered than independent (uniformly distributed) points. If the data lie in the region above the maxima, there is evidence favouring a "repulsion" model for the location of CPs, i.e. they tended to be further apart from each other than independent (uniformly distributed) points.

5. METHODOLOGY

The algorithm above is very simple to implement, and we feel that it might be helpful in the diagnostic and identification of problems regarding the spatial distribution of CPs. (Listings in the C++ programming language available from the authors upon request).

Given n , the number of CPs, and once obtained the values $\text{mean}(j)$, $\min(j)$, $\max(j)$ for every j , these envelopes could be understood as frontiers between the following regions: (i) the "acceptable" ($[\min(j), \max(j)]$); (ii) the "optimal" ($[\frac{1}{2}, \max(j)]$) and (iii) the "unacceptable" ($(0, \min(j))$).

Regions (i), (ii) and (iii) correspond to the labels we would attach to the data sets whose vectors d fall into, with respect to their spatial quality.

6. ANALYSIS OF THREE DATA SETS

In the following we used $M=19$ obtaining, thus, approximated 95% confidence intervals for d_j .

The first data set (W1) was used in the registration of an image of Tucuruí, Para State, Brazil. The whole image matching (including the evaluation of its accuracy) involved seven CPs whose spatial distribution exhibits a strong tendency to the "attraction", as previously defined (see Fig. 1a). The plot 1b shows that most of these points are located in region (iii), considered as "unacceptable".

The second data set (W2) comes from the area of Serra do Roncador, Mato Grosso State, Brazil. Ten CPs were used in this case, where an "attraction" model is also suited to describe their spatial distribution (weaker, though, than the W1 set; see Fig. 2a). The plot 2b shows that most of the points fall into region (i), considered as "acceptable", with a few shifted towards region (iii).

The W1 and W2 data sets correspond to regions in the Brazilian legal Amazonas, and are typical cases of the conditions found in that area: few or no features like crossroads, bridges, etc. preclude

the selection of enough CPs with good spatial distribution, being the minimal quality criterion that all points belong to region (i).

The W3 data set corresponds to a Buritama (Sao Paulo State, Brazil) image. In this case the selection of the nine CPs was easier, rendering an spatial distribution better fitted by a "repulsive" model (Fig. 3a). The plot 3b shows that the spatial quality of these points is mostly "optimal", or "acceptable".

Thus, the presented methodology allowed the assessing of the quality of the spatial distribution of these three data sets. Also, it suggests simple options to improve these qualities.

APPENDIX: COMPUTATIONAL INFORMATION

Hardware: SUN SPARK 1 STATION.
Language/Compiler: SUN C++ Version 2.0.
Special functions used: qsort provided in stdlib.h.
Pseudorandom number generator: g9827 (Bustos, 1990).
CPU time used: about one second for every presented result.

ACKNOWLEDGMENTS

The authors are grateful to Dr. Evelyn M.L. de Moraes Novo and to Ecologist Silvana Amaral for supplying the data sets of the Tucuruí and Serra do Roncador areas, respectively.

REFERENCES

- BUSTOS, O.H., 1990. A TURBO C implementation of a GFSR[0,1] pseudorandom number algorithm. Informes de Matematica B-62, IMPA, Rio de Janeiro, Brazil.
- DIGGLE, P.J., 1983. Statistical analysis of spatial point patterns. Academic Press, London.
- FORD, G.F.; ZANELLI, C.I., 1985. Analysis and Quantification of Errors in the Geometric Correction of Satellite Images. Photogrammetric Engineering and Remote Sensing, 51(11): 1725-1734.
- MASCARENHAS, N.D.A.; VELASCO, F.R.D., 1989. Processamento digital de imagens. Buenos Aires, Kapelusz, 1989. Quarta Escola Brasileiro-Argentina de Informatica, Termas de Rio Hondo, Argentina.
- ORTI, F., 1981. Optimal distribution of control points to minimize Landsat image registration errors. Photogrammetric Engineering and Remote Sensing, 47(1): 101-110.

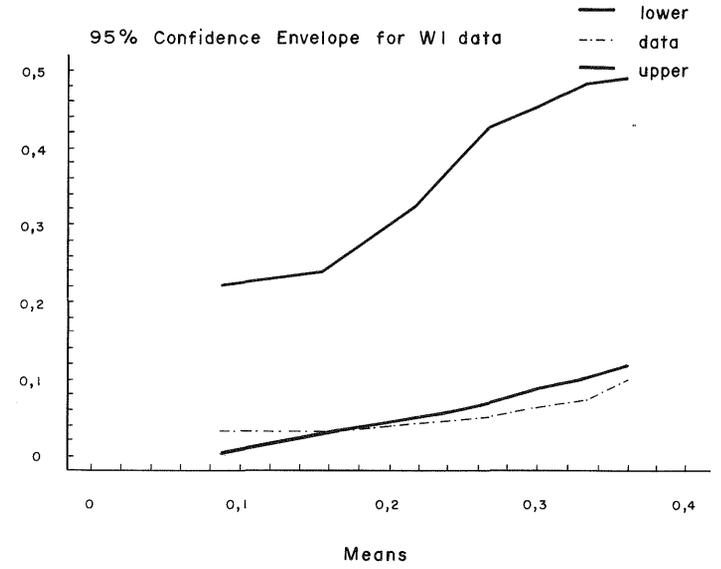
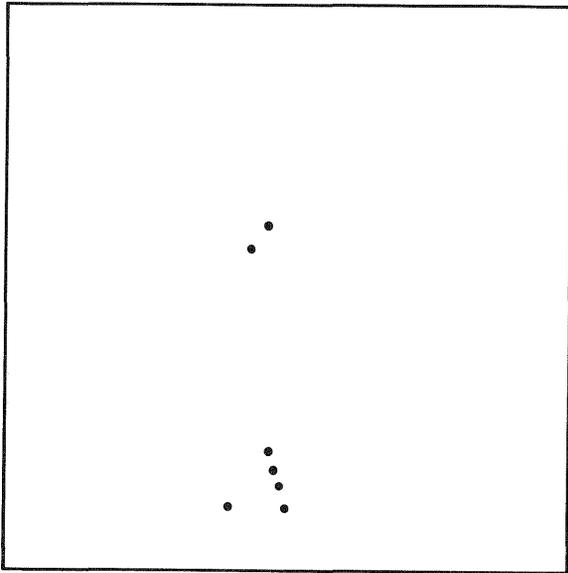


Fig. 1

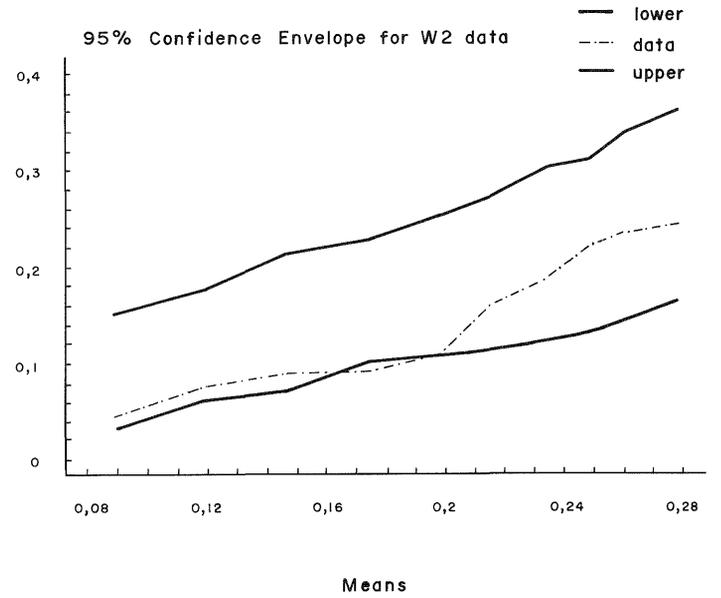
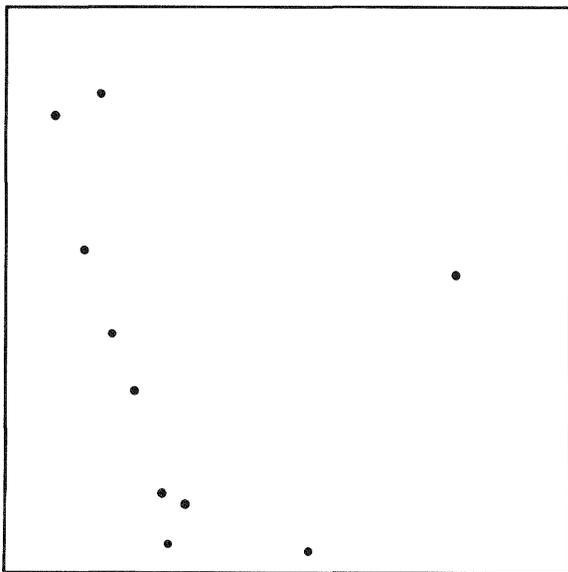


Fig. 2

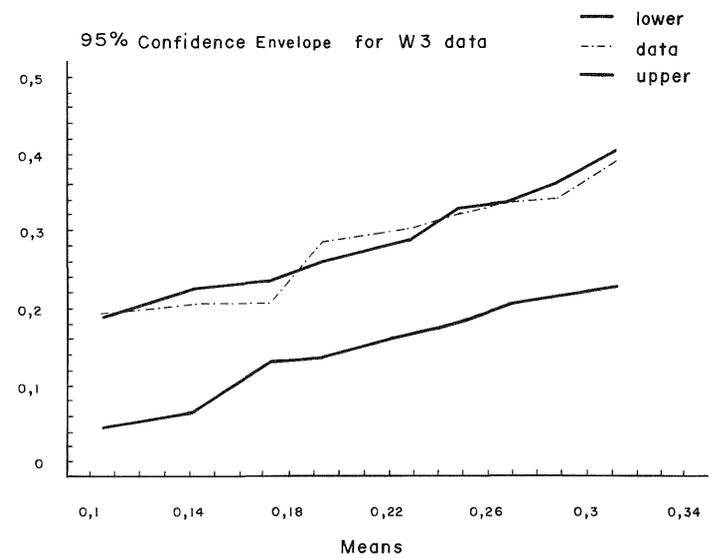
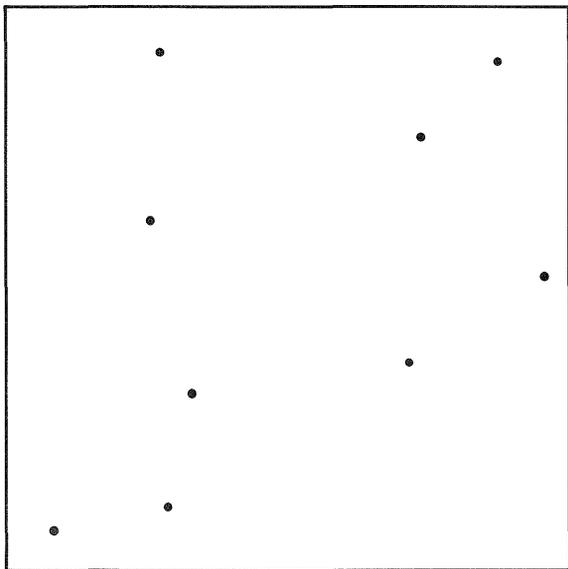


Fig. 3