SOME ASPECTS ON DATA PROCESSING IN REMOTE SENSING AND GIS

T.Bellone*, L. Mussio, C. Nardinocchi*****
***DIGET, Politecnico di Torino, Torino, Italy**
****DIIAR, Politecnico di Milano, Milano, Italy**
email luigi@IPMTF2.TOPO.POLIMI.IT
*****DITS, Università La Sapienza, Roma, Italy**

ISPRS Commission VI, Working Group 3

KEYWORDS: Data Modelling, Data Processing, Mathematics, Statistics, Computer Science.

ABSTRACT

Data processing collects several methodologies and procedures able to analyse a variety of data. In this frame, Remote Sensing and GIS/LIS present a set of soft and relatively low-cost technologies, where specific data processing is necessary in order to obtain the expected results. Referring to data processing, this paper is aimed to presenting an attempt to define a procedure which is central in several experiments.

1. THE EXPECTATIONS OF DEVELOPING COUNTRIES AND LESS FAVOURED AREAS

The world is becoming quickly a global village (McLuhan M., 1989). Nevertheless less favoured areas are broadly diffused and a large number of developing countries also exist, actually in all continents.(Chomsky N., 1993; Chomsky N., 1995).
A philosophy, keeping at the centre of the attention the perspectives of progress, not only in itself, but also connected to the human life, cannot accept ample areas of poverty anymore (Latouche S., 1989; Latouche S., 1995).
It is obviously not easy to say anything about the best way to obtain an effective progress. However it is sure that a correct use of the resources is preliminary to the achievement of the development (Meadows D., 1972; Mesarovic M., 1974).
The knowledge of phenomena and processes, in every part of the world, is a fundamental condition for the (correct) use of resources (Gabor D., 1976; Tinbergen J., 1977). In this frame, sciences and techniques offer proper contributions to this task.
Among the Survey and Mapping disciplines, the Remote Sensing and GIS/LIS are the most promising in order to acquire the early information (data acquisition, data validation)and to organise it (data modelling, data archiving).
Indeed where no data (or very few) are available until now and the amount of necessary information is very large, the time and the power to collect it could become problematic.
Furthermore these new technologies are soft technologies and relatively low-cost, on the contrary requiring specific data processing in order to obtain precise, accurate and reliable results.

2. AN INTRODUCTION TO REMOTE SENSING AND GIS/LIS

In the last twenty five years, the space exploration era and the information technologies have imposed an abrupt change in various fields of the sciences and techniques. Consequently methodologies and procedures in the Survey and Mapping disciplines have changed too.
Many new disciplines have appeared in this context and, among these, Remote Sensing and GIS/LIS. Their fields of application are very broad and, as above said, they could be applied to the specific aim.
Indeed various classes of satellites have been launched, e.g. the family of LANDSAT, SPOT, ERS, etc., and recently some commercial satellites. They are able to watch the earth in its different aspects, by using a lot of multispectral channels.
On the other hand, because Remote Sensing supplies a very large amount of data, there are problems to manage and to process them. The powerful tools of GIS/LIS offer many advantageous contributions to this task. Let remember also that GIS deals with more geometric aspects, whilst LIS deals with the more thematic ones.
Finally the success of these technicalities and their applications is due to their relatively low cost, because these are soft technologies, as already said, even if they require adequate data processing in order to obtain the expected results.

3. SOME ASPECTS ON DATA PROCESSING

A lot of new tools, like data compression and archiving, signal enhancement, data classification and understanding, hypertext and multimedia, need lots of data processing.
Methodologies and procedures used in this class of problems concern different kinds of data segmentation and matching, like clustering, parsing, vectorization, formalization, etc.
The complexity of this approach compels to define the set of data to be processed. Just at the beginning of the process, it is necessary to perform the optimal sampling and the outlier detection, in order to limit the amount of information and to define the message, within an immense data base. Rough data supply very poor information . Moreover anomalous data are not only blunders, but also data which carry information on phenomena and processes not related to the field of interest.
For these reasons, no previous hypotheses about the distribution behaviour of the data can be done: distribution-free inference (more general but less powerful) should normally replace classical statistical multivariate analysis.
At present time, many cases are solved in different ways, by means of different approaches, but the variety of the experiments is still chaotic.
The Bayes' theorem (Bayes., 1764):

$$p(A / B) = \frac{p(A) \cdot p(B / A)}{p(B)}$$

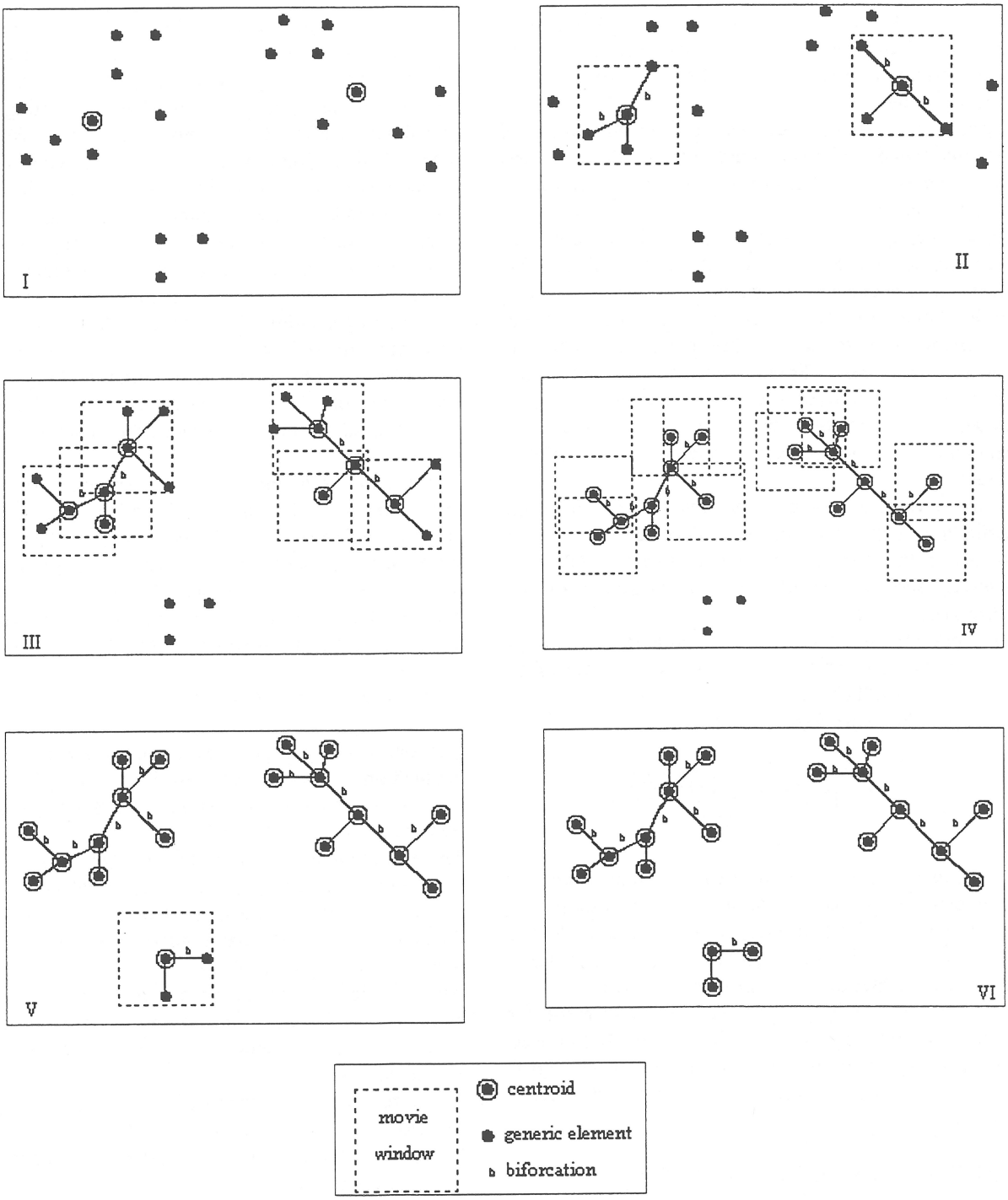
requires that the probability of the a posteriori estimate is larger than the probability of the a priori information. Indeed it is evident that every procedure is applied in order to have an a posteriori probability (and conditioned probability) greater than the a priori probabilities.
This theorem provide a criterium of judgement but doesn't give any suggestion on the way to be followed. The following paragraph means to present an attempt to define a procedure which is central in several experiments.

4. THE NEAREST NEIGHBOUR PROCEDURE

This procedure recalls a set of analogue operations of the human vision. They consist in the capability too watch the world from a given point of view, taking into account the closest objects (Arnheim R., 1954).
Notice that the relational operations, underlaying the techniques of human vision, involve the concept of proximity and likelihood (Arnheiim R., 1966). Indeed the recognition of the sorrounding area requires to identify and classify (likelihood) objects starting from the closest ones (proximity).
The procedure (see the enclosed figure) is initiated by a sample of tentative starting points (e.g. some modal values). In case of tentative starting points unknown or lacking, a generic element could be selected.
The selection of the nearest neighbour data is performed by means of a moving window around these elements. In such a way, new points are linked to the starting points. The function of starting points is then transferred from the original starting points to the new elements. Thus the procedure can be iterated in the same way as before.
The procedure is stopped, after having formed the chain of data related to each other, when no more data could be added and no isolated data remain in the sample.
Notice that, in case of isolated data present in the sample, the procedures should be newly repeated by any isolated element of the sample.
In order to obtain a more refined result, robust procedures of split and merge could be applied and iterated until finding a reproduction point in the final configuration.
To this aim, the definition of a dispersion measure allows to split the set, when the density of the data is too low and the definition of a distance measure lets merging different sets of data, which are too close to each other.
Notice that both the selection of the nearest neighbour and the fusion of the too close sets need to take into account possible bifourcations in the chain of elements.

5. FIELDS OF APPLICATION

The aim of this paragraph is to present a list of fields of application of the above described nearest neighbour procedure which is central in several experiments of Remote Sensing and GIS/LIS.
These applications show different levels of complexity, work either in a geometric frame or in a relational context and are unbedded in the space domain, in the space-time domain or in a particular kind of space with a proper metric and a proper topology.
An operation of image processing useful for feature extraction is called Line Following. This is not too complex, running exactly according to the above explained procedure.
A more complex operation, involving straightline segments (or circular arcs), is performed in Computer Cartography by transforming spaghetti into topologically consistent structures.
An operation of Remote Sensing useful for land classification is called Region Growing It is not too complex too.
A more complex operation, involving patches and edges is performed in Computer Graphics by tesselation of convex hulls and concave stellar or not- objects.
Different operations of image processing deal with the establishment of the correspondence in the geometrical and/or relational matching. There are many types of matching, whose list, comments and remarks are here omitted for sake of brevity.
Anyway one of the principal problems in the



The nearest neighbor procedure, illustrated step by step

establishment of this correspondence is the overcoming of occlusions and lavoyers.
Furthermore the operations of shape from shading and phase unwrapping are moving in this frame, when monoscopic optical imagery or micro-wave complex images are analysed respectively.
Concerning special kinds of space with a proper metric and a proper topology, they are the study of time series and 3D (or 2D) spatial analysis and the reduction to co-planar graphs.

6. PERSPECTIVES

The topics illustrated in this paper should be considered as a reflection, given by the authors, about some aspects on data processing in the frame of Survey and Mapping disciplines and particularly of Remote Sensing and GIS/LIS.
They start from the analysis of many experiments very well done by different authors, in the last fifteen years, and represent an attempt to define a procedure which is central in several experiments.
For these reasons, the exposition cannot be organised in a form which is classical in the field of Mathematics and Physics, like axioms, definitions, theorems, corollaries, counterexamples and remarks, leaving out the presentation of real examples to the specific disciplines.
Indeed, because the plenty of experiments, as already said, is still chaotic, the organisation of a complete and rigorous theory seems to the authors problematic. Consequently, they prefer to consider the Geomatics not yet a new science, but a set of interesting disciplines, methodologies, procedures and techniques very promising for the future.
In this context , the observations which have brought first to recognise the concepts of proximity and likelihood, and later to identify the nearest neighbour procedure as the best algorithm to implement them, could be taken into account, even if they are comparatively quite poor.
Regarding possible perspectives, the authors would state that the design of the future is not a science. Nevertheless analysing again the sample of experiments, some other concepts may be identified. The most important one seems to be the concept of rarity. Furthermore, some additional concepts of percepting grouping, like parallelism, collinearity and contiguity, could be taken into account.
Indeed among the closest objects (proximity) which are similar to a given model (likelihood), it is convenient to begin their analysis starting from the most rare ones (seldomless). This trick lets to avoid, as much as possible, hard decisions in heterogeneous sets of data, reducing the number of blunders in the data base.
The second group of concepts, deriving from percepting grouping theory, supplies more refined information and works at a higher level of complexity. As to the nature of the samples of data to be applied, they are not only rough and primary data, but also data acquired by different kinds of pre-processing.
A possibility to perform suitably the data transformation is given by bijective and surjective operations of data conversion. The former allows passing from a given space into a better configuration one, the second reduces suitably the dimensions of the transformed space to the ones of the original one.
However, since a reflection can be considered as acceptable only after that some results have been ascertained (in case, by a pilot experience), the authors hope that their reflection may be taken as a charge for a future engagement.

7. REFERENCES

Arnheim, R., 1954. Art and visual perception: a psychology of the creative eye, University of California, Berkeley.
Arnheim, R,. 1966. Towards a psychology of art. The Gestalt theory of expression, Univerity of California, Berkeley.
Bayes, T., 1764. Essays towards solving a problem in the doctrine of chances. Philosophical Transactions of the Royal Society of London.
Chomsky, N., 1993. Year 501. The conquest continues, South End Press, Boston.
Chomsky, N., 1995. La societad global. Educacion, mercado y democracia, Joaquin Mortiz, Mexico City.
Gabor, D., 1976. Beyond the age of waste, Report of M.I.T., The Club of Rome, Geneve.
Latouche, S., 1989. L'occidentalisation du monde. La Decouverte, Paris.
Latouche, S.,1995. La megamachine. Raison techno-scientifique, raison economique et le mythe du Progres. La Decouverte, Paris.
McLuhan, M., Powers, B.R., 1989. The global village, Oxford University Press, New York .
Meadows, D., 1972. Limits of growth, Report of M.I.T., The Club of Rome, Geneve.
Mesarovic, M., 1974. Mankind at the turning point, Report of M.I.T., The Club of Rome, Geneve.
Tinbergen, J., 1977. Reshaping the international order, Report of M.I.T., Elsevier, Amsterdam.

For the sake of brevity, the most important references, concerning Remote Sensing and related sciences, aren't quoted, because the readers are supposed to be very well acquainted with the specific literature.