AREA MATCHING IN RASTER MODE UPDATING

J.P. GRELOT, P. CHAMBON and F. MARCHE
Institut Géographique National
136 bis rue de Grenelle
75700 PARIS - FRANCE

ABSTRACT

Cartographic data bases updating is a real challenge for the maintenance of the homogeneity and reliability. The choice has to be made between interactive editing and automatic procedures. An attempt has been made by IGN France on the occasion of a land-use inventory updating. More precisely, the aim was to adjust a new delineation of land use areas to an old one. Digitization is performed in raster mode through a scanner. After skeletonization, control points are interactively identified in order to calculate a global transformation which is applied to the vectorized boundaries, and eliminates main distorsions due to variations in drawing materials. Then the boundaries are processed in raster mode in order to correct new lines which are exactly adjacent to old ones, or which define with an old line a micro-area. The results seem rather good but introduce local distorsions due to the very local approach of the process. To avoid such distorsions, we have to improve the software in using shape considerations in replacing very local operations by more regionalized operations. It is an interesting example of the successive or simultaneous use of several data structures to solve an updating problem.

INTRODUCTION

When a cartographic compilation is made to follow the evolution of a phenomenon, the updating process must be forecasted. Digital cartography allows us to make it easier, solving the problem of bad conservation of documents. However, in the case of a land use inventory, updating informations have to be collected carefully, because they cannot be combined directly with the device that will be used to produce the map : the file on the computer. Therefore, problems appear when trying to superimpose the updating files with the old ones, producing poor cartographic results, incompatible with the rules for representing the phenomenon and which do not have any real meaning. This problem of misregistration of digital areal information has been solved at IGN France for a project concerning a series of maps.

UPDATING A LAND USE INVENTORY

The area matching project began three years ago when we had to update a land use inventory known as "the French Littoral Inventory", which has been designed as a management tool with a periodical depiction of the area of interest (Grelot 1982). Let us have a few words about it. The aim is creating some objective data on the coastal area in order to help in leading a protection policy in this area. You may regard France as a Far-East country, but you also may know that its coastal area gathers about one tenth of the population and has a major touristic appeal for many European people coming there from cloudy and cold countries. The basic tool we designed is a land use digital map at 1:25 000 scale made from aerial photographs interpretation and from subsidiary data compiled by local administrations (Grelot and Chambon 1986).

We defined an initial coverage with about 400 000 polygons and the legend consisted in about thirty land-use categories which had been selected after a long discussion involving many people from local and governmental administrations as well as from technical organisations. However, with the availability of the initial map coverage, some people requested to re-define the legend. In some cases we could subdivide some categories and it was only a change in the attributes allocated to the polygons. But in many cases the result was a new set of polygons we had to substitute to the former one. And everyone knows that modifying boundaries is much more difficult than allocating attributes.

On this background makes the concern in terms of area matching easily understandable. We had to update the data files in two ways : the first one was due to real changes in the countryside during the five-year period and the second one was due to a new classification mainly in urban areas and also in agricultural areas. Because we had a huge number of polygons, we decided to use automated processes instead of interactive ones and we developped an area-matching facility.

However a major hypothesis was absolutely wrong. We thought that the main reason for distorsion during the drawing was graphic uncertaincy which we agreed upon because of the average size of polygons. In fact we had very bad misregistrations due to three main sources : (a) dimensional variations in the base map used as the background for drawing the boundaries, (b) poor drawing quality, and (c) graphic uncertaincy. In that respect a scale factor and some local processing were not sufficient for matching archived and up-to-date boundaries, and we had to design a new software package.

The main requirements for the matching package were :
(a) automated processing with only minor interactive tasks and no editing;
(b) input and output data files in raster mode ;
(c) defining then computing local geometric distorsions ;
(d) preserving old boundaries ;
(e) local capture of up-dated boundaries;
(f) automatic re-allocation of land-use attributes.

THE AREA MATCHING PROCESS

This process deals only with zonal information. These data constitute files in raster mode, made of homogeneous and connex zones where one or more attributes are given.

The initial data set is used as a reference and considered as geometrically exact. The updating document contains only the new information. The boundaries of the new closed areas are then hand-drawed and digitized on a drum scanner. The features are skeletonized and some editing is made on a graphic workstation in order to avoid errors due to digitization and to close every area. This file is then superimposed to the old state one, in raster mode, and the operator can see the quality of registration. Homologous points (20 to 30) are identified. The coordinates of this points are taken interactively in order to calculate a global deformation between the old and the new boundaries.

Here begins a series of functions which are processed automatically one after the other.

The global deformation is represented by a biquadratic polynom whose coefficients are computed by the least squares adjustement method. By applying this polynom to the new boundaries after vectorization, we make them fit the old ones, with the best accurary on the control points.

A check plot is made after this transformation showing :
a) the boundaries of the old areas.
b) the boundaries of the new areas (having changed between the two states).
c) the same boundaries as b), having undergone the deformation.

We can also see the position of the control points on this plot, made on a large format electrostatic colour plotter. That is where we can notice that the final boundaries are well fitted to the old ones near the control points. On the other hand, where fewer control points have been identified, a misregistration remains. But everywhere, this error is reduced to less than half a millimeter, though it could exceed two millimeters initially. If this is not the case, we have to take other control points an remake the process.

At this stage, we can see when superimposing the new and the old boundaries that only local errors are still present and are represented either by small areas (areas having a small surface) or by prolate areas (one or some pixels wide). These parasite zones do not have any real thematic significance and must be suppressed, the aim always being to keep the old boundaries when there is a conflict. These suppressions are made by a special software which has several effects :
a) It detects areas whose surface is less than one square millimeter, or which are less than 4 pixels (0,5 millimeter) wide,

b) an iterative algorithm forces the new boundaries to be just near the old ones (distance is exactly one pixel) along these micro-areas,
c) when such sticked boundaries are encountered, they are merged in favour of the old ones.

Another check plot is made at this stage which shows the final results ; on this plot where the areas are labelled, we can also identify corrections to be made, due to other reasons.

The whole process requires approximately 5 computing hours, and some additional time to check the plots. During this time, up-to-date attributes are interactively given to the new areas defined by the digitized boundaries (only those having changed since the old state). After the matching procedure, the old and the new boundaries are superimposed, defining new areas which all have a thematic significance. These new areas receive automatically the old and the new thematic attributes. That is where we stop the matching process, maps being made as for an usual work (Grelot and Chambon 1986).

ADVANTAGES AND DRAWBACKS OF THE METHOD

The matching procedure merges neighbour boundaries, i.e. made with adjacent pixels, and eliminates small linear areas which look like spaghettis with two parallel boundaries, one old and the other updated. The surface of such areas to be eliminated is less than a given threshold value.

Of course this automatic procedure is not perfect. The tolerance after the distorsion is about two to three pixels which means about a quarter millimetre. There may be some misinterpretations with linear features such as rivers and major roads and railroads. They are removed during the final editing stage.

You may have noticed that any ckecking is made from a graphic plot. In other words it means that only those defaults with graphic significance can be seen. In fact it is not realistic to display every file through small windows with a large zooming factor in order to detect any default. And we all know some defaults which do exist in raster boundaries after skeletonizing hand-drawn lines : there are some random waves and on the other hand some systematic smoothing upon angular shapes. The matching process is not responsible for them and does not try any improvement.

The matching process is perfomed on small areas with a pixel-by-pixel approach. Not only during the comparison stage, but also during the decision and drawing stages. And there is a need for a smarter solution in the area separating a to-be-changed from a to-be-kept boundary set. It means that the local approach induces defaults in lines because the lines are considered as sets of adjacent pixels and not as features with particular shapes. Those defaults are made of very short segments joining the old and the new locations of the two parts of a boundary : they do not create graphic troubles but they are logical errors in the depiction of lines.

Finally the procedure :
(a) adjusts the updating file to the old one through a smooth distorsion to reduce misregistrations to the order of magnitude of graphic uncertaincy ;
(b) matches the boundaries pixel by pixel ;
(c) increases some defaults of raster files such as random waves and smoothed angles.

I do not think that a vector-mode process would give better results. On the contrary I do consider that this process successfully combines vector-mode stages and raster-mode stages. It proves that the link between data structures and processes is more important than the link between structures and models, and how interesting is the capability for using several data structures in the very same data base depending on applications.

Let us have a come-back to look for improvements. First the graphic quality could be improved. A land-use pattern has not the same characterics as a geological one : it is more related to the land ownership and to parcels with straight boundaries and precise corners. There is a close connection between the geometric characteristics of a boundary and the nature of adjacent areas, in other words between geometric and semantic describers. As an application for our purpose we should look for algorithms enhancing the raw data set and getting more straight boundaries and sharper angles. I can add a comment : as this geometric processing is due to the very nature of areas, it can only be performed after attribute allocation.

Second point : the boundary patterns are not similar in the two files. It is evident because the second data set is only drawn for updating the first one, but it has to be reminded and it has some consequences. The main one is that we cannot select a set of points (for instance the nodes) within the updating file and look for the geometric distorsion which maximizes the correlation with a corresponding set of points from the old file. It also means that the result of the distorsion is very dependent upon the interactive selection of control points and of some bad effects of skeletonization. It would have been useful to get a network of fiducial marks displayed on both lineworks for automating and improving the mathematical distorsion.

Third point : a major quality achievement should be made when explicitly considering the shapes of features. This is not easy. It means that the local processing which now removes points should be replaced by only a comparison which determines a displacement to be applied on a boundary or an entire set of boundaries. There are some requirements for making this : for instance, the nodes and corners of polygons and some geometric characteristics of the shapes have to be known in order to avoid large modifications of shapes. The displacements have to be designed as a mechanical stress upon a solid network.

CONCLUSION

I think that these three points are the main directions for improving the area matching procedure. We have seen how fruitful is the capability for using several data structures during a particular process. We also have seen that local processing has to be complemented with a more global approach using the geometry of shapes.

These remarks may look pessimistic. Our process does not work as perfectly as we could expect, but it actually works. We use it for superimposing data sets on the same area related to the same kind of phenomenon such as a land-use updating or related to different matters such as land-use and soils.

A lot of interactive editing is replaced by an automatic process which efficiently eliminates noisy areas created when superimposing the data sets. And this was the purpose: any cartographic superimposition or updating induces a lot of matching work, and the digital challenge is not performing the matching on editing workstations but automating it as much as possible.

REFERENCES

Grelot, J.P., 1982, An Inventory of French Littoral: Auto-Carto 5 Proceedings, pp. 367-373.

Grelot, J.P. and Chambon, P., 1986, Up-dating a Land Use Inventory : Auto-Carto 7 Proceedings, to be published.