BRINGING REMOTELY SENSED DATA CLOSER TO THE USER: THE CAMOTIUS EXPERIENCE

Frans J.M. van der Wel
Utrecht University
Faculty of Geographical Sciences, Cartography Section
P.O.Box 80.115, 3508 TC Utrecht, The Netherlands
e-mail: f.vanderwel@frw.ruu.nl

Abstract
This paper presents an overview of a number of components of an operational CAMOTIUS classification and monitoring tool currently in the making at the Cartography Section. Supported by a number of illustrations, the extra value of user support by means of visualized meta-information is emphasized.

1. Introduction

Earth observation satellites provide valuable source data for a number of cartographic applications. The relevance of remote sensing for thematic mapping applications has been the subject of a number of studies, and can be described according to an exclusive, complementary and substitute use of satellite data ([2]). Besides thematic maps, an increasing number of topographic and image maps demonstrate the importance of satellite image derivatives. The inability to meet the world-wide need for up-to-date cartographic material by means of "traditional" approaches has been mentioned as an incentive for the exploration of satellite remote sensing data by cartographers ([1]). But the involvement of cartographers in remote sensing exceeds the boundaries of mere presentation. Instead, the information process as a whole deserves the attention of such an independent spatial data expert who could provide (preferably visual) decision support.

In deriving information from remotely sensed data, users tend to rely heavily on the functionality being offered by the information system under consideration. Without sufficient knowledge about the potential of data and methods, the serious danger of inducing a *trial-and-error* process is lurking, fueled by a lack of insight into the requirements of the anticipated information.

As a consequence, the justified acceptance of a particular processing result, for instance a land cover classification, is largely dependent on the skills, experience and assumptions of the operator c.q. user, which vary not only per person, but over time as well. Obviously, this dependency hampers the repeatability and objectivity of the information extraction procedure. Instead, it contributes to an ongoing abuse of remotely sensed data in numerous applications - as a consequence of the inability to prove the extra value of the data. Ignorance of the physical characteristics and statistical meaning of these sampled data could result in a senseless, meaningless procedure with possible disappointing output that harms the credibility and acceptance of remotely sensed data in otherwise promising applications.

In this paper it is stated that a certain level of user support is required in order to improve the information content of the processing results. By increasing the understanding of the potential of remotely sensed data and, as such, bringing these data closer to the users, the role of cartographers as spatial information engineers is underlined. Attention is especially drawn to the extraction of land cover information from remotely sensed data by means of classification.

2. The CAMOTIUS project

Within the framework of the research project CAMOTIUS[1], the need for quality information before, during and after the classification of remotely sensed data has been addressed. This type of meta-information offers guidance to users of an information system by evoking rules that direct them to the required/desired information. By matching user requirements and data characteristics (distinguished according to quality components), selection, classification and interpretation of remotely sensed data can be improved.

The main objective of CAMOTIUS is the development of an operational tool that, as part of an information system, provides a grip during the extraction of particular land cover information. This concerns both static (inventoring) and dynamic (monitoring) information. Before selecting source data, a user can explore the available data by considering *visualized* quality information, reflecting the information potential of the database. During *classification*, likelihood vectors are derived which can be

---

[1] CAMOTIUS is both an acronym, meaning CArtography-Assisted MOnitoring - Towards an Interactive Userfriendly System and the Latin name of sixteenth-century Venetian cartographer Giovanni Francesco Camoccio. Thanks to the financial support of the Netherlands Remote Sensing Board, the Cartography Section of Utrecht University is able to cooperate with the International Institute for Aerospace Survey and Earth Sciences (ITC), the Dutch Physical Planning Agency (RPD) and Eurosense b.v.

visualized in order to communicate the strength of class assignments. The *evaluation* of these likelihood vectors can be improved by introducing cost functions, defining the - financial - risks attached to possible misclassifications. Figure 1 gives a complete overview of the components of the CAMOTIUS framework.

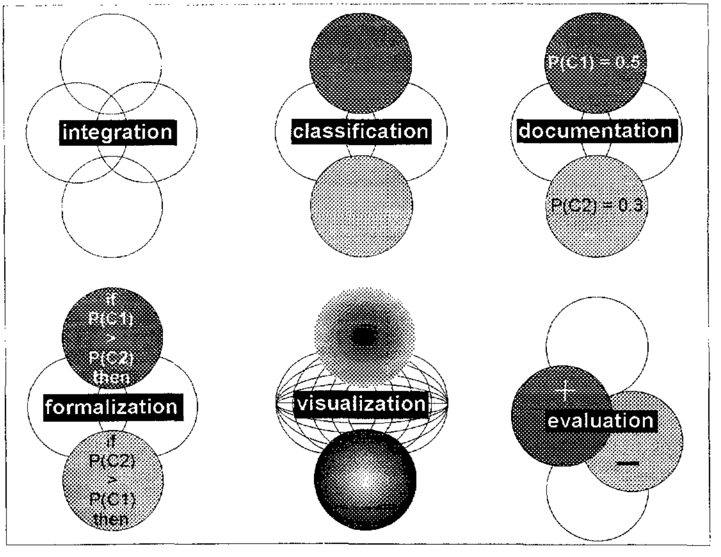

Figure 1: Components of the CAMOTIUS framework

In the following sections these components will be further elucidated as the basis of an anticipated classification tool. Land cover monitoring will be emphasized as one of the applications that could benefit from the proposed strategy.

3. Monitoring as a post-classification comparison

Monitoring is one of the most powerful applications of remote sensing because it takes full advantage of the strong points of this technique: frequent, objective, cost-effective and informative data acquisition. The focus of CAMOTIUS is primarily on the derivation of land cover as a basic information layer in numerous applications. The resulting classifications provide the starting points of a monitoring procedure based on a *post-classi-*

fication comparison. This simply means that the (re)classified remote sensing images are subjected to an overlay procedure that reveals possible changes in land cover between two or more points in time (figure 2).

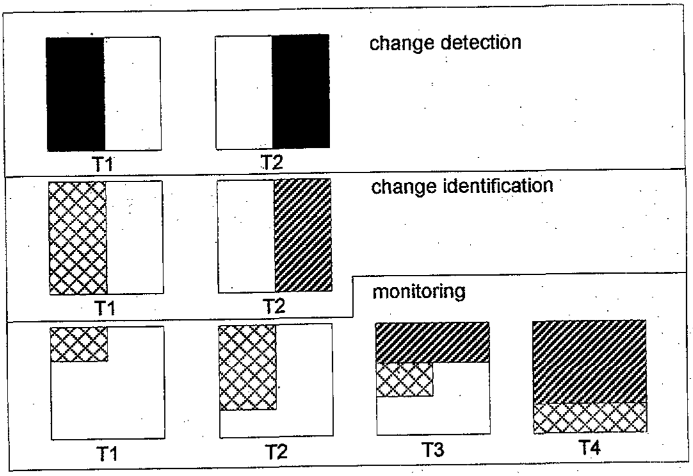

Figure 2: Monitoring as a post-classification comparison.

Monitoring according to this principle requires the derivation of highly accurate classification results in order to reduce the occurrence of *noise*. As a more realistic alternative, though, knowledge about the quality of these source data as well as about the manifestation of particular changes could contribute to a more reliable interpretation of monitoring results. This requires, however, the availability of methods and rules that support...

* ...the derivation of quality information during classification (*"what is the fitness for use"*);
* ...the definition of the risks of wrong class assignment (*"what are the - financial - consequences of confusing classes A and B"*);
* ...the improvement of the classification (*"what must be done in order to conform the results to the user requirements"*);
* ...the evaluation of changes in monitoring results (*"what is the probability of this increase in urban area"*);
* ...the communication of quality and change information to a user (*"communication by visualization"*).

4. Classifications and quality information

Quality reveals information on the appropriateness of data for a certain application. Accuracy, both positional and thematic, is an example of a quality component. From this, it can be stated that instead of maximum accuracy, sufficiently high accuracy is pursued. In order to make the distinction between appropriate and inadequate classifications, user requirements as well as quality statements are needed. A quality statement can be obtained by deriving an error matrix, but this requires some reference data set or field work. As an alternative, a statistical measure can be derived, to wit the *a posteriori* probabilities being calculated as part of a *Maximum Likelihood Classification* procedure (figure 3).

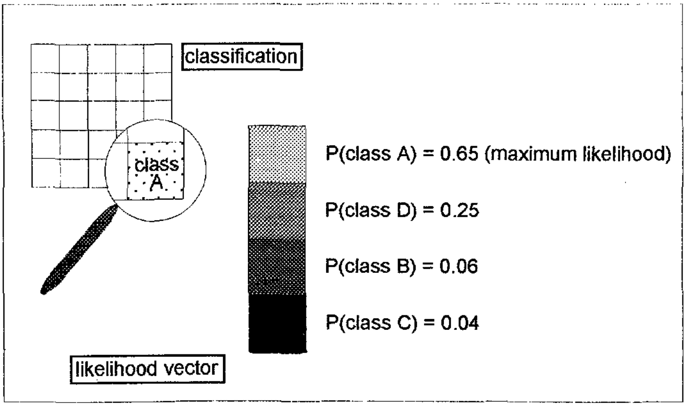

Figure 3: The classification procedure

This classification method attaches a *likelihood vector* to each pixel of the data set, although in general only the maximum likelihood is reported and stored in the information system. CAMOTIUS, on the contrary, provides all likelihoods, as components of likelihood vectors which reveal the strength of the class assignment, possible confusion and incomplete classification schemes. Moreover, they could reflect the extra value of additional data being used during the process (*a priori* knowledge). It is important to note that these probabilities fail to report absolute accuracy information!

The evaluation of this probability information needs the counterweight of user requirements. From experience gained during a number of case studies, it appears that users find difficulties when asked for the limiting conditions of the desired information. Therefore, a call for an extremely high accuracy is not seldom a disguised expression of a lack of understanding. It is, however, difficult to define a relation between application and required accuracies or probabilities. But very often, a user is able to specify the risks of wrongly assigning pixels from one class to another. These risks, or costs if expressed in a financial way, can be summarized in a matrix (figure 4) and together with the probability information this results in a cost image. This leads to a critical consideration of the classification results and possibly radical changes in the procedure to which the user used to adhere.

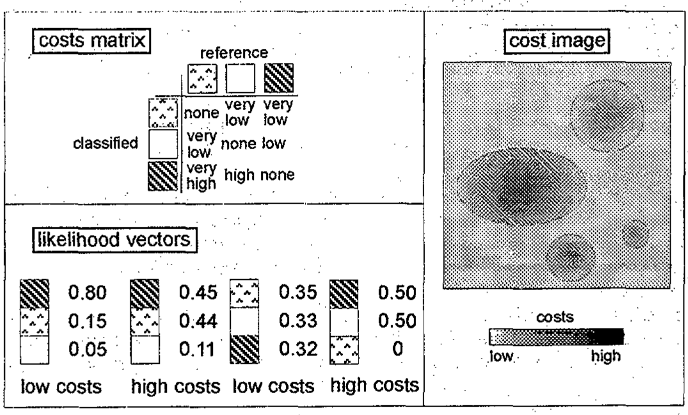

Figure 4: The derivation of a cost image

5. The improvement of classification results

After evaluating a classification a user could decide to improve the result by making an appeal to the extra value of additional knowledge. The Maximum Likelihood Rule allows the definition of *a priori* probabilities, reflecting the likelihood of occurrence of a certain information class in a particular area. The formalisation of ancillary knowledge (from maps, aerial photographs, existing remote sensing classifications, experts, personal experience) is required in order to arrive at these *priors*. Several ways to consider map information in the classification process are described in the literature

([5]). They are often based on the principle of making a stratification according to a certain variable (e.g. soil type). Each of these strata has its own set of priors. In addition, uncertainties can be taken into account as well (figure 5). In view of an optimal classification accuracy (minimum costs, maximum benefits) the following question arises: how to decide on the type and amount of additional data and at which stage of the process? CAMOTIUS attempts to help a user by providing a number of guidelines, such as:

* apply additional data according to a strategy of increasing costs, so start with obvious, ready available, easy-to-process and cheap alternatives - look to earlier classifications first (as stored in the GIS), before purchasing new and expensive data;
* exploit analog approaches before embracing complex (digital) equivalents - interactive, visual editing of a classification is less rigorous and more customized than an automatic majority filter procedure;
* evaluate the *a posteriori* probabilities to monitor the possible improvement of the classification result - a considerable difference between maximum and second likelihood could be an indication of less confusion (figure 3);
* interpret uncertainty patterns (from the likelihood values) in order to assess spots requiring special attention - additional knowledge concerning these areas needs to be collected urgently only if they appear to be essential.

It is obvious that the general suggestions mentioned in the above need some further elaboration, dependent on the application under consideration. CAMOTIUS provides a grip

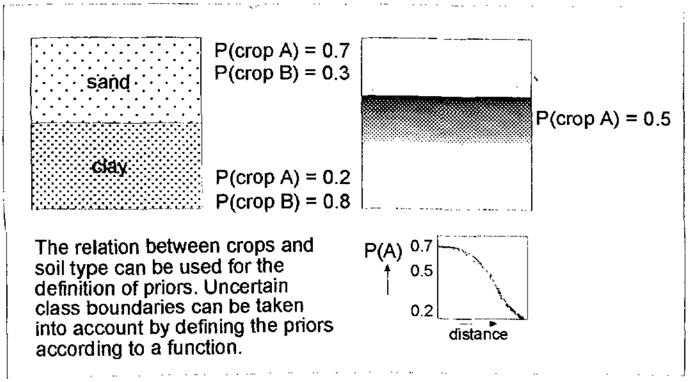

Figure 5: Improving classifications by *a priori* knowledge

for the classification and monitoring of greenhouses as an important expression of "agribusiness", both from a planning and economic point of view. With respect to these objects, guidelines have been formulated based on available and suitable additional data in some selected user environments (planning agencies).

6. The interpretation of monitoring results

After classification, interpretation of the resulting image requires guidelines as well. As pointed out in section 3, a monitoring approach based on the comparison of two or more classifications, links up best to the idea underlying CAMOTIUS: providing user support in order to arrive at optimal classifications, taking advantage of data and meta-data stored in a GIS environment. As a consequence of this monitoring approach, errors and uncertainties are dependent on the quality of the original classifications. Figure 6 shows that consultation of these "source" data and their meta-data contributes to a better interpretation, and therefore justifies the storage of these data.

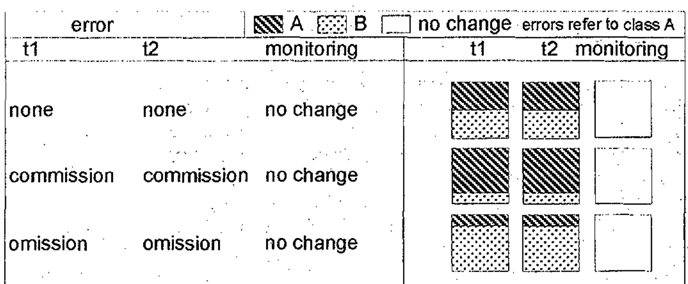

Figure 6: The interpretation of monitoring results can benefit from the original classifications

In addition, interpretation rules can be applied to make the distinction between changes and artefacts. Recognition of the latter is directed by knowledge about the considered information class, e.g. object dimension, shape, manifestation and dynamics (rate and direction of change). Figure 7 illustrates the role of interpretation rules in case monitoring of greenhouses results in three classes: decrease, no change and increase. The original classifications provide valuable back-ground information concerning the character of the change: from pasture to greenhouse or from greenhouse to urban, for example.

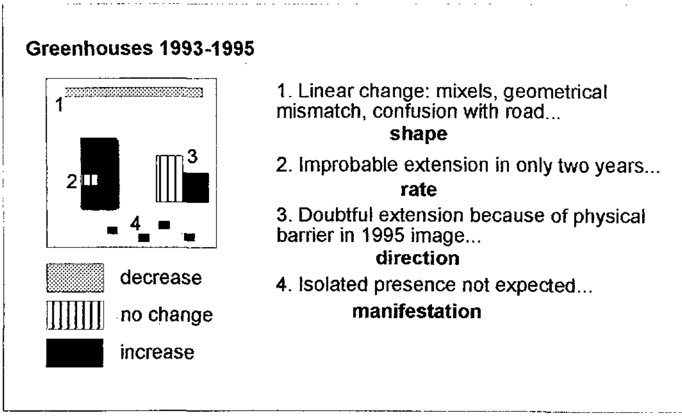

Figure 7: Monitoring of greenhouses: examples of interpretation rules

7. Visualization

Quality information, mainly expressed as probabilities, is a key factor in supporting users during the classification and monitoring process. A proper use of this meta-information hangs on the effectiveness of the way in which it is offered to the user. Viewed in that light, the recent emphasis on cartographic visualization is understandable ([3],[4],[6]). It enables the exploration of the intrinsic value of geographical data and the communication of information about the validity of these data. The probability information derived by CAMOTIUS can be visualized in several ways; bivariate maps revealing both classes (colour hues) and maximum likelihoods (saturation or intensity), black and white probability images, risk images (see figure 4) or dynamic sequences of visualizations representing attribute and positional class uncertainty (figure 8a). As an extension, change information can benefit from dynamic visualizations as well; in a monitoring image, areas characterized by remarkable changes can be highlighted by blinking. Obviously, a transition from water in 1990 to greenhouses in 1995 is dubious, but a change from urban area to greenhouses based on classification likelihoods of 0.3 and 0.4 respectively, is doubtfull as well (figure 8b).

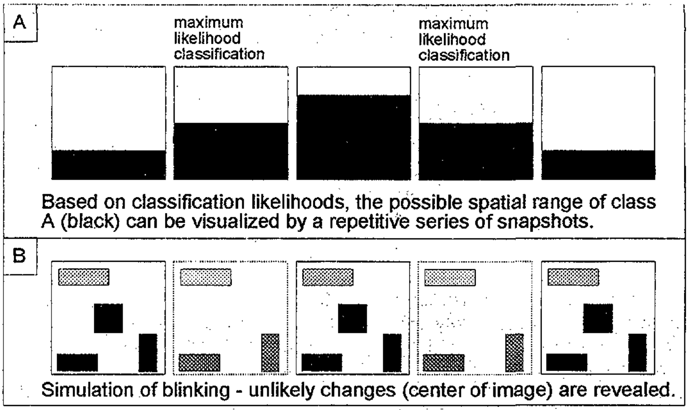

Figure 8: Visualization of data quality
a. monotemporal information b. monitoring information

8. Concluding remarks

The effectiveness of promoting the justified application of remotely sensed data by providing extensive visual user support needs to be tested in a number of selected land cover classification strategies. At the moment, a prototype of CAMOTIUS is still under development, but it is hoped that test results can be revealed at the conference. The bottom line of the CAMOTIUS experience so far, is the catalytic role of cartographic visualization in accessing and exploring information from remotely sensed data.

References

[1] Albertz, J. (1991): *Grundlagen der Interpretation von Luft- und Satellitenbildern. Eine Einführung in die Fernerkundung*. Darmstadt: Wissenschaftliche Buchgesellschaft.

[2] Ciolkosz, A. & A.B. Kesik (1994): Satellite Remote Sensing Imaging and its Cartographic Significance. In J. Denègre (ed.): *Thematic Mapping from Satellite Imagery*. pp. 1-29. Oxford: Pergamon.

[3] Goodchild, M., B. Buttenfield & J. Wood (1994): Introduction to Visualizing Data Validity. In H.M. Hearnshaw & D.J. Unwin (ed.): *Visualization in Geographical Information Systems*. pp. 141-149. Chichester: Wiley & Sons.

[4] MacEachren, A.M., B. Buttenfield, J. Campbell, D. DiBiase & M. Monmonier (1992): Visualization. In R.F. Abler, M.G. Marcus & J.M. Olson (ed.): *Geography's Inner Worlds: Pervasive Themes in Contemporary American Geography*. pp. 99-137. New Brunswick, N.J.: Rutgers University Press.

[5] Strahler, A.H. (1980): The Use of Prior Probabilities in Maximum Likelihood Classification of Remotely Sensed Data. *Remote Sensing of Environment*, 10, pp. 135-163.

[6] Wel, F.J.M., R.M. Hootsmans & F.J. Ormeling (1994): Visualization of Data Quality. In A.M. MacEachren & D.R. Fraser Taylor (ed.): *Visualization in Modern Cartography*. pp. 313-331. Modern cartography series, volume 2. Oxford: Pergamon.