

# Implementation of Temporal Relationships in Knowledge Based Classification of Satellite Images

Hans Middelkoop\*

International Institute for Aerospace Survey and Earth Sciences (ITC), P. O. Box 6, 7500 AA Enschede, The Netherlands

Lucas L. F. Janssen

The Winand Staring Centre for Integrated Land, Soil and Water Research, P.O. Box 125, 6700 AC Wageningen, The Netherlands; Fellow of the Department of Surveying and Remote Sensing, Wageningen Agricultural University, The Netherlands.

**ABSTRACT:** Classification of digital satellite images involves a process similar to classification problems in artificial intelligence. In a knowledge-based classification, ancillary data and knowledge are combined with spectral information. A method of knowledge-based classification based on temporal relationships between classes is introduced. Knowledge about crop rotations is represented by means of state transition matrices. Spectral image information, information stored in a geographic information system, and knowledge as represented in a matrix are combined in a Bayesian maximum-likelihood classification. This method is elaborated for a test area in The Netherlands. Depending on the spectral separation of the classes and the level of detail of the transition matrices, the overall accuracy of the classification increased by 4 percent to 20 percent with respect to the result based on only spectral information.

## INTRODUCTION

IN THE EARTH SCIENCES, digital satellite images have become an important source of information. For many survey applications, these images are spectrally classified into classes which are relevant for the user. When the recognition of classes on the basis of their spectral characteristics becomes difficult, other types of data (thematic or geometric) are included in the classification process. Because this requires knowledge about the relationships between classes and the different ancillary data sources, the latter category is called "knowledge-based classification" (e.g., Skidmore, 1989). Knowledge engineering techniques used in expert systems have common grounds with pattern recognition and can be applied for image interpretation and classification (Middelkoop *et al.*, 1989). The development of a procedure for knowledge-based classification involves knowledge acquisition and requires a knowledge representation formalism.

This paper presents a specific method of knowledge-based image classification. First, a short review of image classification and knowledge-based classifications is given. Then a method of knowledge-based classification based on temporal relationships is presented using a case study in a test area in The Netherlands. It includes (1) knowledge acquisition, (2) knowledge representation, (3) implementation of the classification procedure, and (4) evaluation of the results.

## SPECTRAL IMAGE CLASSIFICATION

After pre-processing of the raw data and extraction of proper features (Mulder, 1985) from the original spectral bands, images can be separated into classes which are significant for the user. Classification of a satellite image involves a decision making process that assigns a class label of real-world object to a raster cell (pixel) on the basis of its spectral values (represented as an observation vector).

\*Present address: Institute of Geographical Research, Faculty of Geographical Sciences, University of Utrecht, P.O. Box 80115, 3508 TC Utrecht, The Netherlands.

When the classes are spectrally well separable, a box-classifier can be applied using a feature space spanned by the different spectral bands. This is equivalent to Boolean IF <condition> THEN <conclusion> rules. Delineation of a box in this space is done by means of an AND function in the condition part; merging boxes is equivalent to an OR expression (Middelkoop *et al.*, 1989).

The usual problem, however, is the spectral variability of the classes and the overlap between the clusters in the feature space. This requires finding the most likely class and providing an estimation of the associated probability of error. An intuitively satisfying and mathematically manageable classification theory is the maximum-likelihood or Bayes' optimal classification (Mather, 1987): i.e.,

$$P(W_i|X) = P(X|W_i) \cdot P(W_i) / \sum_j \{P(X|W_j) \cdot P(W_j)\} \quad (1)$$

$P(W_i|X)$  = the (*a posteriori*) probability that class  $W_i$  occurs, given the observation vector  $X$ ;

$P(X|W_i)$  = the probability that observation  $X$  will occur, given class  $W_i$ ; in image classification: the spectral characteristics of class  $W_i$ .  $P(X|W_i)$  is considered to be constant over the whole scene; and

$P(W_i)$  = the *a priori* probability for class  $W_i$ , which can be considered as a weight factor.

The values of all  $P(X|W_i)$  are assessed using a set of training elements with a known class label. Ideally,  $P(X|W_i)$  is based on relative frequencies of co-occurrence between  $X$  and  $W_i$ . In many expert systems, the total uncertainty in a hypothesis depending on several observations is determined by means of a sequential accumulation of the separate pieces of evidence (e.g., Duda *et al.* (1976), Shafer (1976), and Shortliffe and Buchanan (1986)). For spectral image classification,  $P(X|W_i)$  is determined for the combination of (spectral) observations because in most instances the features are not completely uncorrelated (which is assumed by the current methods for accumulating evidence) and relatively large sample sets are available. When the number of dimensions increases and not enough training samples can be collected,  $P(X|W_i)$  is estimated by an  $n$ -dimensional Gaussian probability density function (Mather, 1987). This function is described by a mean vector and covariance matrix of a representative sample.

Bayes' rule is then rewritten as a decision function  $g_i(X)$  which selects  $W_i$  for which  $g_i(X)$  is the largest; that is,

$$g_i(X) = (2\pi)^{-N/2} \cdot |C_i|^{-0.5} \cdot \exp[-0.5 \cdot (X-M_i)^t \cdot C_i^{-1} \cdot (X-M_i)] \cdot P(W_i) \quad (2)$$

where

$X$  is the observation vector for the object under consideration,  
 $M_i$  is the mean vector for class  $i$ ,

$(X-M_i)^t$  is the transpose of the difference vector between  $X$  and  $M_i$ ,

$C_i$  is the  $N \times N$  symmetric variance-covariance matrix for class  $i$ , and

$|C_i|$  is the determinant of matrix  $C_i$ .

#### KNOWLEDGE-BASED CLASSIFICATION

In practice, the spectral overlap between different classes may be so large that the overall accuracy (expressed as the percentage of correctly classified elements) of a classification result is still unacceptably low. Ancillary data are then used in addition to the spectral information. Such data can be a digital elevation model (elevation, slope angle, and exposition), or thematic maps such as soils and geologic maps. Examples are provided by Kenk *et al.* (1988), Wu *et al.* (1988), and Skidmore (1989).

Ancillary data can be applied effectively only if they have a known relationship to the classes in the image. This means that the classification is extended with a further decision making process, based on the knowledge of the relationship class-ancillary data.

The knowledge can be represented in several ways. Examples include IF <> THEN <> rules, implemented as Boolean look-up tables (e.g., Mulder, 1985) and frames (Wu *et al.*, 1989). Attempts at implementation of uncertain knowledge include certainty factors (Desachy *et al.*, 1988), the Dempster-Shafer theory (Shrinivasan and Richards, 1990), and even a neural network (Hepner *et al.*, 1990).

Kenk *et al.* (1988) used prior probabilities in a vegetation classification. The ancillary data used were terrain elevation, slope, and aspect. The relation between vegetation type and ancillary data was represented by prior probabilities that were assessed by wildlife habitat biologists. This resulted in a modest increase of the classification accuracy (5 percent at most). Strahler (1980) also used elevation data in combination with prior probabilities to improve forest classification. Relationships between forest type, elevation, and aspect were determined using aerial photographs. The classification accuracy of a Landsat MSS image increased 13 percent.

A detailed overview on methodologies of pattern recognition, image analysis, and knowledge based classification is provided by Argialas and Harlow (1990).

To date, only a few studies have been concerned with temporal relationships between classes and ancillary data. Time sequential classifications were mentioned by Simonett *et al.* (1967), Swain (1978), and Strahler (1980). In these classifications, ancillary data were used to determine prior probabilities and these articles refer mainly to multi-season classifications. Van der Laan (1988) used a topographic map from 1980 to improve the classification result of a remotely sensed image acquired in 1986. A look-up table was constructed which defined the possible and impossible land-cover changes between these dates. The table was used to modify the result of the spectral classification.

#### AIM OF THE STUDY

The aim of this study was to develop a method for using land-cover data from preceding years, which are stored in a GIS, to improve the (overall) accuracy of land-cover classifications of a remotely sensed image. The application of temporal thematic information and knowledge meant that the following points had to be investigated:

- Acquisition of knowledge about the temporal relationships between classes from the available data and experts.
- Representation of the knowledge.
- Implementation of the information and knowledge in the image classification mechanism.

This is described for a test area in The Netherlands where there are clear temporal relationships between classes in the form of crop rotation schemes.

#### TEST AREA AND DATA

##### TEST AREA

The test area is located in Oostelijk Flevoland, one of the polders in The Netherlands. A large part of this polder (a tract of low land reclaimed from the IJssel Lake), including the test area, is used for arable agriculture. Temporal relationships between classes exist here in the form of distinct crop rotation schemes. The main crops are grass (GR), cereals (CE), potatoes (PO), sugarbeets (SB), beans (BE), peas (PE), and onions (ON).

##### SATELLITE DATA

A Landsat Thematic Mapper (TM) image of the Flevoland area (acquisition date: 7 July 1987) of good quality was available. Bands 3 (0.63 to 0.69  $\mu\text{m}$ ), 4 (0.76 to 0.90  $\mu\text{m}$ ), and 5 (1.55 to 1.75  $\mu\text{m}$ ) from this image were used.

##### GEOGRAPHIC DATA

The ground reference data originated from a local agricultural institute (Rentambt Oostelijk Flevoland). Until 1988, this institute systematically gathered data about the crops grown in the area. Each year, small maps were sent to farmers who indicated the position and acreage of their crops. The maps were transposed to a 1:10,000-scale topographic map and were subsequently digitized for the situation in 1985, 1986, and 1987, and finally stored in a geographic information system. The digitized area comprises approximately 3,800 ha.

#### KNOWLEDGE ACQUISITION

Before an appropriate formalism for knowledge representation could be defined, it was necessary to know at least qualitatively which crops are grown in the area and what kinds of temporal relationships (rotation schemes) can be found. This information was acquired from different information sources, as listed below.

##### LITERATURE

From literature, it was known which crops were grown in the Flevoland area (RIJP, 1988). In this area, mainly three- and four-year rotation cycles are used. A crop succession scheme indicating suitabilities of crop successions and risk of disease and pests was also available (PAGV, 1989).

##### FOUR-YEAR SERIES OF FIELDS

Ground reference data for a series of four successive years were available for 60 fields. It was found that

- There was a four year cycle (Potatoes - Cereals - SugarBeets - diverse - Potatoes) in this part of the test area.
- Field boundaries may change in time; the temporal relationships are therefore defined per raster cell.

It might be possible to determine empirically prevalent crop transitions or successions by counting transitions over a period of  $N$  successive years for one field. To find prevalent transitions that are significantly different from random transitions, however, a time series of at least 20 to 30 years must be available. These were not available for the test area. Furthermore, it is likely that within such a long period the crop rotation schemes would change.

CROSSING GROUND-COVER MAPS FROM SUCCESSIVE YEARS

Because counting transitions over a long period was not possible, we enlarged our sample by including a large number of land parcels. Because each phase of the cycle was equally presented in the area, we overlaid ("crossed") two rasterized ground-cover maps of two successive years. The transition frequencies from class  $W_j$  at  $t-1$  to  $W_i$  at  $t$  were counted and stored in a matrix. The classes of  $t-1$  was put as the row-index and  $t$  as the column index, and the values are normalized over  $t-1$ . The matrix indicates the prevalent transitions in terms of percentage of class  $W_j$  at  $t-1$  that becomes class  $W_i$  at  $t$ .

INTERVIEWING EXPERTS

In addition to the information from literature and the ground reference data, agricultural experts from a consulting agency (Consulentschap Akker- en Tuinbouw, Lelystad) were interviewed. From these interviews, we learned which crops were relevant for inclusion in the classification. These are shown in Figure 1. At least five different crop rotation schemes are used in this area. The most dominant are a three-year cycle and two four-year cycles. These are illustrated in Figure 2 as 3yr, 4yr1, and 4yr2. Quantitative descriptions of these rotations were obtained and also physical and economic backgrounds of the rotation schemes were explained by the experts. These, however, could not be included in the classification.

REPRESENTATION OF TEMPORAL RELATIONSHIPS

A land-cover class  $W_i$  for a specific object (e.g., pixel, parcel) at a specific time can be defined as the state of that object. The change from a certain class  $W_j$  at  $t-1$  to another class  $W_i$  at  $t$  can be interpreted as a state transition. The transitions from the state  $W_{j,t-1}$  to the state  $W_{i,t}$  can be described by a transition matrix.

The time lapse between  $t$  and  $t-1$  may be a season, a year, or even a five year period. The only condition is that both the land cover data at  $t-1$  and the transition matrix are based on the same time lapse. In this study, yearly land-cover changes were considered.

An example is a three-year crop rotation in which PO (potatoes), SB (sugar beets), and CE (cereals) are grown sequentially. The chain PO-SB-CE-PO-SB-CE-... is called a Markov chain (e.g., Fletcher (1972)) which can be represented by a state transition graph (Figure 3). This graph can be converted to the transition matrix  $M$  (Figure 4) which contains the same information as the graph. The rows in the matrix are transition vectors, indicating the transitions from the state  $W_{j,t-1}$  to state  $W_{i,t}$ .

In the three-year rotation which is used in the study area, not one but several different states may follow a preceding state, as illustrated in Figure 2. It means that it is not certain what the land-cover  $W_i$  at  $t$  will be given the landcover  $W_j$  at  $t-1$ . The three-year crop rotation shown in Figure 2 therefore can be transformed into a probabilistic state transition graph (Figure 5). The widths of the arrows in the graph are proportional to the relative areas of  $W_j$  at  $t-1$  that will be covered by  $W_i$  at  $t$ . Thus, the arrows indicate the probability that a state  $W_j$  at  $t-1$  will be followed by  $W_i$  at  $t$ .

Matrix  $F$  shown in Figure 6 corresponds to the probabilistic transition graph in Figure 5.

In a stochastic matrix, the values of all entries are equal to zero or larger and the sum of the entries in each row equals 1. Matrices  $M$  and  $F$  are examples of a stochastic matrix. The rows in these matrices are called "state transition probability vectors"  $P(W_{i,t}|W_{j,t-1})$ , representing the probabilities that a transition from state  $W_{j,t-1}$  to state  $W_{i,t}$  will occur.

The states in matrix  $F$  are described by both the land-cover class label and the year index, because the same land cover may be found in more than one year of the rotation cycle. In the








| crop type   | icon  |
|-------------|---|
| potatoes    |  |
| cereals     |  |
| sugar beets |  |
| grass       |  |
| beans       |  |
| peas        |  |
| onions      |  |

FIG. 1. Crop types included in the classification.

example, "beans" are grown in both the second and the third years of the cycle. Thus, "beans" in the second year and "beans" in the third year of the cycle have to be interpreted as different states.

The year-indexed matrix demands that the input to this matrix is defined also by both  $W_j$  and a year index. This means that, for this matrix, it is not sufficient to have a ground-cover map indicating  $W_{j,t-1}$ , but this map must also indicate the year of the cycle (1, 2, or 3) to which the crops refer. Because it may happen that these year-indices are not available, as in this study, the year indices were removed from matrix  $F$ , resulting in matrix  $G$  (Figure 7). The latter shows the state transition probabilities without regard to the year of the cycle. In the example, matrix  $N2$  does not distinguish "beans" in the second year from "beans" in the third year. A little information was therefore lost in this matrix.

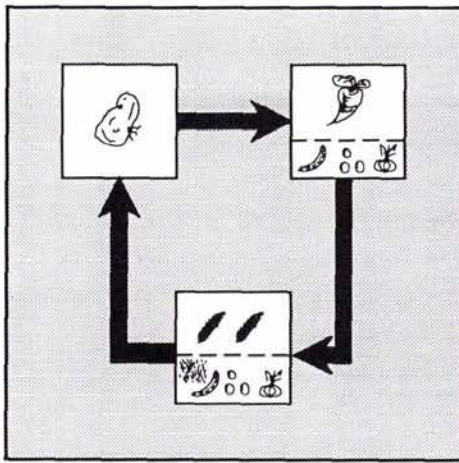
The terms in the  $n$ th power of a transition matrix give the probabilities of a transition taking place in  $n$  time lapses.  $M^2$ ,  $M^3$ , and  $M^4$  are the second, third, and fourth powers of matrix  $M$ , indicating the transitions over two, three, and four years (Figure 8). If "potatoes" was found at a certain time, it can be concluded from  $M^2$  that "cereals" will be found at the same spot two years later. The fourth power of  $M$  is the same as  $M$ , so there is a clear pattern that takes a period of three years.

A transition matrix is said to be regular if all entries of some power are positive. As can be seen from the example, the matrix  $M$  is not regular. Also  $F$  is not regular, but matrix  $G$  is a regular stochastic matrix. An interesting property of a regular stochastic matrix is the following: the sequence of powers  $G^2$ ,  $G^3$ ,  $G^4$ , ...,  $G^n$  ( $n \rightarrow \infty$ ) approaches a matrix whose rows are each the same probability vector  $V$ .

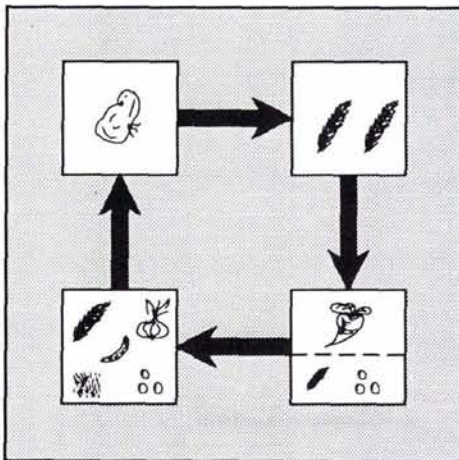
The 100th power of  $G$  gives the matrix shown in Figure 9. The probability vector  $V = (0.33, 0.30, 0.16, 0.03, 0.04, 0.04, 0.10)$  is one of the eigenvectors for this matrix. The eigenvalue corresponding to this eigenvector = 1.0.

In a stable dynamic system such as the rotation system in the example, the relative areas covered by each class do not change. So these relative areas are thus represented by the eigenvector  $V$ . Moreover, the total size of the area is always the same. The eigenvalue of eigenvector  $V$  therefore equals 1.0.

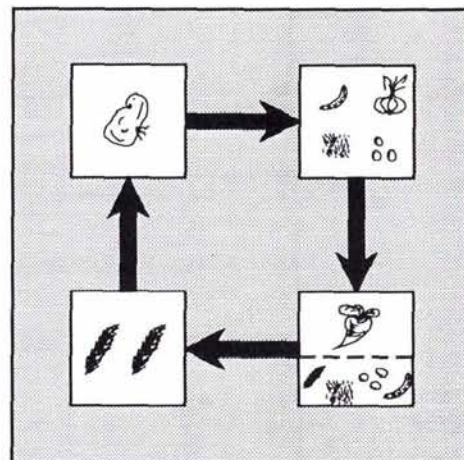
In terms of land-cover changes,  $G^{100}$  gives the probability vectors for a time lapse of 100 years. The matrix can be interpreted as follows: it means that the probability vectors for all



Three year crop rotation (3yr).



Four year crop rotation type 1 (4yr1).



Four year crop rotation type 2 (4yr2).

FIG. 2. Qualitative representation of the most dominant rotation schemes that are used in the study area.

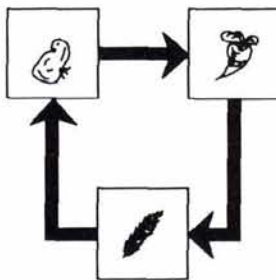


FIG. 3. Schematic representation of a three-year crop rotation cycle.

|                        |  | $W_{i,t} \rightarrow$ |   |   |
|------------------------|--|-----------------------|---|---|
|                        |  |                       |   |   |
| $W_{j,t-1} \downarrow$ |  | 0                     | 1 | 0 |
|                        |  | 0                     | 0 | 1 |
|                        |  | 1                     | 0 | 0 |

FIG. 4. Transition matrix M.

TRANSITION MATRICES

DETERMINATION OF TRANSITION MATRICES

The values in the transition matrices were assessed on the basis of both empirical estimations using the ground reference data from previous years, and expert knowledge.

*Empirical estimation.* Because the crop rotation system(s) in the

$W_k$  at  $t-100$  are the same; so, after too many rotation cycles, a ground cover map (like  $W_{k,t-100}$ ) is no longer relevant as ancillary information. These limiting probability vectors actually represent the prior probabilities that are based on the relative areas of the land-cover types.

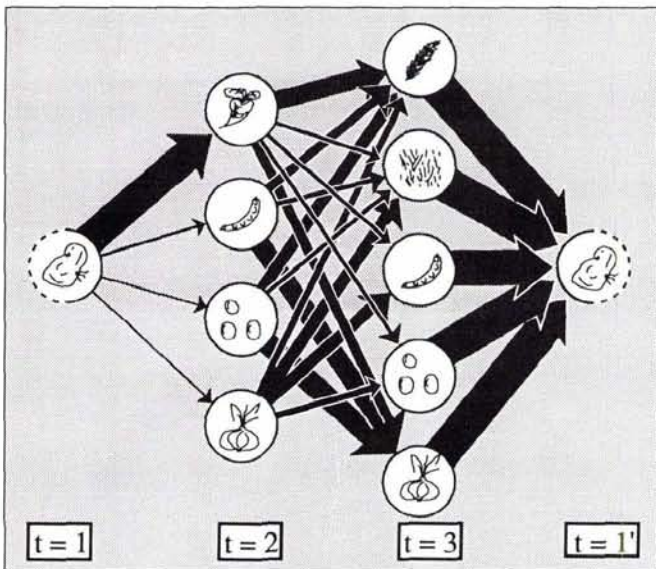


FIG. 5. Probabilistic state transition graph of three-year crop rotation cycle. The width of the arcs is proportional to the transition probabilities.

area are stable, the transition probability vectors between classes were assessed by overlaying rasterized ground-cover maps of two successive years. The frequencies of co-occurrences between  $W_{j,t-1}$  and  $W_{i,t}$  are counted and normalized over  $W_{j,t-1}$ . The resulting matrices are called empirical matrices (MAR8586 and MAR8687). Matrix MAR8586 is shown in Figure 10.

*Expert knowledge.* Because different rotation cycles are used in this area, the empirical matrices reflect a mixture of rotation cycles. The agricultural consultants were therefore asked to estimate the transition vectors for each of the three dominant rotation schemes separately. For each type a year-indexed matrix (Figure 6) was filled in. Because the classification only uses a ground reference map for  $t-1$ , the year indices from these matrices were eliminated and a one-step transition matrix was created.

As a test for consistency of the different information sources, an ensemble matrix was composed on the basis of the three separate dominant matrices. This is done by a weighted addition of the three, where the weight is the relative area of occurrence of each rotation type. The resulting matrix (MEXPERT) is shown in Figure 11.

*Empirical estimation of a single rotation matrix.* Because the empirical matrices MAR8586 and MAR8687 reflect more than rotation cycle, an empirical matrix for one single rotation type was created. On the basis of a land-ownership map and the ground-cover maps of the three successive years (1985 to 1987), the area where the 4yr1 rotation is used could be separated and masked. Within this mask an empirical matrix of the 4yr1 rotation was made (M4YR1).

INTERPRETATION OF THE MATRICES

Comparison of the empirical matrices MAR8586 and MAR8687 revealed only minor differences. These were found for classes which do not occur very frequently: grass, beans, peas.

The 3yr and 4yr1 rotations are clearly reflected by the highest transition probabilities in the empirical matrices MAR8586 and MAR8687. This illustrates that the latter two reflect an ensemble of different rotation matrices.

$(W_{i, Year_j})_t \rightarrow$

|              | <sub>1</sub> | <sub>2</sub> | <sub>2</sub> | <sub>2</sub> | <sub>2</sub> | <sub>3</sub> | <sub>3</sub> | <sub>3</sub> | <sub>3</sub> | <sub>3</sub> |
|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| <sub>1</sub> |              | 0.80         | 0.07         | 0.07         | 0.06         |              |              |              |              |              |
| <sub>2</sub> |              |              |              |              |              | 0.50         | 0.08         | 0.08         | 0.08         | 0.26         |
| <sub>2</sub> |              |              |              |              |              | 0.30         | 0.10         |              |              | 0.60         |
| <sub>2</sub> |              |              |              |              |              | 0.33         | 0.12         |              |              | 0.55         |
| <sub>2</sub> |              |              |              |              |              | 0.17         | 0.33         | 0.33         | 0.17         |              |
| <sub>3</sub> | 1.0          |              |              |              |              |              |              |              |              |              |
| <sub>3</sub> | 1.0          |              |              |              |              |              |              |              |              |              |
| <sub>3</sub> | 1.0          |              |              |              |              |              |              |              |              |              |
| <sub>3</sub> | 1.0          |              |              |              |              |              |              |              |              |              |
| <sub>3</sub> | 1.0          |              |              |              |              |              |              |              |              |              |

$\leftarrow (W_{k, Year_l})_{t-1}$

FIG. 6. Transition matrix F for three-year rotation cycle; in this matrix, a state is defined by both crop type and year.

$W_{i,t} \rightarrow$















|                        |   |   |   |   |   |   |   |
|------------------------|---|---|---|---|---|---|---|
|                        |  |  |  |  |  |  |  |
| $W_{j,t-1} \leftarrow$ |  | 0.89  |   |   | 0.05  | 0.03  | 0.03  |
|                        |  |   | 0.50  | 0.08  | 0.08  | 0.08  | 0.26  |
|                        |  | 0.96  |   | 0.01  | 0.01  | 0.01  | 0.01  |
|                        |  | 1.0   |   |   |   |   |   |
|                        |  | 0.60  |   | 0.12  | 0.04  |   | 0.24  |
|                        |  | 0.82  |   | 0.06  | 0.02  |   | 0.10  |
|                        |  | 0.94  |   | 0.01  | 0.02  | 0.02  | 0.01  |

FIG. 7. Transition matrix **G** representing the three-year rotation cycle. The year-index has been removed.

The MEXPERT matrix is slightly different from the empirical matrices, but the correspondence is satisfactory. The differences can be explained by the following factors:

- The expert matrix represents only the three dominant rotations.
- The expert expresses the situation in 1989, but the matrices are related to 1986 and 1987.
- Small errors occur in the digital ground reference maps.
- The expert did not predict the transition probabilities for these crops very accurately.

It can be easily verified that both empirical matrices (MAR8586 and MAR8687) and the expert matrix are regular. Thus, calculation of the *n*th power of each matrix provides an eigenvector which indicates the relative areas occupied by each ground cover class. As shown in Table 1, they agree well with the relative areas given by agricultural statistics and with the areas indicated on the ground reference maps.

EVALUATION OF KNOWLEDGE SOURCES

Both procedures for assessing of the class transition probabilities have their costs and benefits in terms of effort of acquisition and expected accuracy of the matrices. The most important of them are mentioned below.

The advantages of empirical estimation were

- The procedure is fast and easy when digital data are available;
- The transition probabilities are area-specific, unlike literature references which usually contain more general information; and

- The method is objective, not biased by the expert's opinion or wishes.

The disadvantages of the empirical estimation were

- The ground reference data must be available in detail (in this case, more thematic information or classes than shown on topographic maps have been used);
- The method is laborious if data have to be digitized first;
- Errors in the ground-cover maps may cause incorrect and unlikely transition probabilities; and
- An empirical matrix may reflect an ensemble of different rotations, and it is difficult to separate these without other information.

The advantages of estimation by means of an expert were

- No ground reference data are required for the determination of a transition matrix;
- Acquisition of knowledge by interviewing experts is faster than digitizing detailed ground-cover maps;
- There will be fewer errors caused by errors in the GIS (only for ground cover at *t*-1); and
- An expert can foresee minor yearly fluctuations and trends, so that these can be included in the transition probabilities.

The disadvantages of estimation by means of an expert were

- The expert's knowledge can be different from the real situation. In this study the expert expected some transitions other than the farmers actually used; and
- The expert must be able to express his/her knowledge in terms of transition probabilities.

IMPLEMENTATION IN THE MAXIMUM-LIKELIHOOD CLASSIFICATION

As described above, Bayes' rule provides an appropriate method to assess the most likely class, given an observation vector and the class *a priori* probabilities. In this case study, the *a priori* probability for a class could be specified more accurately than just using the relative occurrence of the class in the whole area. Where there are temporal relationships between the classes, the prior probabilities depend on the land cover at the preceding year.  $P(W_i)$  in Equations 1 and 2 was therefore substituted by  $P(W_{i,t}|W_{j,t-1})$ , which is read from a transition matrix.  $P(W_{i,t}|W_{j,t-1})$  represents the relative area occupied by  $W_i$  within the masked area where  $W$  at *t*-1 =  $W_j$ . This means that  $P(W_{i,t})$  is dependent on the spatial distribution of  $W_j$  at *t*-1, so it may vary per pixel.

The complete classification procedure is shown schematically in Figure 12.

- (1) The mean vectors and covariance matrices for all classes were determined using training samples from fields where the current land cover was known.
- (2) From the mean vectors and covariance matrices,  $P(X|W_{i,t})$  was calculated for each class.
- (3) On the basis of two overlaid ground cover maps of successive years (A) or interviews with experts (B), a transition matrix containing probability vectors  $P(W_{i,t}|W_{j,t-1})$  was created.
- (4) For each pixel the land-cover class  $W_j$  at *t*-1 was looked-up in the GIS. This class pointed to a probability vector in the transition matrix.

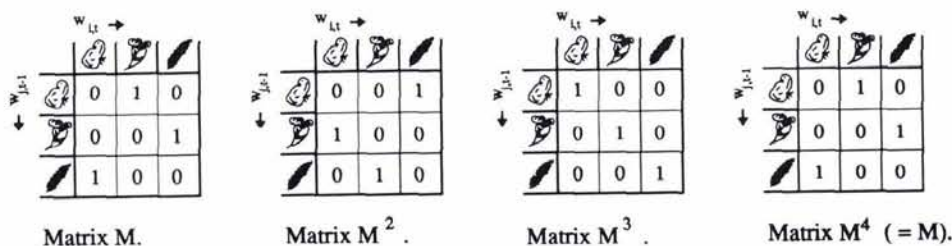


FIG. 8. Sequence of powers of matrix **M**.

|                        |  |                       |      |      |      |      |      |      |
|------------------------|--|-----------------------|------|------|------|------|------|------|
|                        |  | $w_{i,t} \rightarrow$ |      |      |      |      |      |      |
|                        |  |                       |      |      |      |      |      |      |
| $w_{j,t-1} \downarrow$ |  | 0.33                  | 0.30 | 0.16 | 0.03 | 0.04 | 0.04 | 0.10 |
|                        |  | 0.33                  | 0.30 | 0.16 | 0.03 | 0.04 | 0.04 | 0.10 |
|                        |  | 0.33                  | 0.30 | 0.16 | 0.03 | 0.04 | 0.04 | 0.10 |
|                        |  | 0.33                  | 0.30 | 0.16 | 0.03 | 0.04 | 0.04 | 0.10 |
|                        |  | 0.33                  | 0.30 | 0.16 | 0.03 | 0.04 | 0.04 | 0.10 |
|                        |  | 0.33                  | 0.30 | 0.16 | 0.03 | 0.04 | 0.04 | 0.10 |
|                        |  | 0.33                  | 0.30 | 0.16 | 0.03 | 0.04 | 0.04 | 0.10 |

FIG. 9. 100th power of regular matrix **G**. Each row contains the same probability vector.

|                        |  |                       |      |      |      |      |      |      |
|------------------------|--|-----------------------|------|------|------|------|------|------|
|                        |  | $w_{i,t} \rightarrow$ |      |      |      |      |      |      |
|                        |  |                       |      |      |      |      |      |      |
| $w_{j,t-1} \downarrow$ |  |                       | 0.56 | 0.18 | 0.05 | 0.08 | 0.08 | 0.05 |
|                        |  | 0.42                  | 0.04 | 0.50 | 0.01 | 0.01 | 0.01 | 0.01 |
|                        |  |                       | 0.61 |      | 0.02 | 0.02 | 0.30 | 0.05 |
|                        |  | 0.70                  | 0.08 | 0.20 |      | 0.01 |      | 0.01 |
|                        |  | 0.68                  | 0.05 | 0.21 | 0.01 |      |      | 0.05 |
|                        |  | 0.67                  | 0.05 | 0.21 | 0.02 |      |      | 0.05 |
|                        |  | 0.74                  | 0.01 | 0.22 | 0.01 | 0.01 | 0.01 |      |

FIG. 11. Transition matrix **MEXPERT**, based on expert knowledge.

|                        |  |                       |      |      |      |      |      |      |
|------------------------|--|-----------------------|------|------|------|------|------|------|
|                        |  | $w_{i,t} \rightarrow$ |      |      |      |      |      |      |
|                        |  |                       |      |      |      |      |      |      |
| $w_{j,t-1} \downarrow$ |  | 0.02                  | 0.69 | 0.15 | 0.06 | 0.01 | 0.05 | 0.02 |
|                        |  | 0.27                  | 0.03 | 0.67 |      | 0.01 | 0.01 | 0.01 |
|                        |  | 0.10                  | 0.28 | 0.02 |      | 0.04 | 0.27 | 0.29 |
|                        |  | 0.31                  | 0.04 | 0.33 | 0.17 |      | 0.02 | 0.13 |
|                        |  | 0.90                  | 0.06 | 0.04 |      |      |      |      |
|                        |  | 0.94                  | 0.01 | 0.01 |      |      |      | 0.04 |
|                        |  | 0.87                  | 0.04 | 0.08 | 0.01 |      |      |      |

FIG. 10. Empirical transition matrix **MAR8586**.

TABLE 1. RELATIVE AREAS OF CROP TYPES (RIJP, 1988) AND EIGEN VECTORS OF TRANSITION MATRICES.

|             | relative area occupied in |      |      | eigen vector after 100th power |         |         |
|-------------|---------------------------|------|------|--------------------------------|---------|---------|
|             | 1985                      | 1986 | 1987 | MAR8586                        | MAR8687 | MEXPERT |
| potatoes    | 0.28                      | 0.28 | 0.26 | 0.28                           | 0.25    | 0.26    |
| sugar beets | 0.25                      | 0.26 | 0.27 | 0.25                           | 0.26    | 0.24    |
| cereals     | 0.31                      | 0.27 | 0.27 | 0.27                           | 0.27    | 0.31    |
| grass       | 0.01                      | 0.02 | 0.02 | 0.02                           | 0.02    | 0.02    |
| beans       | 0.03                      | 0.01 | 0.01 | 0.02                           | 0.02    | 0.03    |
| peas        | 0.05                      | 0.08 | 0.08 | 0.08                           | 0.09    | 0.1     |
| onions      | 0.08                      | 0.08 | 0.09 | 0.09                           | 0.1     | 0.04    |

the Landsat TM image) and a situation with poor spectral class discrimination (using band 4 only). Also, different transition matrices were evaluated.

For every class, three to seven different fields (125 to 460 pixels) were randomly selected to determine the class mean vector and covariance matrix. The overall accuracies of the classification results were assessed by calculating a cross tabulation of the rasterized ground reference map and the classification result. They are shown in Table 2.

MAXIMUM-LIKELIHOOD CLASSIFICATION USING BANDS 3, 4, AND 5

When only spectral information was used (all prior probabilities were equal), the overall accuracy of the classification was 76 percent.

The improvement using class prior probabilities was very little: only 1 percent. This is striking, because the prior probabilities range from 0.01 to 0.30. Apparently, only a few pixels had values that fall into the overlap zones between clusters in the feature space.

The improvement using class priors based on the empirical matrices **MAR8586** and **MAR8687** was larger: 4 percent. Both empirical matrices provided about the same result, but the improvement using the pixel priors from the matrix **MEXPERT** was less: 1.8 percent.

The best result was obtained with matrix **M4YR1** for the area with only the 4yr1 rotation: the overall accuracy was 81.9 percent.

- (5) The probability vector corresponding to  $w_{j,t-1}$  was read from the matrix.
- (6) The computer program **BAYES** determined on the basis of  $P(X|w_{i,t})$  (from step 2) and  $P(w_{i,t}|w_{j,t-1})$  (from step 5) for each pixel which was the most likely class.
- (7) The result was stored in the GIS.

TESTING

The performance of the knowledge based classification was tested by comparison with "standard" classification results using equal prior probabilities and class prior probabilities (using the relative frequency of the classes in the whole area).

The classifications were carried out for a situation with a good spectral separation between the classes (using three bands of

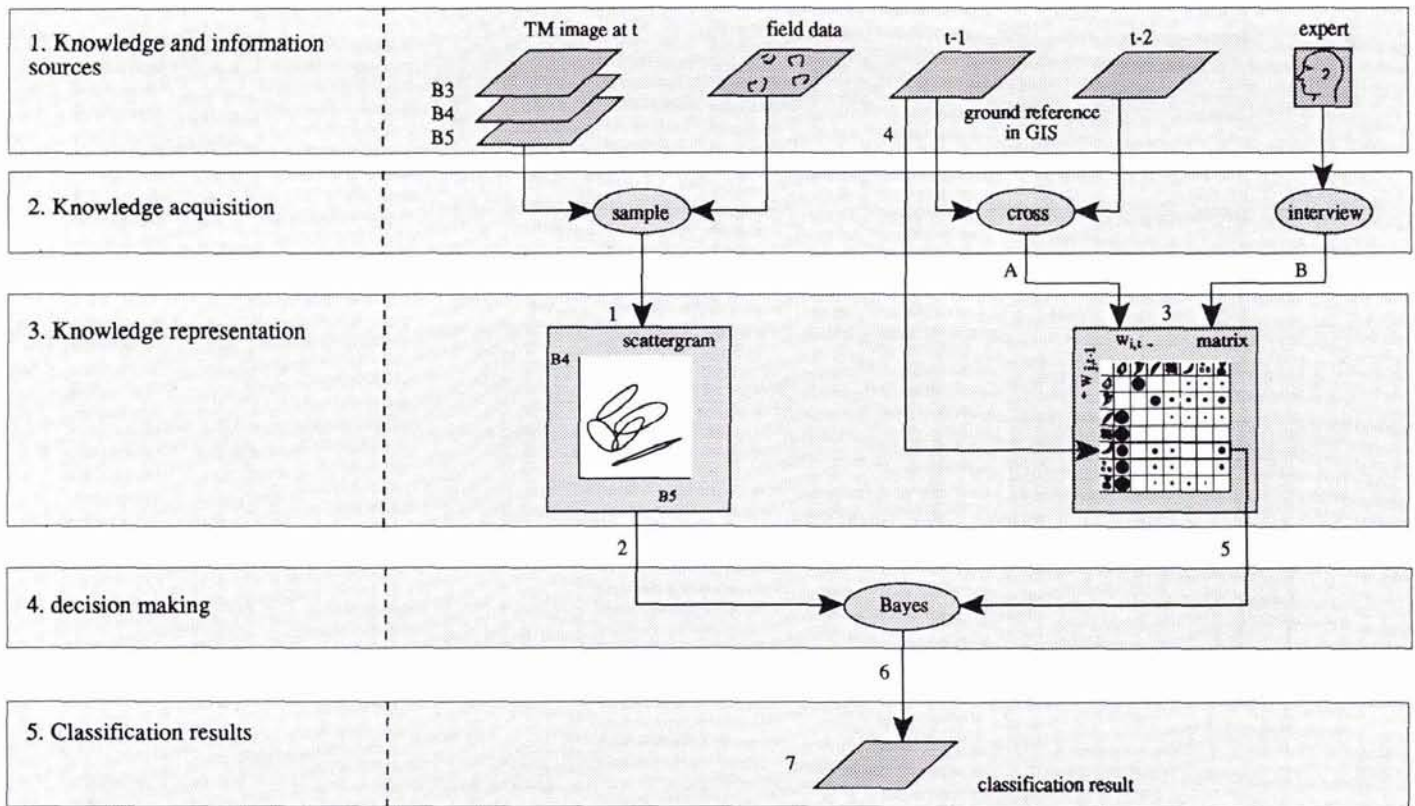


FIG. 12. Scheme of the classification procedure.

But even with these most area-specific priors, the improvement is not more than 5.9 percent.

From this it can be concluded that, if there is a good spectral separation between the classes, little is gained by incorporating additional data and knowledge.

**MAXIMUM-LIKELIHOOD CLASSIFICATION USING ONLY BAND 4**

The second series of tests was carried out in a situation with a poor spectral separation between the classes. This was simulated by classifying using only band 4.

The overall accuracy using equal prior probabilities decreased to 61.4 percent. Also the other classifications provided a lower overall accuracy than when using three bands, but the effect of the prior probabilities was much larger here.

The application of class priors gave a 3.6 percent better result, which was still rather small.

The classifications based on the pixel priors, however, showed an important improvement of the overall accuracy. Using the empirical matrices for the whole area (MAR8586 and MAR8687), a 12.9 percent better result was obtained, and matrix MEXPERT yielded a 12.6 percent better result.

The most interesting result was based on the matrix M4YR1 for this area. The overall accuracy here was still more than 80 percent, which is 19.5 percent better than using equal priors, and 14.1 percent better than using class priors. In fact, this result was only 1 percent worse than the same classification using three bands. Because the spectral information from band 3 and band 5 added only 1 percent to the accuracy, this additional spectral information can be seen as redundant.

This classification was not only good just because the ground cover was usually as predicted by the matrix. A classification on the basis of the transition probabilities alone yielded an overall accuracy of only 60 percent. It was the combination of spectral

TABLE 2. CLASSIFICATION RESULTS

|  | prior probability       | TM B3,4,5 | TM B4         |
|--|-------------------------|-----------|---------------|
|  | equal for all classes   | 76.0      | 61.4          |
|  | relative area per class | 77.0      | 65.0          |
|  | MAR8586                 | 79.6      | 74.3          |
|  | MAR8687                 | 79.9      | not evaluated |
|  | MEXPERT                 | 77.8      | 73.6          |
|  | M4YR1                   | 81.9      | 80.9          |

information with proper predictions which led to a high overall accuracy of the classification result.

This means that, where there is an important spectral overlap between classes, this application of pixel priors may be worth the effort, especially when a specific matrix, not reflecting an ensemble of rotations, can be found.

**CONCLUSIONS**

Markov chains and probabilistic transition matrices form a useful extension of the knowledge representation methods currently used for knowledge-based image classification. They provide a proper representation formalism for temporal relationships. These temporal relationships can be implemented as context ( $C_{t-1}$ ) dependent sets of prior probabilities. In this study we used crop rotations to develop the method. Other applications, however, of transition probabilities could be thought of, for example, developments of (natural) vegetation, forest management. For the assessment of the transition matrices, a lot of data are needed, but an expert may be available to provide the transition probabilities. The improvements which can be expected from this application of temporal relationships depend on two factors:



- (1) Spectral separation of classes. When there is a good spectral separation between the classes, there is little improvement from adding ancillary data and knowledge. On the other hand, a considerable increase of the overall accuracy can be obtained if there is a large spectral overlap between the classes. These situations occur when images with a low spectral resolution (such as SPOT data) are used. SPOT data lack a middle-infrared spectral band which is very important for land-cover discrimination. Images outside the optimal (growing) season are also possible candidates for this approach.
- (2) The best results can be expected when a transition matrix can be found which reflects only one distinct Markov chain. Such a matrix will contain probability vectors containing values close to either 0.0 or 1.0.

The necessary knowledge about temporal relationships can be obtained both from analysis of a time series of ground reference data or from a human expert. The first requires a lot of data in digital format, which may be (time) expensive. Underlying processes and the existence of different transition chains are difficult to discover from these data. The expert's knowledge can be acquired easily and quickly, and background information will be available, but the expert should be able to provide reliable and accurate information.

#### REFERENCES

- Argialas, D. P., and C. A. Harlow, 1990. Computational Image Interpretation Models: An Overview and a Perspective. *Photogrammetric Engineering & Remote Sensing*, Vol. 56, No. 6, pp. 871-886.
- Desachy, J., P. Debord, and S. Castan, 1988. An expert system for satellite image interpretation and GIS based on problem solving. *ISPRS proceedings Kyoto*, Comm. 4/5, Vol 27, B4.
- Duda, R. O., P. E. Hart, and N. J. Nilsson, 1976. *Subjective Bayesian Methods for Rule-Based Inference Systems*. Artificial Intelligence Center, Tech. Note 124, SRI International, Menlo Park, California.
- Fletcher, T. J., 1972. *Linear Algebra through its Applications*. Van Nostrand Reinhold Company, London.
- Hepner, G. F., T. Logan, N. Ritter, and N. Bryant, 1990. Artificial Neural Network Classification Using a Minimal Training Set: Comparison to Conventional Supervised Classification. *Photogrammetric Engineering & Remote Sensing*, Vol. 56, No. 4, pp. 469-473.
- Kenk, E., M. Sondheim, and B. Yee, 1988. Methods for improving accuracy of thematic mapper ground-cover classifications. *Can. Jnl. of RS*, Vol. 14, No. 1.
- Laan, van der, F. B., 1988. Improvement of classification results of satellite imagery using context information from a topographic map. *Proc. ISPRS Symp. Comm. VII, Kyoto*, pp. 98-108.
- Mather, P. M., 1987. *Computer Processing of Remotely-Sensed Images. An Introduction*. John Wiley & Sons, Chichester.
- Middelkoop, H., J. W. Miltenburg, and N. J. Mulder, 1989. Knowledge engineering for image interpretation and classification: a trial run. *ITC Journal*, 1989-1, pp. 27-32.
- Mulder, N. J., 1985. Classification and decision-making. *Photogrammetria* 40, pp. 95-116.
- PAGV, 1989. *Bouwplan en vruchtopbevolging*. PAGV publicatie nr. 44, Lelystad.
- RIJP, 1988. *Overzicht van de gewassen keuze op de landbouwbedrijven in Oostelijk Flevoland, tijdreeks 1965-1985*. RIJP rapport 1988 - 30 cbw, Lelystad.
- Shafer, G., 1976. *A Mathematical Theory of Evidence*. Princeton Univ Press, Princeton, N.J.
- Shortliffe, E. H., and B. G. Buchanan, 1986. A model of inexact reasoning in predictive expert systems. *Uncertainty in Artificial Intelligence*. (L. N. Kanal and J. F. Lemmer, eds.): North-Holland Publ, Amsterdam, pp. 43-52.
- Shrinivasan, A., and J. A. Richards, 1990. Knowledge-based techniques for multi-source classification. *International Journal of Remote Sensing*, Vol. 11, No. 3, pp. 505-525.
- Simonett, D. S., J. E. Eagleman, A. B. Erhart, D. C. Rhodes, and D. E. Schwartz, 1967. *The Potential of Radar as a Remote Sensor in Agriculture: 1. A Study with the K-Band Imagery in West-Kansas*. CRES Report No. 16-21, University of Kansas, Lawrence.
- Skidmore, A. K., 1989. An Expert System Classifies Eucalypt Forest Types Using Thematic Mapper Data and a Digital Terrain Model. *Photogrammetric Engineering & Remote Sensing*, Vol. 55, No. 1, pp. 133-146.
- 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.
- Swain, P. H., 1978. Bayesian classification in a time-varying environment. *IEEE transactions on Systems, Man and Cybernetics*, Vol. SMC-8, No. 12, pp. 879-883.
- Wu, J. K., D. S. Cheng, W. T. Wang, and D. L. Cai, 1988. Model Based remotely sensed imagery interpretation. *International Journal of Remote Sensing*, Vol. 9, No. 8, pp. 1347-1356.

(Received 17 September 1990; accepted 18 October 1990; revised 6 November 1990)

## Air, Marine, and Land Radionavigation Systems Users

### 1990 Federal Radionavigation Plan 1991 Conferences

The U.S. Department of Transportation is conducting open meetings for all users of U.S. government-provided radionavigation systems. The purpose of the meetings is to obtain user perspectives on federal policies and future plans for these services. Federal radionavigation policies and plans are outlined in the 1990 DOD/DOT Federal Radionavigation Plan, single copies of which are available from the VOLPE National Transportation Systems Center. Users are encouraged to attend the meetings to provide inputs for the 1992 plan.

LORAN-C • OMEGA • TRANSIT • RADIOBEACONS • VOR/DME • MLS/ILS • GPS

**Sponsors:** Research & Special Programs Administration; Federal Aviation Administration; U.S. Coast Guard

**Dates/Locations:** 19-20 Nov. 1991, Alexandria, Virginia; 5 Dec. 1991, Seattle, Washington.

**Information:** *Federal Radionavigation Plan:* Elisabeth J. Carpenter, Volpe National Transportation Systems Ctr., Ctr. for Navigation (DTS-52), 55 Kendal Square, Cambridge, MA 02142-1093, tel. 617-494-2126.

*Conferences:* Conference Office (DTS-930), Attn: Radionavigation Users Conference, 55 Kendal Square, Cambridge, MA 02142-1093, tel. 617-494-2307.