Multitemporal fuzzy classification model based on class transition possibilities

Guilherme L.A. Mota [a,b], Raul Q. Feitosa [a,b,*], Heitor L.C. Coutinho [c], Claus-Eberhard Liedtke [d], Sönke Müller [d], Kian Pakzad [e], Margareth S.P. Meirelles [b,c]

[a] *Catholic University of Rio de Janeiro, Department of Electrical Engineering, Rua Marquês de São Vicente 225, Room 401 L, Gávea, Rio de Janeiro, RJ 22453-900, PO Box 38097, Brazil*
[b] *Rio de Janeiro State University, Department of Systems and Computing Engineering, Rua São Francisco Xavier 524, Room 5028 D, Maracanã, Rio de Janeiro, RJ 20550-900, Brazil*
[c] *Embrapa, Soils Research Center, Rua Jardim Botânico 1024, Jardim Botânico, Rio de Janeiro, RJ 22460-000, Brazil*
[d] *University of Hannover, Institut für Informationsverarbeitung, 9a Appelstrasse, 13th floor, Hannover 30167, Germany*
[e] *University of Hannover, Institute of Photogrammetry and GeoInformation, 1 Nienburger Street, Hannover D-30167, Germany*

Abstract

This paper proposes a new method to model temporal knowledge and to combine it with spectral and spatial knowledge within an integrated fuzzy automatic image classification framework for land-use land-cover map update applications. The classification model explores not only the object features, but also information about its class at a previous date. The method expresses temporal class dependencies by means of a transition diagram, assigning a possibility value to each class transition. A Genetic Algorithm (GA) carries out the class transition possibilities estimation. Temporal and spectral/spatial classification results are combined by means of fuzzy aggregation. The improvement achieved by the use of multitemporal knowledge rather than a pure monotemporal approach was assessed in a real application using LANDSAT images from Midwest Brazil. The experiments showed that the use of temporal knowledge markedly improved the classification performance, in comparison to a conventional single-time classification. A further observation was that multitemporal knowledge may subsume the knowledge related to steady spatial attributes whose values do not significantly change over time.
© 2007 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

Keywords: Remote sensing; Knowledge-base representation; Multitemporal interpretation; Fuzzy logic

* Corresponding author. Catholic University of Rio de Janeiro, Department of Electrical Engineering, Rua Marquês de São Vicente 225, Room 401 L, Gávea, Rio de Janeiro, RJ 22453-900, PO Box 38097, Brazil.

E-mail addresses: guimota@gmail.com (G.L.A. Mota), raul@ele.puc-rio.br (R.Q. Feitosa), heitor@cnps.embrapa.br (H.L.C. Coutinho), liedtke@tnt.uni-hannover.de (C.-E. Liedtke), mueller@tnt.uni-hannover.de (S. Müller), pakzad@ipi.uni-hannover.de (K. Pakzad), maggie.simoes@gmail.com (M.S.P. Meirelles).

1. Introduction

One of the most important applications of remote sensing technology is the regular update of land-cover land-use (LCLU) maps. In addition to the image being classified, a previous image of the same area and its classification are usually available in these applications. A photointerpreter often uses this prior information to

aid in the visual interpretation process. Similarly, many automatic image interpretation approaches try to explore the temporal correlation between images of the same geographical area acquired at different dates.

Every image classifier incorporates some form of knowledge representation. This work considers three knowledge modalities: spectral, spatial and temporal. The spectral knowledge relies on the spectral appearance, which characterizes the object by itself opposed to characterizing its component parts or its relationships with other objects. This is the simplest and most widely used form of knowledge. By lacking an explicit high-level knowledge representation, such classifiers are seldom labelled as knowledge-based.

The second category considered in this work is the spatial knowledge. It incorporates some higher level knowledge representation, in addition to an object's overall characteristics. This may include a description of its parts, as well as its relationships with other objects. Input data for a spatial classifier can be derived from different sensors, GIS databases or the results of a spectral classifier. Spatial knowledge can be described by different techniques such as predicate logic (Epstein, 2000), frames (Minsky, 1974), production rules (Clément et al., 2003) and semantic nets (Liedtke et al., 2001).

Generally, spatial knowledge is used in combination with spectral knowledge, usually referring to object features on the image being classified and not taking the history of the object into account. These two forms of knowledge are termed non-temporal. A third knowledge modality is temporal. It is related to the ability of a photointerpreter to utilize an earlier data set of the same area during the visual interpretation taking into account the dynamics of the objects and classes of the target area. The temporal knowledge usually is derived from classifiers combined with other knowledge forms to create a multitemporal classifier (Growe, 1999).

The multitemporal classification can be treated as a multisource classification problem (*e.g.* Lee et al., 1987) where the central issue is how to combine the temporal data sets. Most solutions proposed thus far (see next section for references) rely on the simplifying assumption that transition probabilities are independent from the spectral/spatial information (*e.g.* Jeon and Landgrebe, 1999).

The present work proposes a new multitemporal image interpretation framework for the LULC map update application. The method uses class transition diagrams that express temporal class dependencies in a fuzzy way. No intersource independence is assumed so that the method has the potential to capture the correlation between spectral, spatial and temporal infor-

mation. Class transition possibilities can be estimated by means of a Genetic Algorithm. This approach was adopted due to its ability to find, based on examples, appropriate solutions for complex relations without the need for an explicit model. After estimating the classifier's parameters, the Genetic Algorithm is substituted by a full deterministic procedure.

The main contributions of the present work are:

- A temporal knowledge modeling method containing a novel procedure to estimate class transition possibilities,
- A method to integrate information from distinct time instants that is able to capture the temporal intersource correlation, and
- An evaluation of the potential contribution of the temporal knowledge in comparison to other knowledge forms, particularly with respect to steady features.

The present method uses fuzzy logic concepts and techniques (Zadeh, 1978; Kuncheva, 2000; Mendel, 1995) instead of a probabilistic approach. This choice is mainly due to their ability to represent vague and imprecise knowledge by means of fuzzy rules in such a way that even those that are not familiar with the underlying theory are able to understand the general meaning of the knowledge being represented.

The present proposal is validated on a set of LANDSAT-TM images from Midwest Brazil acquired over three consecutive years during the dry season. Temporal relationships between classes were determined with the help of an expert well acquainted with the class dynamics of the test area. The experimental results indicated that the temporal knowledge may subsume the spatial knowledge in order to simplify the usually troublesome process of knowledge acquisition.

The remaining parts of this paper are organized as follows. Section 2 presents an overview of related works on multitemporal knowledge representation. Section 3 describes the fuzzy classification framework with emphasis on the proposed temporal knowledge representation method. Section 4 describes the experiments and discusses their results. Section 5 presents some concluding remarks and a discussion of future research directions.

2. Related research

The literature discusses different approaches for the multitemporal interpretation of remote sensing data with change detection being the most common. A recent

survey that clearly summarizes the most relevant change detection algorithms was presented by Lu et al. (2004).

Change detection plays an important role in land cover map updates. A time intensive task in this type of application is the gathering of a sufficient number of training samples. Generally, the class labels in a previous image are already available, but one still has to select training samples from the new image to be classified. These partially unsupervised techniques perform this task automatically (Bruzzone and Prieto, 2001). Such approaches use no "ground truth" information from the specific image to be classified, but instead use a training set related to a previous image of the same geographical area.

A partially unsupervised multitemporal classification methodology has been introduced by Bruzzone and Prieto (2001, 2002) that estimates the class conditional probability densities in the image by using a specific formulation of the Expectation-Maximization (EM) Algorithm (Dempster et al., 1977).

Another partially unsupervised proposal (Cazes et al., 2004) uses change detection as the initial step. Objects in the most recent image, set as "stable" by a change detection procedure, are selected as training samples carrying the same labels they had in the previous image.

Another group of techniques explores temporal relations and approaches image interpretation as a data fusion problem. In the remote sensing arena, research on data fusion was initially restricted to multisensor/multisource image classifications. One early report proposes a contextual classifier that considers both spatial and temporal interpixel class dependencies (Jeon and Landgrebe, 1992). This classifier uses Random Gibbs Fields to model class coherence among neighbouring pixels in terms of prior probabilities, while the temporal aspect is modelled by class transition probabilities.

A method for data fusion was developed that incorporates the multisensor/multisource, as well as the temporal aspects (Solberg et al., 1994; Jeon and Landgrebe, 1999). In this method, the *a priori* information on the probabilities of class changes between image acquisition epochs is incorporated into a single-time model. A methodology to estimate the class transition probabilities is a central issue in many proposals about multitemporal classification. Most approaches assume class conditional independence in the time domain (Bruzzone et al., 1999; Bruzzone and Prieto, 2002). Other works present methods that do not assume independence and try to capture intersource (spectral, spatial and temporal) correlations by means of neural networks (Melgani et al., 2001, 2003).

A further group of related approaches compares images of different epochs at the semantic level with different conditions for possible changes between objects from one epoch to another being postulated. These are described by means of a class transition diagram; temporal part of the prior knowledge (Pakzad, 2002; Pakzad et al., 2003). The interpretation process takes advantage of the knowledge on possible class changes. Some kind of estimation of class transition probabilities could add significant value to the method.

These diagrams can be used to identify possible class changes and to restrict the number of classes being considered for a given image segment (Bückner et al., 1999; Growe et al., 2000). Pakzad (2001) and Growe (2001) associated each class transition with a value that expresses the probability that it might occur within a given time period. However, these probabilities merely establish the search order for a solution through a semantic network. In such approaches, transition probabilities do not take part in the computation of any discriminant function.

The present work proposes a fuzzy multitemporal classification procedure that combines non-temporal and temporal knowledge. The non-temporal approach can consider both its spectral and spatial aspects. Nonetheless, while spatial knowledge is mostly time invariant, spectral knowledge can vary significantly over time due to many factors, such as differences in atmospheric conditions. This problem is avoided in the present proposal by using a partially unsupervised approach as proposed by Cazes et al. (2004). Another novelty of the present methodology is the use of a possibilistic approach (Mota, 2004) instead of a probabilistic one. Moreover, by not assuming data source independence, the method is able to exploit the correlation among spectral, spatial and temporal data. This is achieved by using a Genetic Algorithm to estimate the class transition possibilities based on examples without assuming any pre-defined distribution of the data sets. In addition, it is worth mentioning that our method can be applied to both pixel-wise and object-oriented classification (Blaschke and Strobl, 2001).

3. Multitemporal classification model

3.1. Problem formulation

Suppose a multitemporal data set comprising an image $\mathbf{I}_t$ acquired at time t. Let $\mathbf{\Omega} = \{\omega_1, \omega_2, \ldots, \omega_n\}$ be the set of distinguishable LULC classes at both t and $t - \Delta t$. Let $\mathbf{x}_t(o)$ denote a p-dimensional vector in $\mathbf{R}^p$, composed of spectral and spatial features ($x_{t,i}(o)$) for

$i=1,...,p$) extracted from an object o in image $\mathbf{I}_t$, and eventually from a GIS database. With $\mathbf{w}_{t-\Delta t}(o)$ we denote the *crisp* label vector of object o at a previous time $t-\Delta t$. Therefore, $\mathbf{w}_{t-\Delta t}(o)$ is a vector in space $\mathbf{W}$, defined as follows:

$$\mathbf{W} = \{\mathbf{w} \in \mathbf{R}^n : w_i \in [0, 1] \ \forall \ i = 1, ... n, \sum_{i=1}^{n} w_i = 1\}$$

In other words, $\mathbf{w}_{t-\Delta t}(o)$ is a *n*-dimensional unitary vector of the form $[0 \ ... \ 1 \ ... \ 0]^T$ having "1" in its *i*th component ($w_{t-\Delta t,i}(o)=1$) and "0" ($w_{t-\Delta t,j}(o)=0$, for $j \neq i$) otherwise, indicating that the object o belongs to the class ω_i at time $t-\Delta t$. Henceforth, we drop the explicit indication to the object by writing its feature vector and its crisp label vector at a previous time as $\mathbf{x}_t$ and $\mathbf{w}_{t-\Delta t}$, respectively.

The problem consists of identifying the crisp vector label $\mathbf{w}_t$ of each image object at time t based upon $\mathbf{x}_t$ and $\mathbf{w}_{t-\Delta t}$. Thus, we are searching for a function $\mathbf{M}$ such that:

$$\mathbf{w}_t = \mathbf{M}(\mathbf{x}_t, \mathbf{w}_{t-\Delta t}), \tag{1}$$

where $\mathbf{M}$ is a function of the form $\mathbf{M}: \mathbf{R}^{p+n} \rightarrow \mathbf{W}$.

3.2. General model description

Fig. 1 shows an overview of the solution proposed. The classification occurs in four main steps:

− Non-temporal classification
− Temporal classification
− Decision fusion
− Defuzzification.

Two types of classifiers cooperate during the recognition task. The first, called non-temporal, is based exclusively on spectral/spatial object descriptions ($\mathbf{x}_t$) at the image acquisition date (t). The second type of classifier, temporal, classifies image objects based solely on the previous object class ($\mathbf{w}_{t-\Delta t}$) at $t-\Delta t$. The non-temporal and the temporal classifiers are combined to form a potentially more accurate classifier called multitemporal.

Following a fuzzy approach, the n classes in Ω are modelled at each classifier by fuzzy sets whose membership functions map points in the corresponding input feature space into the interval [0 1].

The non-temporal classifier produces a n-dimensional fuzzy label vector denoted by $\boldsymbol{\alpha}=[\alpha_1, \alpha_2, \ldots, \alpha_n]^T$ where α_i stays for the membership of the image object assigned by the non-temporal classifier to the class ω_i, for $i=1,2,\ldots,n$ and for at least one i, $\alpha_i \neq 0$. It can be described by a function of the form $\mathbf{A}: \mathbf{R}^p \rightarrow \mathbf{F}$, such that:

$$\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_n] = \mathbf{A}(x_t), \tag{2}$$

where the space $\mathbf{F}$ is given by:

$$\mathbf{F} = \{\boldsymbol{\alpha} \in \mathbf{R}^n : \alpha_i \in [0, 1] \forall \ i = 1, ... n,$$
$$\text{and } \alpha_i > 0 \text{ for at least one } i\} \cdot$$

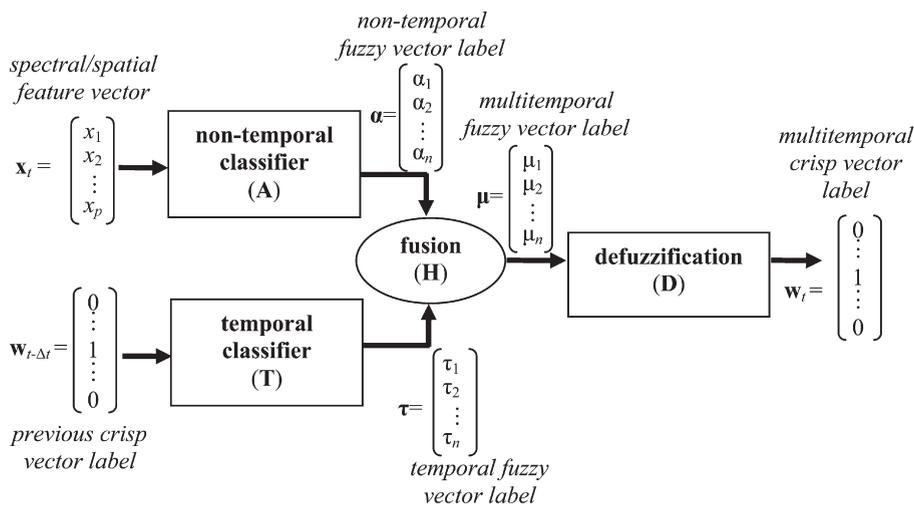In the temporal classifier, classes are analogously represented by fuzzy sets. The temporal classifier is

Fig. 1. The multitemporal classification procedure.

given by a function $\mathbf{T}$ of the form $\mathbf{T}: \mathbf{W} \rightarrow \mathbf{F}$. The value $\boldsymbol{\tau} = \mathbf{T}(\mathbf{w}_{t-\Delta t})$ is a fuzzy label vector $\boldsymbol{\tau} = [\tau_1, \tau_2, \dots, \tau_n]^T$ where τ_j is the membership of the image object assigned by the temporal classifier to the class ω_j.

The two fuzzy classification outcomes $\boldsymbol{\alpha}$ and $\boldsymbol{\tau}$ are then combined in the next step labelled in Fig. 1 as fusion. As a result, a multitemporal fuzzy label vector

$$\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_n]^T = \mathbf{H}(\boldsymbol{\alpha}, \boldsymbol{\tau}) \tag{3}$$

is produced by a function of the form $\mathbf{H}: [0\ 1]^n \times [0\ 1]^n \rightarrow [0\ 1]^n$.

The final step of the classification method depicted in Fig. 1 is the defuzzification that transforms a fuzzy vector label into a crisp one. The defuzzification function of the form $\mathbf{D}: \mathbf{F} \rightarrow \mathbf{W}$ assigns the image object being classified to the class ω_j that has the highest membership values. So $\mathbf{D}$ is given by the formula:

$$\begin{aligned} \mathbf{w}_t &= (w_{t,1}, \dots, w_{t,n})^T \\ &= \mathbf{D}(\boldsymbol{\mu}), \text{where} \begin{cases} w_{t,i} = 1, \text{for } \mu_i = \max_j \{\mu_j\} \ . \\ w_{t,i} = 0, \text{otherwise} \end{cases} \end{aligned} \tag{4}$$

Replacing Eqs. (2) and (3) in Eq. (1) results in the multitemporal classifier $\mathbf{M}$ proposed in this paper expressed as:

$$\mathbf{w}_t = \mathbf{M}(\mathbf{x}_t, \mathbf{w}_{t-\Delta t}) = \mathbf{D}[\mathbf{H}[\mathbf{A}(\mathbf{x}_t), \mathbf{T}(\mathbf{w}_{t-\Delta t})]]. \tag{5}$$

Indeed, many well known classification methods fit in the description of the non-temporal classifier $\mathbf{A}$ presented above and can be used in this scheme. Similarly, the defuzzification method $\mathbf{D}$ portrayed thus far has been used in many pattern recognition applications. The design of the temporal classifier $\mathbf{T}$ and the fusion function $\mathbf{H}$ are described in detail in the next sections.

3.3. Design of the temporal classifier

The temporal knowledge is represented by the possibilities P_{ij} that an image object belonging to a class ω_i at a previous time $t - \Delta t$ changes to the class ω_j at a later time t for a given Δt and $i, j = 1, 2, \dots, n$. This can be represented pictorially by a class transition diagram.

The class transition diagram is a graph whose nodes correspond to the classes and links of the plausible class transitions between $t - \Delta t$ and t. Each link is labelled with a value P_{ij} ($0 \le P_{ij} \le 1$) that expresses the possibility that an object of class ω_i changes to class ω_j within Δt. For simplicity, links with $P_{ij} = 0$ are not drawn.

The temporal knowledge can be modelled equivalently by n discrete fuzzy sets. As proposed by Mendel (1995), discrete fuzzy sets are represented by a sequence of pairs composed of a fuzzy membership value, a slash, and the discrete value of the linguistic variable, which in this case is the previous classification at time t. The distinct pairs are separated by the symbol + that in this notation is a mere separator. Hence, temporal knowledge may be represented by a set of n expressions as in the example below:

$$\begin{aligned} \mathbf{P}_j &= P_{1j}/\omega_1 + P_{2j}/\omega_2 + \dots + P_{nj}/\omega_n, \\ &\quad \text{for } j = 1, 2 \dots n. \end{aligned}$$

An even more compact representation can be formulated by writing each set as a distinct column of a $n \times n$ matrix $\mathbf{P} = \{P_{ij}\}$ called a Transition Possibility Matrix (TPM).

The temporal classifier can be derived from the TPM in the following way. Given the class ω_i of an image object at time $t - \Delta t$, the temporal classifier produces a membership vector:

$$\begin{aligned} \boldsymbol{\tau} &= \mathbf{T}(\mathbf{w}_{t-\Delta t}) = [\tau_1, \tau_2, \dots, \tau_n]^T \\ &= [P_{i1}, P_{i2}, \dots, P_{in}]^T \text{ where } w_{t-\Delta t, i} = 1 \end{aligned} \tag{6}$$

that is the i-th row of the transition matrix $\mathbf{P}$.

3.3.1. Fusion of non-temporal and temporal fuzzy labels

The temporal fuzzy label $\boldsymbol{\alpha}$ is combined with the result $\boldsymbol{\tau}$ of the non-temporal classifier into a single meaningful fuzzy label vector $\boldsymbol{\mu}$ that expresses the overall membership of a given image object to each of the n classes. In this study, the outcome μ_i of the fusion step relative to class ω_i is a function h of α_i and τ_i, of the form:

$$h : [0\ 1]^2 \rightarrow [0\ 1].$$

This function is called aggregation function or aggregation operation by the fuzzy set theory literature. According to Klir and Yuan (1995), an aggregation function must satisfy the following axioms:

Axiom 1. $h(0,0) = 0$ and $h(1,1) = 1$ (boundary conditions)

Axiom 2. For any pair (a_1, a_2) and (b_1, b_2) such that $a_i, b_i, \in [0\ 1]$ if $a_i \le b_i$, for $i = 1,2$, then:

$$h(a_1, a_2) \le h(b_1, b_2),$$

that is, h is monotonic, increasing in all its arguments.

Axiom 3. h is a continuous function.

Aggregation functions are usually expected to meet two additional conditions, namely:

Axiom 4. h is symmetric in its arguments; that is:

$$h(a_1, a_2) = h(a_2, a_1).$$

Axiom 5. h is idempotent,[1] that is:

$$h(a, a) = a$$

for all $a \in [0\ 1]$.

A family of functions that meet all five axioms is the generalized means, which is defined by the formula:

$$h_\gamma(a_1, a_2) = \left(\frac{a_1^\gamma + a_2^\gamma}{2} \right)^{1/\gamma} \tag{7}$$

where $\gamma \in R$, $(\gamma \neq 0)$ and $a_i \neq 0$ for $i = 1, 2$ when $\gamma < 0$.

It can be demonstrated (Klir and Yuan, 1995) that the functions max, arithmetic mean, geometric mean[2], harmonic mean and min are special cases of the generalized mean, for $\gamma \to +\infty$, $\gamma = 1$, $\gamma \to 0$, $\gamma = -1$ and $\gamma \to -\infty$, respectively.

Other functions, apart from the generalized means, could be used for aggregation with different levels of performance. An extensive discussion of this topic goes beyond the scope of the present work. We limit ourselves to consider the geometric mean as an aggregation function. In this case, the final multitemporal fuzzy classification will be given by:

$$\begin{aligned}
\boldsymbol{\mu} = \mathbf{H}(\boldsymbol{\alpha}, \boldsymbol{\tau}) &= [\mu_1, \mu_2 \ldots, \mu_n]^T \\
&= [(\alpha_1 \tau_1)^{1/2}, (\alpha_1 \tau_2)^{1/2}, \ldots, (\alpha_n \tau_n)^{1/2}]^T.
\end{aligned} \tag{8}$$

This choice is inspired by an analogy with the maximum probability classifier where the prior probability scales the probability density to build the discriminant function (see *e.g.* Shackelford and Davis, 2003). However, the product cannot be qualified as an aggregation function since it returns a value inferior (or equal) to both its arguments. Therefore, the use of the product as an aggregation function would have as a counterintuitive consequence that the more knowledge we combine, the lower the membership values become. The geometric mean circumvents this inconvenience by computing the square root of the product. The experimental analysis reported in Section 4.4 provides further

empirical elements favoring the choice of the geometric mean as an aggregation function. An investigation of aggregation functions other than the geometric function is the subject of future research.

3.3.2. Estimating the model parameters

Class transitions can be treated by two ways. The first, called a crisp transition, considers the possibilities as either "1" or "0", depending whether the transition is possible or not. Such information can be easily delivered by a photointerpreter with some experience in the test area without any need for training data.

A more powerful alternative, called a soft transition, considers the transition possibilities as real numbers in the interval [0 1]. A key issue in this proposal is the estimation of the soft transition possibilities P_{ij}. This can be done in three sequential steps:

1) Trainingset selection

   The necessary training data to estimate soft transition possibilities becomes apparent in Fig. 1. One needs a representative set of samples for which the features $\mathbf{x}_t$ at a time t, and their crisp label $\mathbf{w}_t$ and $\mathbf{w}_{t-\Delta t}$ at time t and $t - \Delta t$ are known in advance. For an accurate estimation, all possible class transitions must be represented in the selected samples. Such training objects may be chosen manually by a photointerpreter according to a conventional supervised approach. For LULC update applications, this can be done automatically by applying a partially unsupervised training set selection as mentioned in Section 2.

2) Training of the non-temporal classifier

   The non-temporal classifier parameters are adjusted by using the training samples selected in the previous step according to the particular classifier design elected for this task.

3) Estimation of the Transition Possibility Matrix

   Once the design of the non-temporal classifiers has been completed, the Transition Possibility Matrix TPM is estimated. By using the same training set selected in Step 1, one searches for the values P_{ij} that maximize the performance measure computed on the known crisp classification of the training objects. The photointerpreter can indicate the impossible transitions (whose possibilities are set to zero) and, for each source class, the outgoing transition with the highest possibility. This *a priori* knowledge may considerably accelerate the search for optimum and plausible possibility values. In addition, depending on the selected aggregation function, the number of parameters to estimate can be reduced further. Let's consider again the geometric mean. Recall that, for

---

[1] Notice that Axiom 5 is implied in Axiom 1.

[2] To compute this limit, just take the natural logarithm of the generalized mean, apply l'Hôpital rule and calculate the exponential of the result.

the defuzzification function **D**, only the ordering of the final memberships μ_i matters. This does not change if the vector τ in Eq. (3) is scaled by any positive real number. Therefore, each row of the matrix **P** can be estimated only up to a scale factor. We can eliminate this ambiguity without affecting the final crisp result by setting the highest transition possibility at each row of the matrix **P** to "1". This reduces the number of transition possibilities to estimate by n, the number of classes.

Training of the non-temporal classifier (Step 2) requires samples to represent each of the n classes at a given time t. However, training of the temporal classifier (Step 3) requires samples representing all possible class transitions whose number can be as large as n^2. Collecting samples for all possible transitions may therefore be a troublesome task for the photointerpreter. We can overcome this hindrance by estimating the TPM upon data sets ($\mathbf{x}_{t'}$, $\mathbf{w}_{t'}$ and $\mathbf{w}_{t'-\Delta t}$) relative to a pair of preceding dates t' and $t'-\Delta t$ ($t'<t$) also separated by Δt. In LULC map update applications, such data sets of previous epochs are usually available. It is important to use images taken at the same time of the year to avoid seasonal effects. Steps 1–3 are performed on the earlier data sets to estimate the matrix **P**. Notice that in this case all objects in the entire image may be selected for training. The estimate of **P** is then stored to be used for the classification of the image $\mathbf{I}_t$. We are in this way implicitly assuming that the transition possibilities in images from the same season in the same geographical area do not significantly change over time. Certainly the closer t' is to t the more plausible this assumption is.

The temporal model described so far is not bound to any particular optimization method. The most appropriate technique will depend on the objective function to be optimized. In many image analysis problems, the objective function that truly expresses the goodness of the solution does not meet the properties required by most conventional calculus-based optimization methods. Gradient descent methods, for instance, require that the objective function is differentiable. In this study, we use a Genetic Algorithm for this task (Schmiedle et al., 2002). It does not impose any assumption concerning the class probability distribution and can be applied to any performance metric used as an objective function.

It is important to emphasize that the proposed method produces transition possibility estimates that are specific to a particular target site, as well as to a particular classifier design. Possibility values estimated upon one area, and for a particular classifier, may not be applicable to other geographical areas or to a distinct non-temporal classifier design.

It is worth mentioning that the ability of our method to capture intersource correlation is a result of the optimization procedure described above. The Genetic Algorithm (GA) enables one to search for possibility values that optimize the performance of the entire classifier comprising both the temporal and non-temporal data sets. It is important to stress why not to use a more intuitive maximum likelihood based technique to estimate transition possibilities based on the available ground truth at both dates ($\mathbf{w}_{t'}$ and $\mathbf{w}_{t'-\Delta t}$). In this case, the possibility values would be adjusted to maximize the performance of the temporal classifier alone. The disadvantage of this method is that it disregards the temporal classifier and consequently fails to capture the temporal correlation between both data sets.

4. Experiments

A software prototype was written in order to validate the proposed method in a real application. This section describes how the experimental evaluation was designed and its results.

4.1. Description of the data set

4.1.1. Test area

Our test area is situated in the County of Alcinópolis in the State of Mato Grosso do Sul, Brazil. It is covered by a single LANDSAT 7 scene (224-073). Two subsets were used covering around 950 km^2. The segments picked up from both subsets were treated in our experiments as if they belonged to a single contiguous area.

The streams in Alcinópolis are located in the Taquari River sub-basin, part of the Upper Paraguai River basin, and the headwaters of the Pantanal wetlands, one of the most important ecosystems in South America. Much of Alcinópolis County is used to raise cattle for the meat industry. Given the region's soil, topography, geology, and climate, the excessive numbers of cattle placed on pasturelands makes it highly vulnerable to erosion processes. Therefore, environmental recovery efforts and the monitoring of land cover changes in theses areas are urgently needed.

4.1.2. Image pre-processing and segmentation

Bands 5, 4 and 3 of three LANDSAT-7 images were used in the experiments. The images were acquired during the dry seasons of 1999 (August 5), 2000
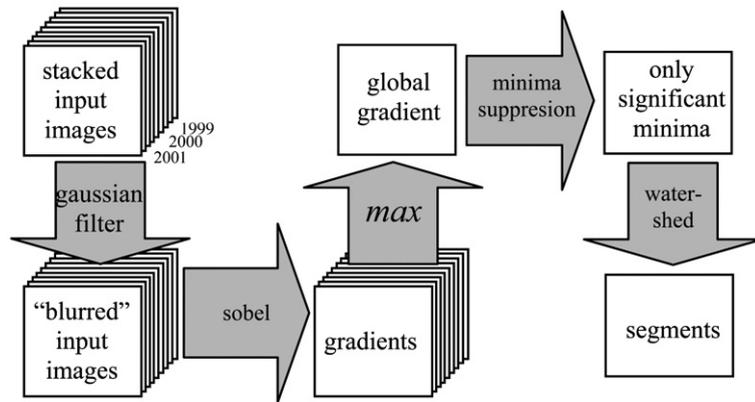
Fig. 2. The watershed based segmentation procedure.

(August 7), and 2001 (August 10), being co-registered using a GIS database as reference data.

The registered images were segmented producing spectrally homogeneous objects through the following sequential steps (see Fig. 2):

a) The three bands of all three images were stacked forming an artificial nine-band image. A spatial Gaussian lowpass filter with a given standard deviation σ was applied to each band in order to eliminate noise effects and small details.
b) After lowpass filtering, the gradient of each band was computed by using the Sobel operator.
c) The maximum value of the gradient magnitude across all bands was then computed resulting in a two-dimensional matrix. This guarantees that a significant discontinuity on any band will suffice to bring about a segment border on all images.
d) All local minima in the gradient matrix whose depth is lower than a given value δ were suppressed by

applying the h-minima transformation (Soille, 2003). This is intended to avoid oversegmentation.
e) Finally, the Watershed Algorithm (Vincent and Soille, 1991) was applied to the result of the previous step.

The values of the segmentation parameters σ and δ were selected empirically by a photointerpreter.

4.1.3. Validation data

Table 1 describes the land-use classes considered in our experiments. To assess the performance of our method a reference classification for all three years was created visually by a photointerpreter with expertise in land cover classification. This was achieved using as ancillary data a videography taken in October 2001, the LANDSAT images, a drainage map, and a digital elevation model. Only segments covered by the flight lines of the videography were considered in this analysis. Segments set as "unclassified" by the photointerpreter in

Table 1
Land-use classes

Label	Class	Description	Number of segments		
			1999	2000	2001
Bare soil (ω_1)	Bare soil	A soil that has been degraded by erosion or that is being prepared for cultivation.	96	84	61
Riparian (ω_2)	Riparian forest	Dense woodland alongside rivers and streams.	62	62	57
Pasture (ω_3)	Pasture	Cultivated pasture for cattle nutrition.	496	510	538
Water (ω_4)	Water bodies	Water or swampland	28	28	28
Savannah (ω_5)	Dense savannah	Formation of low trees (8–12 m) densely packed, but without significant contact between their crowns so that the shading effect is not complete allowing development of an understory vegetation containing grasses, dwarf palm trees and ground woody plants - Brazillian cerrado (Coutinho, 1978).	134	132	132
Regeneration (ω_6)	Dense savannah in regeneration	An area used previously for pasture that was left aside by the farmer and is regenerating its native vegetation.	6	6	6

Table 2a
Number of possible class transitions that occurred in 1999–2000

1999	2000						Total
	Bare soil	Riparian	Pasture	Water	Savannah	Regeneration	
Bare soil	71		25				96
Riparian	0	62	0				62
Pasture	11		485				496
Water				28			28
Savannah	2		0		132		134
Regeneration	0		0		0	6	6
total	84	62	510	28	132	6	822

any of the three years were also discarded from the experiments. Table 1 also shows the number of segments used in the experiments for each class.

Tables 2a,b show the class transitions observed in two consecutive years. The blank table positions correspond to class transitions considered to be impossible by the photointerpreter.

4.2. Non-temporal classification

One objective of this work is to assess the contribution of the multitemporal knowledge to the classification performance in comparison to other forms of knowledge. For this assessment, two non-temporal classifiers were implemented according to the approach described in Section 3 using a pure spectral classifier and a spectral/spatial classifier. In both cases, the training samples were selected from the image taken in 2001 using a partially unsupervised method as proposed in Cazes et al. (2004) and based on the image taken in 2000 and its classification.

4.2.1. Spectral classification

Feature vectors $\mathbf{x}$ were built for each segment by stacking the mean spectral values of each available band. It was assumed that all classes ω_i can be appropriately modelled by a normal distribution $N(\overline{\mathbf{x}}, \Sigma_{\omega i})$.

So, a Gaussian-shaped membership function $\mathrm{SMF}_{\omega i}(\mathbf{x})$ was selected for all classes ω_i, given by the formula below.

$$\mathrm{SMF}_{\omega i}(\mathbf{x}) = \exp\left[-\frac{(\mathbf{x} - \overline{\mathbf{x}}_{\omega i})^T \sum_{\omega i}^{-1}(\mathbf{x} - \overline{\mathbf{x}}_{\omega i})}{2} \right]. \quad (9)$$

For $\omega_i \in$ {baresoil, riparian, pasture, water, savannah, regeneration}; $\overline{\mathbf{x}}_{\omega i}$, and $\Sigma_{\omega i}$ correspond respectively to the mean and to the covariance matrix of the class ω_i. Estimates for these parameters were computed by standard procedures based on the training samples selected according to the automatic method proposed in Cazes et al. (2004). Replacing Eq. (9) into Eq. (2), the non-temporal classifier will be given by:

$$\boldsymbol{\alpha} = \mathbf{A}(\mathbf{x}) = \begin{bmatrix} \mathrm{SMF}_{\mathrm{baresoil}}(\mathbf{x}) \\ \mathrm{SMF}_{\mathrm{riparian}}(\mathbf{x}) \\ \mathrm{SMF}_{\mathrm{pasture}}(\mathbf{x}) \\ \mathrm{SMF}_{\mathrm{water}}(\mathbf{x}) \\ \mathrm{SMF}_{\mathrm{savannah}}(\mathbf{x}) \\ \mathrm{SMF}_{\mathrm{regeneration}}(\mathbf{x}) \end{bmatrix}. \quad (10)$$

In order to compare the performance of a simple spectral approach with the proposed multitemporal method, the membership vector $\boldsymbol{\alpha}$ produced by the spectral classifier shown in the equation above was

Table 2b
Number of possible class transitions that occurred in 2000–2001

2000	2001						Total
	Bare soil	Riparian	Pasture	Water	Savannah	Regeneration	
Bare soil	49		35				84
Riparian	5	57	0				62
Pasture	7		503				510
Water				28			28
Savannah	0		0		132		132
Regeneration	0		0		0	6	6
total	61	57	538	28	132	6	822

submitted to the defuzzification step producing a crisp result. The performance obtained in this way is presented in Section 4.4.

4.2.2. Spectral/spatial classification

The spatial/spectral knowledge is represented by a rule set. An analysis of the crisp results produced by the spectral classifier revealed that it often confuses riparian forest and dense savannah; 21% of the riparian forest segments were assigned to dense savannah, while 26% of the dense savannah were labelled as riparian forest. Both classes are dense forest and have, therefore, similar spectral responses. However, since riparian forest occurs alongside rivers and dense savannah in less humid places far from the watercourses, they are characterized by different species of vegetation. This confusion can be solved by taking the distance d of the object to the closest water body into consideration. This information can be obtained from the drainage map.

Hence, a fuzzy set named *short* is created whose membership is a function of the distance d to the closest water body as shown in Fig. 3.

Further confusion is due to shaded areas in mountains or dense savannah that become spectrally similar to water bodies. This confusion can be attenuated by noticing that water bodies in the test areas occur only at low altitudes. Thus, the crisp set *high* was created which is true, if and only if the average elevation e of the image object is greater than a threshold E defined by the photointerpreter. So,

$$high(e) = \begin{cases} 0, \text{if } e \leq E \\ 1, \text{if } e > E \end{cases}. \tag{11}$$

The elevation information was obtained from the digital terrain model. The overall reasoning explained above can be modelled by the fuzzy rules presented in Table 3.

Notice that rules R_2, R_4, R_5, and R_6 express the strategy just proposed to solve the main sources of confusion. Rules R_1, R_3, and R_7 are evidence that the
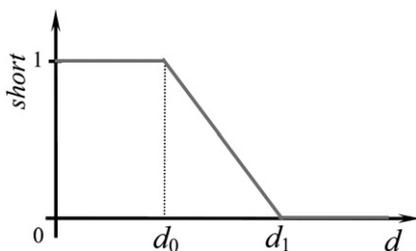

Fig. 3. Membership functions of the fuzzy set short.

Table 3
Rule base for the spectral/spatial classification

Label	Rule
R_1	IF y IS $SMF_{baresoil}$ THEN object IS bare soil
R_2	IF (y IS $SMF_{riparian}$ OR y IS $SMF_{savannah}$) AND d IS short THEN object IS riparian
R_3	IF y IS $SMF_{pasture}$ THEN object IS pasture
R_4	IF (y IS $SMF_{riparian}$ OR y IS $SMF_{savannah}$) AND d IS NOT(short) THEN object IS savannah
R_5	IF (y IS SMF_{water} OR y IS $SMF_{savannah}$) AND e IS NOT(high) THEN object IS water
R_6	IF (y IS SMF_{water} OR y IS $SMF_{savannah}$) AND e IS high THEN object IS savannah
R_7	IF y IS $SMF_{regeneration}$ THEN object IS regeneration

spatial knowledge is not explored in order to refine the classification of the classes bare soil, pasture and dense savannah in regeneration.

As noted in Jang and Sun (1995), the object membership to the class in the THEN clause (the consequent) of each rule will be given by computing the fuzzy formula in the IF clause (the antecedent). In the cases where there is more than one rule with the same consequent, such as in rules R_4 and R_6, the final membership will be given by the maximum membership value derived from these rules.

Using the function product, max and $(1-x)$ to implement respectively the fuzzy AND, OR and NOT operators, the fuzzy spectral/spatial classifier defined in Eq. (2) produces the fuzzy label vector $\boldsymbol{\alpha} = (\alpha_1, \dots \alpha_6)^T$ according to the equations below.

$$\alpha_1 = SMF_{baresoil}(\mathbf{x})$$
$$\alpha_2 = \max[SMF_{riparian}(\mathbf{x}), SMF_{savannah}(\mathbf{x})] \cdot short(d)$$
$$\alpha_3 = SMF_{pasture}(\mathbf{x})$$
$$\alpha_4 = \max[SMF_{water}(\mathbf{x}), SMF_{savannah}(\mathbf{x})] \cdot [1 - high(h)]$$
$$\alpha_5 = \max\{\max[SMF_{ripparian}(\mathbf{x}), SMF_{savannah}(\mathbf{x})] \cdot [1 - short(d)],$$
$$\quad \max[SMF_{water}(\mathbf{x}), SMF_{savannah}(\mathbf{x})] \cdot high(h)\}$$
$$\alpha_6 = SMF_{regeneration}(\mathbf{x})$$
$$\tag{12}$$

The parameters of the spectral membership functions $SMF_{\omega i}$, were estimated in the same manner as in Section 4.2.1.

The values of the parameters d_0 and d_1 in Fig. 3 were obtained by repeating the same training procedure explained above on the 2000 image and searching

Table 4
Possible class transitions for $\Delta t = 1$ year

Class at $t - \Delta t$	Class at t	Process
Savannah	Pasture	Deforestation for introduction of a new pasture.
	Bare soil	Land preparation
Regeneration	Pasture	Deforestation for introduction of a new pasture.
	Bare soil	Land preparation
	Savannah	Favorable environmental conditions allowing the rehabilitation of the original vegetal cover.
Riparian	Pasture	Deforestation for introduction of a new pasture.
	Bare soil	Land preparation for planting a new pasture.
Pasture	Bare soil	Land preparation for planting or enewal of pastures or the offset of soil degradation process (erosion)
Bare soil	Pasture	New pasture or renewal of degraded pasture

for the values that maximize the mean class recognition rate F for the image objects of 2000 as given in Eq. (13),

$$F = \frac{1}{n} \sum_{i=1}^{n} \frac{nc_i}{n_i} \qquad (13)$$

where nc_i is the number of training objects of class ω_i correctly assigned by the spectral/spatial classifier, n_i is the total number of training objects of class ω_i, and n is, as before, the number of classes in the application. A Genetic Algorithm was used to compute the values for d_0 and d_1. The algorithm was executed through a number of generations until it was observed that there was no improvement in the class recognition rate F. This was reached for $d_0 = 1$ pixel and $d_1 = 21$ pixels.

As in the preceding section, the fuzzy classification α given in Eq. (12) was defuzzified in order to compare the performance of the spectral/spatial classifier with the multitemporal approach proposed in this paper. The results are presented in Section 4.4.

4.3. Multitemporal classification

The multitemporal model was built from interviews with an agronomic engineer well acquainted with the class dynamics in the test area. The possible class transitions within $\Delta t =$ one year, as well as the processes that cause such transitions are described in Table 4.

Fig. 4 presents the corresponding class transition diagram. The objects tend to continue in the same class as expressed in the diagram by the possibilities being equal to "1". Disregarding the impossible transitions, a total of nine possible transitions still had to be estimated. The geometric mean (see Eq. (3)) was used as the aggregation function.

The temporal classifier was implemented in crisp and soft versions. In both cases, this was done in combination with pure spectral and spectral/spatial classifiers. Thus, four different multitemporal classifiers were investigated. To evaluate the performance of a crisp class transition diagram, as introduced in Section 3.3.2, an experiment was performed setting all unknown possibility values in Fig. 4 to "1" (see experimental results in Columns 4 and 5 of Table 5).

To assess the soft class transition approach, additional training data were required to estimate the transition possibilities. We followed the scheme described in Section 3.3.2 using all image segments for 2000 and their reference classifications in 2000 and in 1999 for training. Accordingly, the non-temporal membership degrees relative to 2000 were computed as described in Sections 4.2.1 and 4.2.2. Next, the transition possibilities were adjusted to maximize the agreement between the classification produced by the classifier and the reference classification in the 2000 image. The possibility values were estimated by a Genetic Algorithm (GA) using the average class recognition rate given in Eq. (13) as an objective function.

The GA was configured in the following way. Every gene of the chromosome is represented by a double
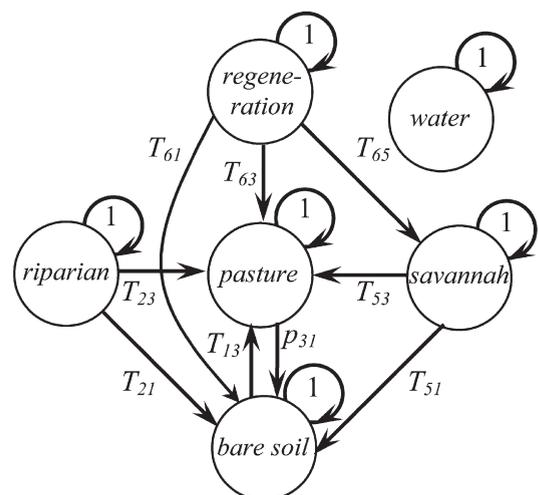

Fig. 4. Class transition diagram for the test area.

Table 5
Experimental average class recognition rate (%) in the 2001 image for different classification approaches

Class	Non-temporal Classification		Multitemporal Classification			
			Crisp transitions		Soft transitions	
	Spectral	Spectral/spatial	Spectral+temporal	Spectral/ spatial+temporal	Spectral+temporal	Spectral/ spatial+temporal
Bare soil	28	28	30	30	89	89
Riparian forest	75	93	98	98	100	100
Pasture	75	75	77	77	93	93
Water bodies	100	100	100	100	100	100
Dense savannah	50	64	93	93	100	100
Dense savannah in regeneration	67	83	67	83	100	100
Mean	66	74	77	80	97	97

precision real number, one value for each parameter being estimated. A population of 100 individuals is assumed. For the next generation, the 50 worse chromosomes will be substituted by reproduction. Reproduction considers the genetic material of one or two individuals selected by a "roulette wheel". The selection is bounded by the individuals' fitness measurements, normalized between 1 and 100. The need for one or two parents depends on the requirements of the selected operator. If crossover, the selection of two parents will be demanded, while if mutation, only one parent will be required. In a linear fashion, during the course of evolution, the probability of selecting a mutation operator increases, while the probabilities associated to crossover operators decrease. In the first generation, the probabilities are: 0.3 for a simple crossover, 0.3 for an arithmetic crossover, 0.2 for a simple mutation, 0.1 for a small creep mutation and 0.1 for a big creep mutation. In the last generation, these values were respectively 0.1, 0.1, 0.3, 0.2, and 0.3. This favors the diversification of the search foci in the early generations, while at the end of the evolution the exploitation of the space around the more potential solutions.

Once the transition possibilities have been identified, the final evaluation was performed on the objects from the 2001 image. In this testing step, the same non-temporal classifier, designed in the experiments described in Section 4.2 for the 2001 image were employed. The result of the temporal classifier was aggregated to the non-temporal outcome according to Eq. (3) producing the overall fuzzy classification, which was subsequently defuzzified (see Fig. 1). The performance of the testing step was measured again by computing the average class recognition rate defined in Eq. (13). This evaluation was conducted starting with both of the non-temporal classifiers presented in Sections 4.2.1 and 4.2.2.

4.4. Results and discussion

The results obtained are summarized in Table 5. It shows the mean class recognition rate achieved in the testing step for the 2001 image. All image objects listed in Table 2b were used in the test.

By comparing the second and third columns relative to the two non-temporal classifiers one can observe the improvements achieved by introducing the spatial knowledge. Typically, the performance increased from 66% to 74%. The performance of the classes dense savannah, dense savannah in regeneration and, specially, riparian forest were the most affected. In fact, not much spatial knowledge can be explored at medium resolution images. A comparatively higher improvement is expected by using spatial knowledge at higher resolutions.

The last four columns of Table 5 refer to the four multitemporal classifiers analyzed in our experiments. Columns labelled as "crisp transitions" show the performance of multitemporal classifiers built by aggregating the crisp temporal knowledge to both non-temporal classifiers. The temporal knowledge enhanced the performance for all classes, in particular for dense savannah.

It is important to observe that there is little change between these two columns. The improvement achieved for dense savannah in regeneration (from 67% to 83%) is actually not so significant because it corresponds to a single additional segment that was correctly classified with the temporal knowledge. Therefore, one could have renounced the spatial reasoning and kept the temporal knowledge without incurring a significant loss in performance. As a matter of fact, the rule base built in Table 3 considers only long-lasting spatial attributes whose values do not change in the time interval

considered ($\Delta t = 1$ year). In the case of rules R_2 and R_4, for instance, the fact that an object was assigned in the past to the class riparian forest or dense savannah implies that the object stays at short and not short distances from river, respectively. Therefore, as far as the separation of these classes is concerned, the temporal and the spatial knowledge were mostly equivalent. A similar reasoning can be formulated for the rules R_5 and R_6. This shows that the crisp multitemporal knowledge can at least partially embody the spatial knowledge within it in that refers to the permanent attributes.

It is also noteworthy to analyze the last two columns of Table 5 that present the results associated with the soft multitemporal knowledge. The mean class recognition rate increased from 77% to 80% in the crisp multitemporal knowledge and to around 97% in the soft approach. These remarkable improvements in performance could be credited to the scarce use of spatial knowledge in our experiments. Columns 6 and 7 again show that there is no significant contribution from the spatial knowledge when the multitemporal knowledge is already being used.

After having emphasized the contribution of the multitemporal approach, some comments concerning the estimation of transition possibilities through Genetic Algorithms must be introduced.

Genetic algorithms (GA) do not guarantee that the global optimum of the objective function will be found. Indeed, GA may encounter some consistent solution that is only locally optimal. The results shown in Table 6 correspond to the best value observed through 20 runs of the GA, each time with a randomly selected start solution. The worst, the mean, and the best class recognition rates found by the GA throughout the 20 experiments are presented in Table 6. By comparing this figure with the number of objects for each class in Table 2, it becomes clear that the largest discrepancy between the worst and the best performance is 5% for the class bare soil, which corresponds to about ±2 ob-

jects. This indicates that the GA was quite stable in our experiments.

Table 2a shows that six possible class transitions do not occur in the training set comprising objects from 1999 and 2000. The corresponding transition possiblities were therefore underestimated by the GA. Looking at Table 2b, we see that five of these six transitions also do not occur between 2000 and 2001 and consequently did not affect the performance measured in our experiments. The exception is the transition riparian→baresoil that does not occur in 1999–2000, but occurs five times in 2000–2001. In this case, the transition possibility was estimated as being near zero. As a consequence, the five objects in the class riparian in 2000 that changed to the class baresoil in 2001, according to the reference classification, were maintained in the class riparian by our multitemporal classifier. This has certainly contributed to the comparatively poor performance obtained for the class baresoil in 2001.

This observation stresses the need of having training examples for all class transitions we want to detect in the image. Indeed, the soft multitemporal approach proposed in this work demands a comparatively large training set in order to perform well. When such training data is not available, the crisp multitemporal model is, according to our experimental results, worth trying.

5. Conclusions and future research

The present paper proposes a novel method to represent temporal knowledge and to combine it with other knowledge forms within an integrated automatic image interpretation framework for LULC map update applications. The method expresses temporal class dependencies by means of a transition diagram that assigns to each class transition a possibility value.

The improvement brought by the temporal knowledge in relation to a pure non-temporal approach was assessed in a real application using medium-resolution LANDSAT images from Midwest Brazil. The multitemporal model parameters were estimated by a Genetic Algorithm based on a pair of images from 1999 and 2000 and tested on a 2001 image of the same area.

The experiments showed that the use of prior knowledge markedly improved the classification performance in comparison to a conventional spectral classification. A further observation was that the temporal knowledge for the most part subsumed the spatial knowledge. This is explained by the fact that the spatial knowledge model built in the experiments was based on steady spatial attributes whose values do not significantly change over time.

Table 6
The worst, mean and best class recognition rate (%) obtained through 20 runs of the Genetic Algorithm

Class	Worst case	Average	Best case
Bare soil	84	84	89
Riparian forest	98	99	100
Pasture	94	94	94
Water bodies	100	100	100
Dense savannah	100	100	100
Dense savannah in regeneration	100	100	100
Mean class recognition rate	96	96	97
Global recognition rate	95	95	95

The ability of the proposed multitemporal knowledge representation to capture long lasting spatial relationships may considerably simplify the usually troublesome task of spatial knowledge acquisition. In contrast, the class transition diagrams used in our experiments were set up as a result of a single interview with a photointerpreter.

Spatial knowledge can play a more important role when high resolution images are the objects of analysis. Assessing the relative contribution of multitemporal knowledge for high resolution image analysis is certainly a promising topic for future research.

The experiments reported in this paper successfully applied a Genetic Algorithm to estimate class transition possibilities. This technique has the advantage of accepting virtually any objective function, as well as any solution constraint. It has the disadvantage of being computationally intensive. Despite the ever increasing availability of processing power, an investigation towards a computationally more efficient optimization technique would be also welcome.

Acknowledgements

References

Blaschke, T., Strobl, J., 2001. What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. GIS — Zeitschrift für Geoinformationssysteme 14 (6), 12–17.

Bruzzone, L., Prieto, D.F., 2001. Unsupervised retraining of a maximum-likelihood classifier for the analysis of multitemporal remote-sensing images. IEEE Transactions on Geoscience and Remote Sensing 39 (2), 456–460.

Bruzzone, L., Prieto, D.F., 2002. A partially unsupervised approach to the automatic classification of multitemporal remote-sensing images. Pattern Recognition Letters 23 (9), 1063–1071.

Bruzzone, L., Prieto, D.F., Serpico, S.B., 1999. A neural-statistical approach to multitemporal and multisource remote-sensing image classification. IEEE Transactions on Geoscience and Remote Sensing 37 (3), 1350–1359.

Bückner, J., Jung, S., Pakzad, K., 1999. Image interpretation and GIS analysis as an approach for moor monitoring. Proceedings International Conference on Dynamic and Multi-Dimensional GIS (DMGIS 99), Peking, China, 4–6 October 1999, pp. 341–349.

Cazes, T.B., Feitosa, R.Q., Mota, G.L.A., 2004. Automatic selection of training samples for multitemporal image classification. Proceed-

ings International Conference on Image Analysis and Recognition (ICIAR 2004) - Part II, Porto, Portugal, 29 September–1 October 2004, pp. 389–398.

Clément, V., Giraudon, G., Houzelle, S., Sandakly, F., 2003. Interpretation of remotely sensed images in a context of multi-sensor fusion using a multispecialist architecture. IEEE Transactions on Geoscience and Remote Sensing 31 (4), 779–791.

Coutinho, L.M., 1978. O conceito de Cerrado. Revista Brasileira de Botânica 1 (1), 17–23.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society 39 (1), 1–38.

Epstein, R.L., 2000. Predicate Logic: The Semantic Foundations of Logic, second ed. Wadsworth Publishing Co., Belmont, CA.

Growe, S., 1999. Knowledge based interpretation of multisensor and multitemporal remote sensing images. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 32 (Part 7-4-3W6), 130–138.

Growe, S., 2001. Wissensbasierte Interpretation Multitemporaler Luftbilder. PhD Thesis, Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, University of Hannover, Hannover, Germany.

Growe, S., Schröder, T., Liedtke, C.E., 2000. Use of Bayesian networks as judgement calculus in a knowledge based image interpretation system. International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences 33 (Part B3/1), 342–350.

Jang, J.S.R., Sun, C.T., 1995. Neuro-fuzzy modeling and control. Proceedings of the IEEE 83 (3), 378–406.

Jeon, B., Landgrebe, D.A., 1992. Classification with spatio-temporal interpixel class dependency contexts. IEEE Transactions on Geoscience and Remote Sensing 30 (4), 663–672.

Jeon, B., Landgrebe, D.A., 1999. Decision fusion approach for multitemporal classification. IEEE Transactions on Geoscience and Remote Sensing 37 (3), 1227–1233.

Klir, G.J., Yuan, B., 1995. Fuzzy Sets and Fuzzy Logic — Theory and Applications. Prentice Hall Inc., Upper Saddle River, NJ.

Kuncheva, L.I., 2000. How good are fuzzy if–then classifiers? IEEE Transactions on Systems, Man and Cybernetics. Part B. Cybernetics 30 (4), 501–509.

Lee, T., Richards, J.A., Swain, P.H., 1987. Probabilistic and evidential approaches for multisource data analysis. IEEE Transactions on Geoscience and Remote Sensing GE-25 (3), 283–293.

Liedtke, C.E., Büchner, J., Pahl, M., Stahlhut, O., 2001. Knowledge based system for the interpretation of complex scenes. Automatic Extraction of Man-Made Objects from Aerial and Space Images (III). Balkema Publishers, Ascona, Switzerland.

Lu, D., Mausel, P., Brondízio, E., Moran, E., 2004. Change detection techniques. International Journal of Remote Sensing 25 (12), 2365–2407.

Melgani, F., Serpico, S.B., Vernazza, G., 2001. Fusion of multitemporal contextual information by neural networks for multisensor image classification. Proceedings Geoscience and Remote Sensing Symposium (IGARSS '01), Sydney, Australia, 9–13 July 2001, pp. 2952–2954.

Melgani, F., Serpico, S.B., Vernazza, G., 2003. Fusion of multitemporal contextual information by neural networks for multisensor remote sensing image classification. Integrated Computer-Aided Engineering 10 (1), 81–90.

Mendel, J.M., 1995. Fuzzy logic systems for engineering: a tutorial. Proceedings of the IEEE 83 (3), 345–377.

Minsky, M., 1974. A framework for representing knowledge. In: Winston, P. (Ed.), The Psychology of Computer Vision. McGraw-Hill, New York, pp. 211–277.

Mota, G.L.A., 2004. Interpretação Baseada em Conhecimento Aplicada a Imagens Multitemporais de Satélite de Baixa Resolução. PhD Thesis, Department of Electrical Engineering, Pontifical Catholic University, Rio de Janeiro, Brazil.

Pakzad, K., 2001. Wissensbasierte Interpretation von Vegetations-flächen aus Multitemporalen Fernerkundungsdaten. PhD Thesis, Institut für Theoretische Nachrichtentechnik und Informationsver-arbeitung, University of Hannover, Hannover, Germany.

Pakzad, K., 2002. Knowledge based multitemporal interpretation. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 34 (Part 3A), 234–239.

Pakzad, K., Mota, G.L.A., Meirelles, M.S.P., Coutinho, H.L., Feitosa, R.Q., 2003. Automatic interpretation of vegetation areas in Brazil. Proceedings Joint ISPRS/EARSeL Workshop — High Resolution Mapping from Space 2003, Hannover, Germany, 6–8 October 2003. 7 pp. (on CDROM).

Schmiedle, F., Drechsler, N., Grosse, D., Drechsler, R., 2002. Heuristic learning based on genetic programming. Genetic Programming and Evolvable Machines 3 (4), 363–388.

Shackelford, A.K., Davis, C.H., 2003. A hierarchical fuzzy classification approach for high resolution multispectral data over urban areas. IEEE Transactions on Geoscience and Remote Sensing 41 (9), 1920–1932.

Soille, P., 2003. Morphological Image Analysis Principles and Application, 2nd ed. Springer-Verlag, Berlin.

Solberg, A.H.S., Jain, A.K., Taxt, T., 1994. Multisource classification of remotely sensed data: fusion of Landsat TM and SAR images. IEEE Transactions on Geoscience and Remote Sensing 32 (4), 768–778.

Vincent, L., Soille, P., 1991. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Transactions on Pattern Analysis and Machine Intelligence 13 (6), 583–598.

Zadeh, L., 1978. Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems 1 (1), 3–28.