4
Evaluation of the proposed multitemporal classification method

The experiments described in the next sections were devised to evaluate the multitemporal classification method proposed in Chapter 3. The data set used in the experiments described in this chapter is basically the same as the one used in (Mota et al., 2007). In this document only the most important characteristics of the data set and of the preprocessing methods applied to it will be presented, further details can be found in (Mota et al., 2007).

Following the description of the data set, the monotemporal and multitemporal classifier designs, as well as the optimization procedures used for estimating the transition possibilities, will be described. To close the chapter, the results of the application of the method in several different experiments will be presented and commented on. Finally, a comparison of the results with those obtained through alternative cascade multitemporal classification methods will be made.

4.1.
Description of the data set

The test area is situated in the Municipality of Alcinópolis, in the State of Mato Grosso do Sul, Brazil. 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 and endangerous ecosystems in South America. It is covered by a single LANDSAT 7 scene (224-073).

Upper left corner		Lower right corner	
Latitude	Longitude	Latitude	Longitude
18º12’01,65”S	53º51’23,90”W	18º21’23,97”S	53º37’29,03”W

Table 1. Geographical limits of the images used in the experiments.

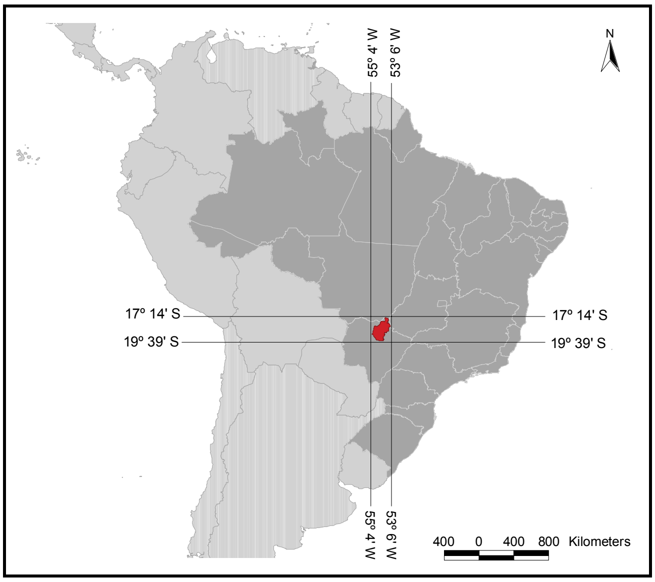

Figure 15. Taquari River sub-basin.

Three LANDSAT 7 images were acquired during the dry seasons of the particular region, respectively in 1999 (August 5), 2000 (August 7), and 2001 (August 10). The same subset of the LANDSAT 7 scene was extracted from each image, and those were the images subjected to classification in the experiments described below. Table 1 shows the geographical limits of the subsets, which cover an area of 459 km^2.

4.1.1.
Segmentation procedure

Bands 5, 4 and 3 of the three images were used in the experiments. The images were co-registered and then segmented producing spectrally homogeneous objects through the following steps:

a) The bands of all 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.

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).

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.

The procedure described above generated, therefore, the same set of segments for all images. Figure 16 shows the segmentation results, composed of a total of 18385 segments.

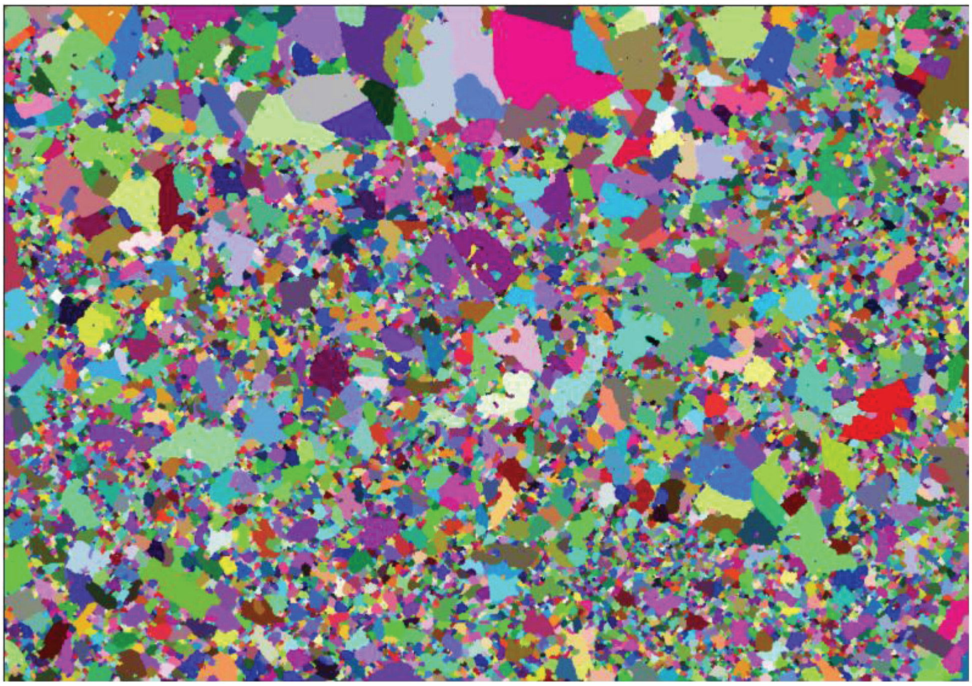

Figure 16. Segmentation procedure result.

4.1.2.
Validation data

To assess the performance of the method a reference classification for all three years was created by a human specialist, using as ancillary data a videography produced in October 2001, the LANDSAT images, a drainage map, and a digital elevation model.

The specialist classified a total of 442 segments in each of the three years. Table 2 contains a description of the land-cover classes considered in the experiments and the respective number of segments in each year. Figure 18 depicts the reference segments and their respective classifications for each of the three images.

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.	51	45	31
Riparian (ω_2)	Riparian forest	Dense woodland alongside rivers and streams.	33	32	30
Pasture (ω_3)	Pasture	Cultivated pasture for bovine cattle nutrition.	264	274	291
Water (ω_4)	Water bodies	Water or swampland.	15	15	15
Savannah (ω_5)	Dense savannah	Formation of low trees (8-12m) 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 – Brazilian Cerrado (Coutinho, 1978).	73	70	69
Regeneration (ω_6)	Dense savannah in regeneration	An area used previously for pasture that was left aside by the farmer, and is under regeneration of its native vegetation.	6	6	6

Table 2. Land cover classes considered in the experiments and number of reference segments in each year.

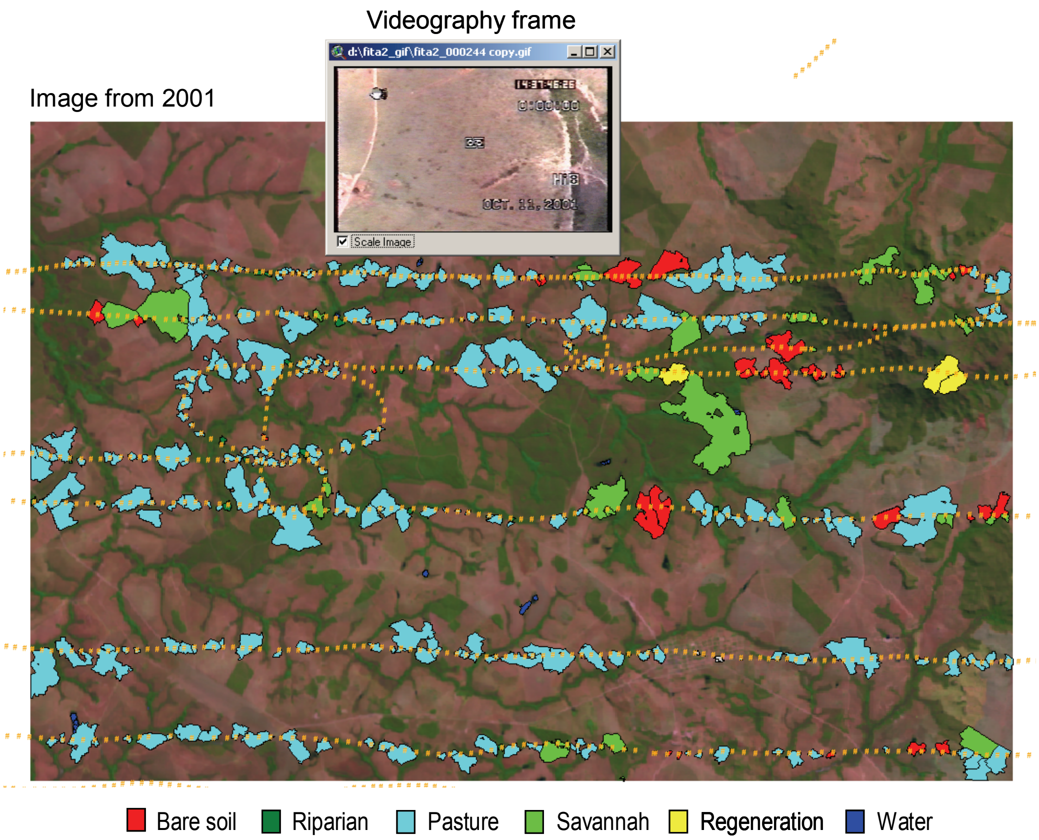

Figure 17. Reference segments and videography flight line. Adapted from (Mota, 2004).

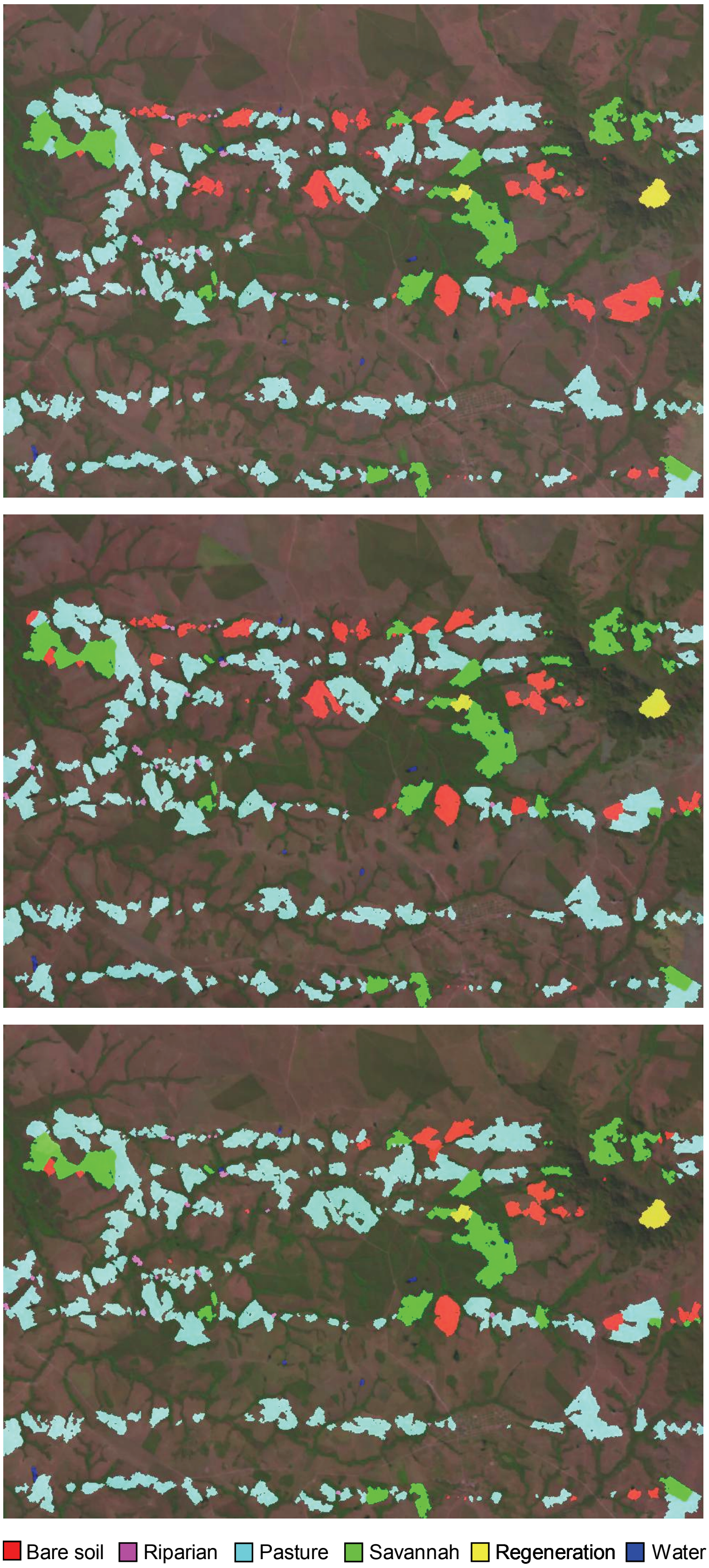

Figure 18. Classes assigned to the reference segments in each year: top 1999, center 2000, bottom 2001.

The image segments used as the reference classification are shown in Figure 17, colored according to the classes they were assigned to by the specialist for the year 2001. The dashed orange line in the figure represents the flight path of the videography. Tables 3 and 4 show the class transitions observed for each pair of years.

	2000					
1999	Bare soil	Riparian	Pasture	Water	Savannah	Regeneration
Bare soil	38		13			
Riparian	0	32	1			
Pasture	6		258			
Water				15		
Savannah	1		2		70	
Regeneration	0		0		0	6

Table 3. Class transitions from 1999 to 2000.

	2001					
2000	Bare soil	Riparian	Pasture	Water	Savannah	Regeneration
Bare soil	26		19			
Riparian	0	30	2			
Pasture	4		270			
Water				15		
Savannah	0		3		67	
Regeneration	0		0		0	6

Table 4. Class transitions from 2000 to 2001.

4.2.
Monotemporal classifier design

In the experiments the monotemporal classifiers for the earlier and later image (${}^{E}\mathbf{C}$ and ${}^{L}\mathbf{C}$) have the same design, originally proposed in (Mota et al., 2007). They consist of a fuzzy classifier that takes into consideration spectral features and spatial relations of the segments.

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 modeled by a normal distribution $N\left(\bar{\mathbf{x}}_{\omega i}, \mathbf{\Sigma}_{\omega i}\right)$. Hence, a Gaussian-shaped membership function $\mathbf{SMF}_{\omega i}(\mathbf{x})$ was defined for all classes ω_i, given by the formula below.

$$SMF_{\omega i}(\mathbf{x}) = \exp\left[-\frac{\left(\mathbf{x}-\bar{\mathbf{x}}_{\omega i}\right)^{T} \Sigma_{\omega i}^{-1}\left(\mathbf{x}-\bar{\mathbf{x}}_{\omega i}\right)}{2}\right] \tag{30}$$

For $\omega_i \in \{baresoil,\ riparian,\ pasture,\ water,\ savannah,\ regeneration\}$, $\bar{\mathbf{x}}_{\varpi i}$ and $\mathbf{\Sigma}_{\varpi i}$ 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 (reference segments).

An analysis of the crisp results (in which the class assigned to a segment is the one with the highest membership value) produced by the classification based solely on the spectral features revealed a great deal of confusion between riparian forest and dense savannah, as both classes present similar spectral responses. However, since riparian forest occurs alongside rivers and dense savannah in less humid places, far from the watercourses, the confusion can be minimized 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* was created, associated to a membership function based on the distance *d* separating the segment border and the closest water body, as shown in Figure 19.

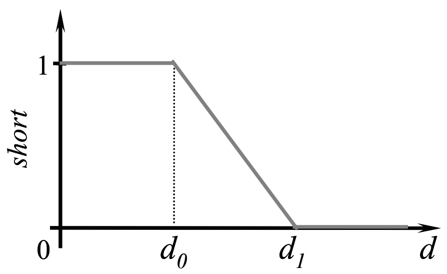

Figure 19. Membership function for fuzzy set short.

The values d_0 and d_1 in the *short* fuzzy set were defined empirically as d_0 =2.1 pixels and d_1=2.5 pixels. Further confusion was due to shaded areas near mountains or in dense savannah, spectrally similar to water bodies. This confusion can be attenuated by the use of elevation data. Thus, the crisp set *high* was created (Equation (31)), which is true if and only if the average elevation *e* of the image object is greater than a threshold *E*, defined by the photo interpreter.

$$high(e) = \begin{cases} 0, \text{if } e \leq E \\ 1, \text{if } e > E \end{cases} \tag{31}$$

The overall reasoning explained above is modeled by the fuzzy rules presented in Table 5. According to (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.

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

Table 5. Rule base of the monotemporal classifier.

Using the function *product*, *max* and $(1-x)$ to implement respectively the fuzzy *AND*, *OR* and *NOT* operators, the fuzzy classifier produces the fuzzy label vector $\boldsymbol{\alpha}=(\alpha_1, \ldots \alpha_6)^T$ according to the equations below.

$$\begin{aligned}
\alpha_1 &= SMF_{baresoil}(\mathbf{x}) \\
\alpha_2 &= max\ [SMF_{riparian}(\mathbf{x})\ ,\ SMF_{savanna}(\mathbf{x})] \cdot short\ (d) \\
\alpha_3 &= SMF_{pasture}(\mathbf{x}) \\
\alpha_4 &= max\ [SMF_{water}(\mathbf{x})\ ,\ SMF_{savanna}(\mathbf{x})] \cdot [1 - high\ (h)] \\
\alpha_5 &= max\ \{\ max\ [SMF_{ripparian}(\mathbf{x})\ ,\ SMF_{savanna}(\mathbf{x})] \cdot [1 - short\ (d)], \\
&\qquad max\ [SMF_{water}(\mathbf{x})\ ,\ SMF_{savanna}(\mathbf{x})] \cdot high\ (h)\ \} \\
\alpha_6 &= SMF_{regeneration}(\mathbf{x})
\end{aligned} \tag{32}$$

4.3. Multitemporal classifier design

The multitemporal model was built based on interviews with an agronomic engineer well acquainted with behavior concerning class changes along the time in the test area.

The possible class transitions within Δt=one year are depicted in Figure 20. Disregarding the impossible transitions, a total of 15 possible transitions still had to be estimated.

The transition possibility values were estimated by an analytical procedure and by a Genetic algorithm (GA), as described in Section 3.5.2, using the average class accuracy and the overall accuracy functions (see equations (24) and (25)) as the accuracy measures for the optimization procedures.

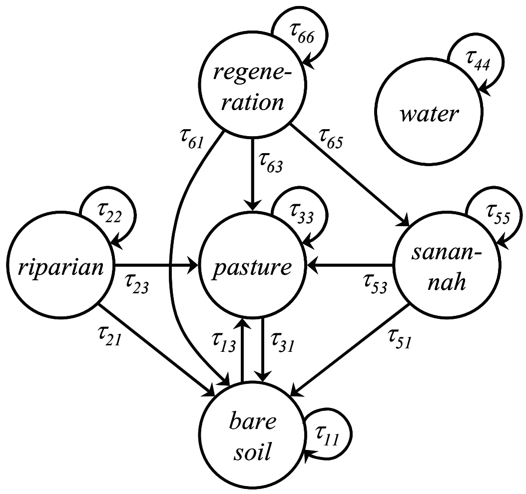

Figure 20. Class transition diagram for the test area.

4.4. Estimation procedures

The objective functions used for the GA were based on the accuracy functions introduced in Section 3.5.1. Each *individual* corresponds to the matrix **T** of transition possibilities, so that its *genes* correspond to the transition possibility values, τ_{ij}. The fitness of each individual was calculated according to the accuracy metric desired at each experiment – average class (Equation (24)) or overall (Equation (25)) – considering the experiment's training set (a number of reference segments chosen from two epochs). The most important GA options were set as:

- Initial population: all[7] transition possibilities to be estimated (genes) defined randomly.
- Population size: 20.

- Stop condition: up to 100 generations or 20 consecutive generations without improvement.
- Elite count: the best solution of each generation are kept to build the new one.
- Percentage of crossover: 80%.
- Percentage of mutation: 1%.

The GA with the configuration above was executed 5 times. The final population of each execution was used as initial population for the next one, and the best result through the 5 runs was taken as the final solution.

For the analytical procedure introduced in Section 3.5.2, the equation systems given by equations (28) and (29) were created by fixing the indexes i and j to the true class indexes and varying l and m for all possible class indexes, i.e. for all $(l,m) \neq (i,j)$, for all objects in the training set. As it was discussed in Section 3.5.2, the equation system based on Equation (28) was used when the performance metric was overall accuracy, and the system based on Equation (29) was used when considering average class accuracy.

The equation systems were solved by a nonlinear least-squares subspace trust-region method based on the interior-reflective Newton method described in (Coleman and Li, 1996).

4.5.
Experimental results

The benchmark for the analysis reported in the subsequent sections is the outcome of the monotemporal classifiers that take part of the multitemporal schemes depicted in Figures 11, 12, 28 and 29. As the object of comparison is the crisp classification of the later image segments, a defuzzification step was attached to the output of the fuzzy monotemporal classifiers ${}^{E}\mathbf{C}$ and ${}^{L}\mathbf{C}$ described in Section 4.2, as shown in Figure 21.

[7] The possibilities of forbidden transitions are set to zero.

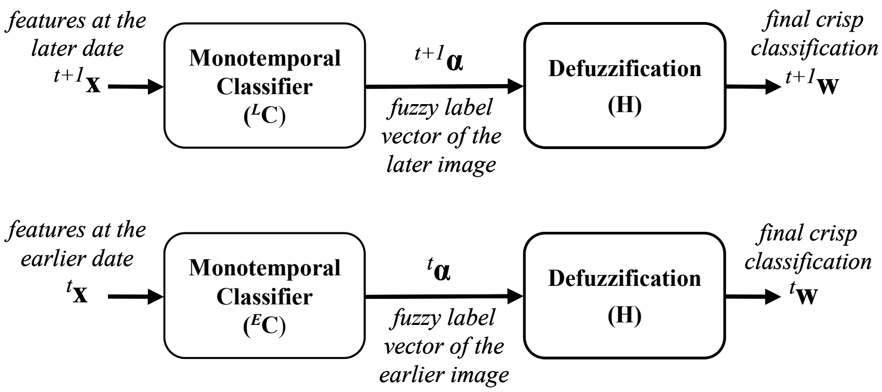

Figure 21. The monotemporal classifiers used as benchmark in the performance analysis.

4.5.1. Transition possibilities estimated on the average class accuracy

In a sequence of experiments using the average class accuracy to estimate transition possibilities, the training set was built in the following way. The reference image segments were first separated in groups according to the class transition they undergone in two different years. To estimate the parameters of the monotemporal classifiers as well as the transition possibilities as described in previous sections, around 50% of the objects in each group are *randomly selected* to form the training set. The remaining 50% of the objects were used to evaluate the method in terms of average class accuracy on both dates.

Table 6 shows the performance achieved by both the monotemporal and multitemporal methods for three pairs of dates. For the multitemporal model, both the analytic least-squares (LS) and the stochastic (GA) transition possibilities estimation procedures were used. The values in Table 6 are averages computed over 100 executions of the same experiment, each time with a distinct random selection of training and testing reference segments.

Note that these performances could be eventually improved if a special technique for selecting training samples were used, e.g. selecting the samples that best represent the classes of concern. But actually what the experiments try to show is the performance improvement brought by the multitemporal scheme in comparison to the monotemporal classifications.

Dates		Average Class Accuracy (%)					
Training/Test		Monotemporal		Multitemporal LS		Multitemporal GA	
Earlier	Later	Earlier	Later	Earlier	Later	Earlier	Later
1999	2000	57.7	55.0	63.4	64.4	63.4	64.4
2000	2001	54.0	54.4	66.7	66.9	66.4	65.8
1999	2001	54.1	56.4	67.9	69.9	64.6	65.4
Average		55.3	55.3	66.0	68.0	64.8	65.2

Table 6. Average class accuracy for both dates (averages over 100 experiments).

The results show that the outcome of the multitemporal method was very similar for both estimation procedures, with a slight advantage of the LS-based one. The multitemporal approach has, in average, improved the average class accuracy by approximately 12 percentage points[8] when using the analytic method for the estimation of the transition matrix, what amounts to a performance increase of around 21%. Considering the results of the experiments in which the GA-based procedure was used, the improvement was close to 10 points[8], with a performance gain of approximately 18%.

Similar experiments were carried out with the earlier monotemporal classifier replaced by an ideal classifier. Recall (see Section 3.1) that this corresponds to applications in which the classification of the earlier date is known, and this is fit into the classification model by setting the output of the earlier monotemporal classifier equal to the crisp label vector that represents the true class of the object at the earlier date, that is, ${}^{t}\boldsymbol{\alpha}={}^{t}\mathbf{W}$. Table 7 shows the results.

Dates		Average Class Accuracy (%)					
Training/Test		Monotemporal		Multitemporal LS		Multitemporal GA	
Earlier	Later	Earlier	Later	Earlier	Later	Earlier	Later
1999	2000	100	57.3	100	87.9	100	95.6
2000	2001	100	58.2	100	89.2	100	93.4
1999	2001	100	57.3	100	89.1	100	91.2
Average		100	57.6	100	88.7	100	93.4

Table 7. Average class accuracy for both dates (average over 100 experiments) using an ideal classifier at the earlier date.

In the experiments reported in Table 8 the time flow was reversed. The idea was to classify the earlier image using the reference information – the ideal classifier – for the later image, i.e. ${}^{t+1}\boldsymbol{\alpha}= {}^{t+1}\mathbf{W}$. In such experiments the temporal model represented by the class transition diagram depicted in Figure 20 was inverted with respect to time, which means that its links were reversed.

Dates		**Average Class Accuracy (%)**					
Training/Test		**Monotemporal**		**Multitemporal LS**		**Multitemporal GA**	
Earlier	**Later**	**Earlier**	**Later**	**Earlier**	**Later**	**Earlier**	**Later**
1999	2000	58.5	100	89.1	100	94.3	100
2000	2001	55.3	100	89.2	100	91.9	100
1999	2001	55.7	100	88.5	100	91.5	100
Average		56.5	100	88.9	100	92.6	100

Table 8. Average class accuracy for both dates (average over 100 experiments) using an ideal classifier at the later date.

In this case – using the true classification for one of the epochs – the multitemporal scheme brought a noteworthy performance improvement in comparison to the monotemporal classifier. The results were moderately different depending on the procedure used for the estimation of the transition possibilities, with advantage this time for the GA estimation method. Anyhow, the gain in performance by using the multitemporal approach with the transition possibilities estimated by the LS-based procedure was in average of 32 points[9], and of 36 points when the transition possibilities were estimated through the GA. The gain in performance was of approximately of 56% and 63%, respectively.

A direct comparison with the performance of the multitemporal method proposed in (Mota et al. 2007), which demands the knowledge of the true classes of the segments at a previous time, is not possible since the experiments in that work were designed in a different way, especially regarding the selection of training samples, however, the results in both studies are consistent.

One possible explanation for the advantage of the GA over the LS-based estimation method in this case has to do with the fact that while the search space

---

[8] Considering the average of the earlier and later image scores.
[9] Considering the average of the scores for all images (earlier or later) actually classified.

becomes smaller, which is in theory positive for both methods, the introduction of crisp values may create flat regions in the search space where partial derivatives of the objective function in relation to the transition possibilities being estimated are equal to zero. Gradient descent methods tend to get stuck in such regions. Furthermore, it must be noted that, differently from the LS method, the objective function of the GA was the exact same function used for the accuracy assessment (Equation 24). With that in mind, and observing the similarity among the results presented in Table 6, it seams that in that case both methods were able to find solutions very near to the optima.

Comparing the results of tables 6 to 8 one is inclined to the intuitive conclusion that the more accurate the earlier monotemporal classifier, the higher is its contribution to the multitemporal classification accuracy. Furthermore, cascade multitemporal methods are effective only if a correlation between the temporal data sets exists. The monotemporal classifiers of our cascade scheme may mask out or enhance this correlation. This explains the substantial performance difference observed between tables 6 and tables 7 and 8 for the multitemporal classifier.

Tables 9, 10 and 11 show the confusion matrices generated in these experiments for the 2001 image, using reference classifications from 1999 and 2001 and the LS-based method for transition matrix estimation. The rows represent the true classes and the columns the classes assigned by the classifiers. The values correspond to the average of all 100 experiments – that is why the entries are not integer numbers.

According to the confusion matrix for the monotemporal classifier (Table 9), the main confusion occurs for objects of class *bare soil*, which tend to be assigned to class *pasture*. The multitemporal approach manages to reduce partially this type of error, but with an increase in the omission and commission errors for *riparian* (Table 10). Moreover, the use of an ideal classifier for the earlier date improved the performance of the multitemporal method in terms of omission as well as of commission errors for all classes in relation to the monotemporal method (Table 11).

Classes	Bare Soil	Riparian	Pasture	Water	Savannah	Regeneration	*Omission*
Bare Soil	12.47	0	8.2	0	0.33	0	*40.6*
Riparian	0	10.33	0	0	0.67	0	*6.1*
Pasture	42.73	1.4	98	0.6	0.27	0	*31.5*
Water	0.2	0	0.07	6.67	0	0.07	*4.8*
Savannah	0.6	4.73	1.07	22.73	4.13	0.73	*87.8*
Regeneration	0.2	0	0	1	0.33	0.47	*76.7*
Commission	*77.8*	*37.2*	*8.7*	*78.5*	*27.9*	*63.2*	
Accuracy	*59.4*	*93.9*	*68.5*	*95.2*	*12.2*	*23.3*	

Table 9. Confusion matrix for the monotemporal classifier for the image from 2001.

Classes	Bare Soil	Riparian	Pasture	Water	Savannah	Regeneration	*Omission*
Bare Soil	12.6	0	8.07	0	0.33	0	*40.0*
Riparian	0	10.03	0.3	0	0.67	0	*8.8*
Pasture	41.6	1.93	98.33	0.07	1.07	0	*31.2*
Water	0.2	0	0.07	6.7	0.03	0	*4.3*
Savannah	0.4	5.13	0.47	0.07	27.33	0.6	*19.6*
Regeneration	0.07	0	0.07	0	1.4	0.47	*76.6*
Commission	*77.0*	*41.3*	*8.4*	*2.0*	*11.4*	*56.1*	
Accuracy	*60.0*	*91.2*	*68.8*	*95.7*	*80.4*	*23.4*	

Table 10. Confusion matrix for the multitemporal classifier for the image from 2001, using average class accuracy and the LS-based method for transition matrix estimation – prior classification not known.

Classes	Bare Soil	Riparian	Pasture	Water	Savannah	Regeneration	*Omission*
Bare Soil	13.07	0	7.6	0	0.33	0	*37.8*
Riparian	0	11	0	0	0	0	*0*
Pasture	38.67	0	103.67	0	0.67	0	*27.5*
Water	0	0	0	7	0	0	*0*
Savannah	0	0	0	0	34	0	*0*
Regeneration	0	0	0	0	0	2	*0*
Commission	*74.7*	*0*	*6.8*	*0*	*2.9*	*0*	
Accuracy	*62.2*	*100*	*72.5*	*100*	*100*	*100*	

Table 11. Confusion matrix for the multitemporal classifier for the image from 2001, using average class accuracy and the LS-based method for transition matrix estimation – prior classification known.

In this group of experiments the average class accuracy was used for the estimation of transition possibilities. Let's now consider how these estimates would affect the overall accuracy measured on the testing data. The confusion matrixes in tables 9 and 10 reveal that the multitemporal approach with a fuzzy earlier classifier improved the class accuracy for *bare soil*, *pasture*, *water*, *regeneration* and especially for *savannah* while for *riparian* the class accuracy worsened. In view of Equation (25) this implies that the multitemporal classifier could be in this case even inferior to the multitemporal counterpart in terms of overall accuracy, if most of the image objects in the data set belonged to *riparian*. This reasoning tries to demonstrate that the proposed multitemporal method may perform poorly if the transition possibilities are estimated upon a metric other than the one that actually expresses the intended classifier performance.

Figures 22, 23 and 24 provide a graphical representation of the results of some experiments over the images of 1999, 2000 and 2001, respectively (the selected cases attained performances above the average of the sequence of experiments). They show the misclassified segments by the monotemporal classifier, by the multitemporal classifier using a fuzzy earlier classifier, and by the multitemporal classifier using an ideal earlier classifier. The correctly classified segments are shown in white and the misclassified segments are shown in red. In those figures, the classifications produced using a fuzzy classifier for the earlier/later dates (center images) were carried out with transition matrixes estimated through the least squares-based method. Classifications produced with an ideal classifier for the earlier/later dates (bottom images) used transition matrixes estimated through the GA-based method. The classification accuracies of the selected experiments were similar to those shown in tables 6 to 8.

Figure 22. Typical results for the classification of the 1999 image: top, monotemporal; center multitemporal with fuzzy classifier for later date; bottom multitemporal with ideal classifier for later date (average class accuracies: 64%, 79% and 95%, respectively).

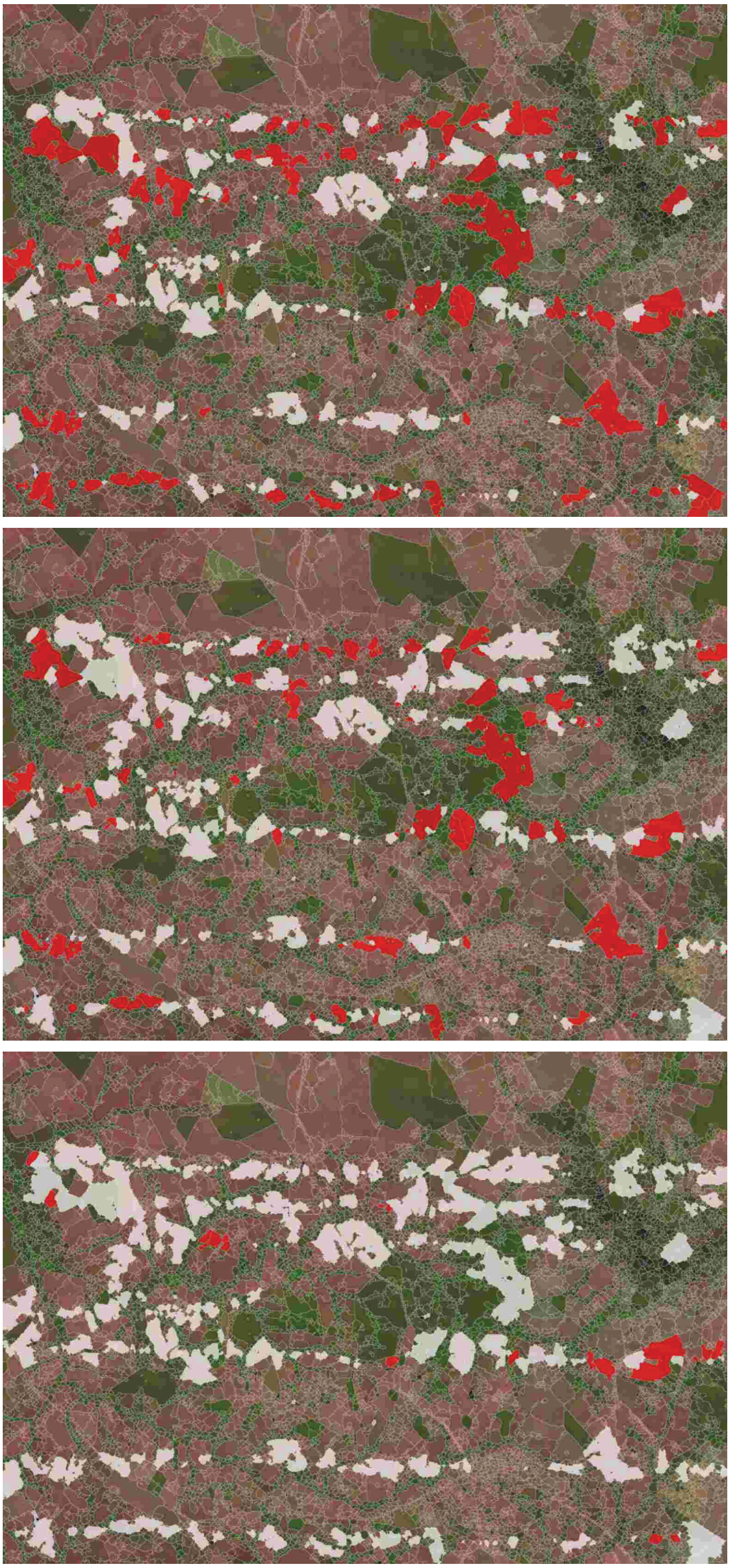

Figure 23. Typical results for the classification of the 2000 image: top, monotemporal; center multitemporal with fuzzy classifier for earlier date; bottom multitemporal with ideal classifier for earlier date (average class accuracies: 60%, 67% and 96%, respectively).

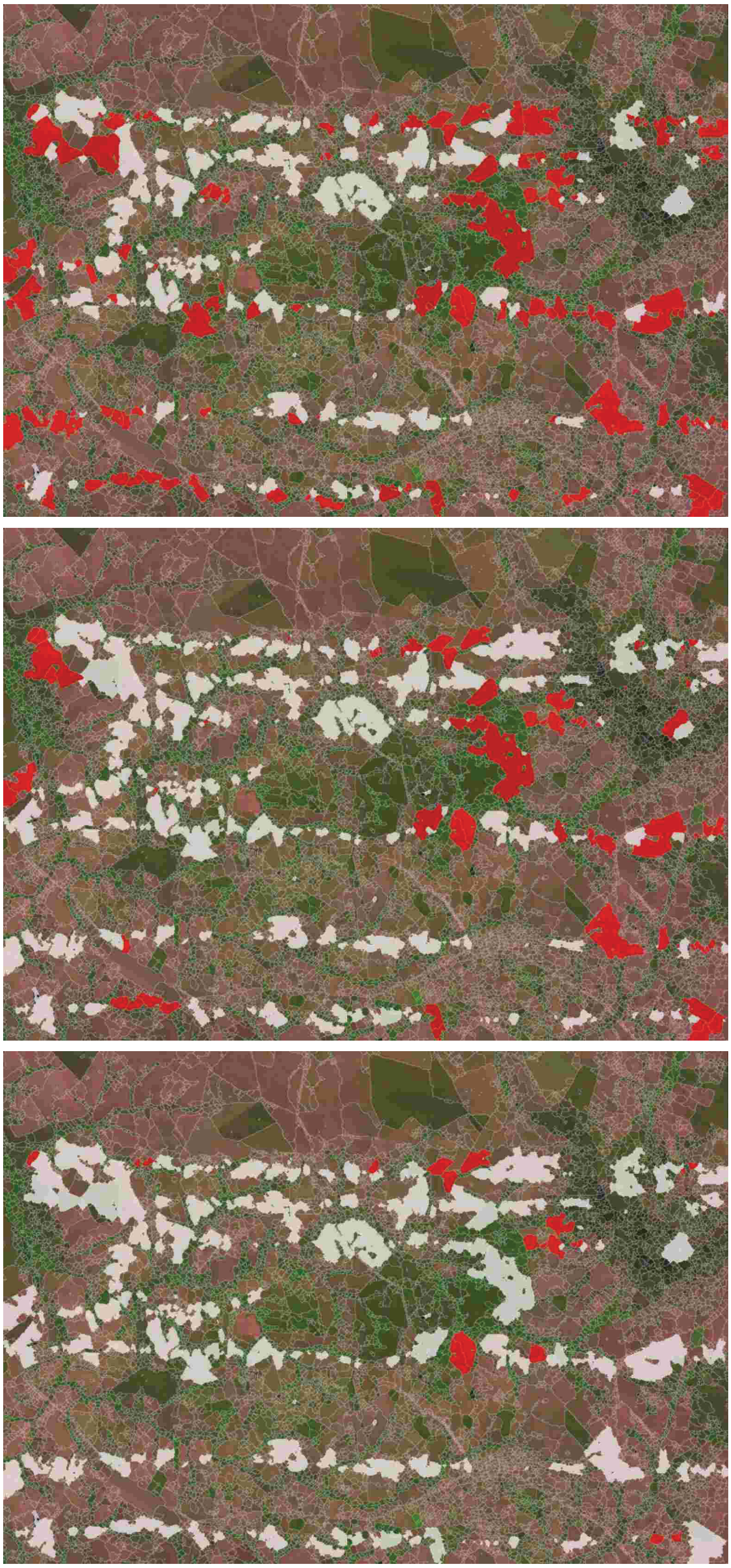

Figure 24. Typical results for the classification of the 2001 image: top, monotemporal; center multitemporal with fuzzy classifier for earlier date; bottom multitemporal with ideal classifier for earlier date (average class accuracies: 58%, 71% and 94%, respectively).

4.5.2.
Transition possibilities estimated on the overall accuracy

Another set of experiments has been conducted using the overall accuracy as performance metric. The purpose of these experiments was to assess how the proposed multitemporal method behaves when information about the distribution of objects among the classes is explored.

It was assumed that the frequency of classes and class transitions in the target area is nearly constant over time. This may be a reasonable assumption in some applications, especially if the images were acquired approximately in the same period of the year (as in our dataset), hence avoiding seasonal effects, and if the rate of change among the land cover classes does not accelerate abruptly. Thus, class frequencies and class transitions could be captured by estimating the transition possibilities upon all image objects of the same target area in an earlier (or later) pair of dates separated by the same time.

Two groups of experiments were devised: the first is about the classification of the pair of images from 2000 and 2001 using transition possibilities estimated from 1999 and 2000 reference data; in the second group, time flow was reversed, meaning that the 1999-2000 image pair of was classified using transition possibilities estimated from 2000-2001 data.

The segments were grouped according to their known classes at the pair of dates to be classified, i.e. 1999-2000 or 2000-2001. Approximately 50% of each group was selected in a random fashion to train the earlier/later monotemporal classifiers and to estimate transition possibilities. The performance was then measured on the remaining 50% segments of the respective pair of dates.

Table 12 shows the performance achieved by the monotemporal and multitemporal methods for two pairs of dates. For the multitemporal model, both the analytic least-squares (LS) and the stochastic (GA) transition possibilities estimation procedures were used. The values in Table 12 are averages computed over 100 executions of the same experiment, each time with a distinct random selection of training and testing reference segments.

The results show that the outcome of the multitemporal method was again very similar for both estimation procedures, with a slight advantage this time for the GA-based. Considering the classification of the 2000-2001 image pair, the multitemporal approach has in average improved the overall accuracy by approximately 19 percentage points[10], amounting to a 31% performance increase for both transition matrix estimation methods. In the case of the 1999-2000 pair classification, the improvement in performance was of approximately 17 percentage points[10] (31% increase) when using the LS-based estimation method, and of 20 points (36% increase) when using the GA-based method.

Dates				Overall Accuracy (%)					
Training		Test		Monotemporal		Multitemporal LS		Multitemporal GA	
Earlier	Later	Earlier	Later	Earlier	Later	Earlier	Later	Earlier	Later
1999	2000	2000	2001	62.0	61.1	78.9	83.1	80.0	82.1
2000	2001	1999	2000	52.3	58.8	72.0	73.3	74.1	77.6

Table 12. Overall accuracy for both dates (averages over 100 experiments).

The lower absolute results for the multitemporal classification of the 1999-2000 pair, when compared to the 2000-2001 classification, can be credited to the also lower performance of the monotemporal classification, especially for the 1999 image (52.3%). Nevertheless, this brings another evidence that a better monotemporal classifier accuracy will bring on a better multitemporal classification performance.

Once again, the slightly better results associated to the stochastic procedure could be attributed to the difficulty of the LS-based method in dealing with the zero valued partial derivatives, as mentioned in the last section, and also to the fact that the objective function of the GA was the exact same function used for accuracy assessment (Equation (25)).

Table 12 shows results that are considerably better than those reported in Table 6. This may be credited to the fact that in the experiments regarding overall accuracy not only the temporal correlation between the data sets have been captured by the transition possibility estimates, but also the relative frequency of classes and class transitions, as described in Section 3.5.1.

Table 13 shows the results of experiments in which the earlier or later monotemporal classifier was replaced by an ideal classifier.

[10] Considering the average of the earlier and later image scores.

Dates				Overall Accuracy (%)					
Training		Test		Monotemporal		Multitemporal LS		Multitemporal GA	
Earlier	Later	Earlier	Later	Earlier	Later	Earlier	Later	Earlier	Later
1999	2000	2000	2001	100	58.2	100	89.2	100	93.4
1999	2000	2000	2001	59	100	90.3	100	94.0	100
2000	2001	1999	2000	57.0	100	87.5	100	94.3	100

Table 13. Overall accuracy for both dates (average over 100 experiments) using an ideal classifier at the earlier or later dates.

Tables 14 and 15 show the confusion matrices generated for the 2001 image, using reference classifications from 1999 and 2000 and the LS-based method for transition matrix estimation. Again, the entries correspond to mean values obtained from 100 trials of the experiment.

Classes	Bare Soil	Riparian	Pasture	Water	Savannah	Regeneration	*Omission*
Bare Soil	9.33	0	26.33	0	0.33	0	*74.1*
Riparian	0	9.97	0.37	0	0.5	0.17	*9.4*
Pasture	19.37	2.47	104.97	0	0.53	0.67	*18*
Water	0.6	0	0.07	6.3	0	0.03	*10*
Savannah	0.03	4.7	0.6	0	26.3	2.37	*22.6*
Regeneration	0.03	0	0.1	0	1.73	0.13	*93.3*
Commission	*68.2*	*41.8*	*20.7*	*0*	*10.5*	*96*	
Accuracy	*25.9*	*90.6*	*82*	*90*	*77.4*	*6.7*	

Table 14. Confusion matrix for the multitemporal classifier for the image from 2001, using overall accuracy and the LS-based method for transition matrix estimation – prior classification not known.

Classes	Bare Soil	Riparian	Pasture	Water	Savannah	Regeneration	*Omission*
Bare Soil	15.63	0	5.37	0	0	0	*25.6*
Riparian	0	11	0	0	0	0	*0*
Pasture	20.3	0	122.7	0	0	0	*14.2*
Water	0	0	0	7	0	0	*0*
Savannah	0	0	0	0	34	0	*0*
Regeneration	0	0	0	0	0	2	*0*
Commission	*56.5*	*0*	*4.2*	*0*	*0*	*0*	
Accuracy	*74.4*	*100*	*85.8*	*100*	*100*	*100*	

Table 15. Confusion matrix for the multitemporal classifier for the image from 2001, using overall accuracy and the LS-based method for transition matrix estimation – prior classification known.

Most of the previous comments on the results in Tables 10 and 11 can also be made for Tables 14 and 15. However, some additional interesting conclusions may be drawn by comparing Tables 9, 10, 11, 14 and 15. Tables 14 and 15 shows that the multitemporal approach, for both types of earlier classifier, was successful in improving the class accuracy for *pasture* and *savannah*, with respect to the monotemporal classification (Table 9). Based on Tables 9 and 14, it can be seen that the multitemporal approach with the fuzzy earlier classifier substantially improved the class accuracy for *pasture* (from 68.5% to 82%). In contrast, the class accuracy dropped significantly for *bare soil* (from 59.4% to 25.9%). A different behavior is observed when comparing Tables 9 and 10. Class accuracy for *bare soil* show a modest improvement (from 59.4% to 60%), while the accuracy for *pasture* stays approximately the same (from 68.5% to 68.8%). This behavior can be understood in view of the particular accuracy function used to estimate the transition possibilities in either case. Note that *pasture* is the class with the highest number of objects in 1999 and in 2000, the dataset used to estimate the transition possibilities, and is approximately three to four times larger than *bare soil*. Thus, the class accuracy for *pasture* has an impact that is approximately three to four times greater than the class accuracy for *bare soil* in the computation of the overall accuracy, according to Equation (25). Hence, the transition possibility values were estimated so as to favor the class *pasture* and to penalize the class *bare soil*. This did not happen in the experiment reported in the previous section, when the average class accuracy was the selected performance metric.

In this group of experiments, the overall accuracy was used for the estimation of transition possibilities, but from the confusion matrices presented in Tables 9 and 14, it is also possible to calculate the average class accuracies. The monotemporal classifier reaches approximately 58.7% (see Table 9), while the multitemporal classifier achieves 62.1% (see Table 14) in terms of average class accuracy, considerably less than the 69.9% achieved when using average class accuracy as the estimation metrics (Table 10). This example reinforces the importance of using the same accuracy function for the estimation of transition possibilities that will be later used to measure the classifier performance.

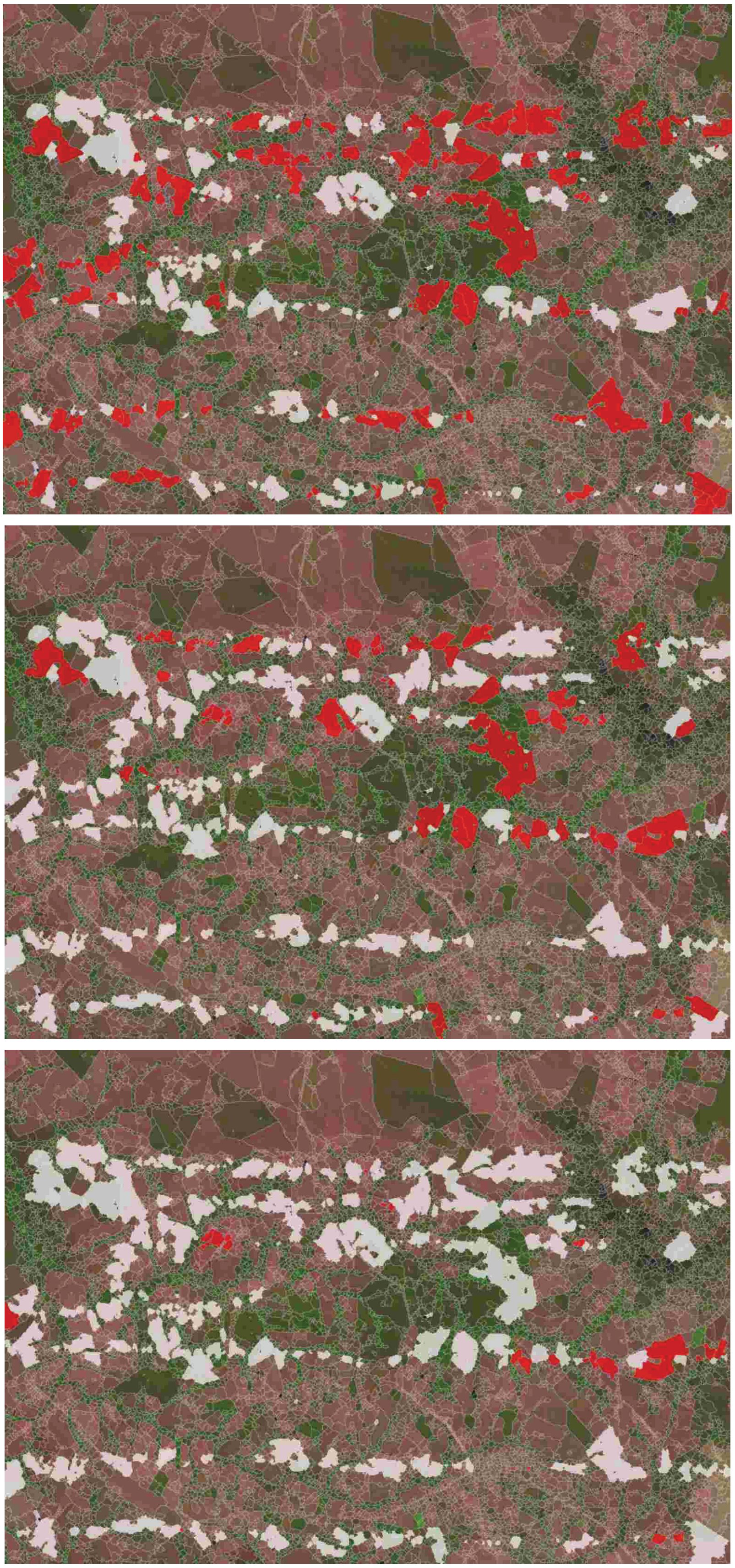

Figure 25. Typical results for the classification of the 1999 image: top, monotemporal; center multitemporal with fuzzy classifier for later date; bottom multitemporal with ideal classifier for later date (overall accuracies: 53%, 76% and 96%, respectively).

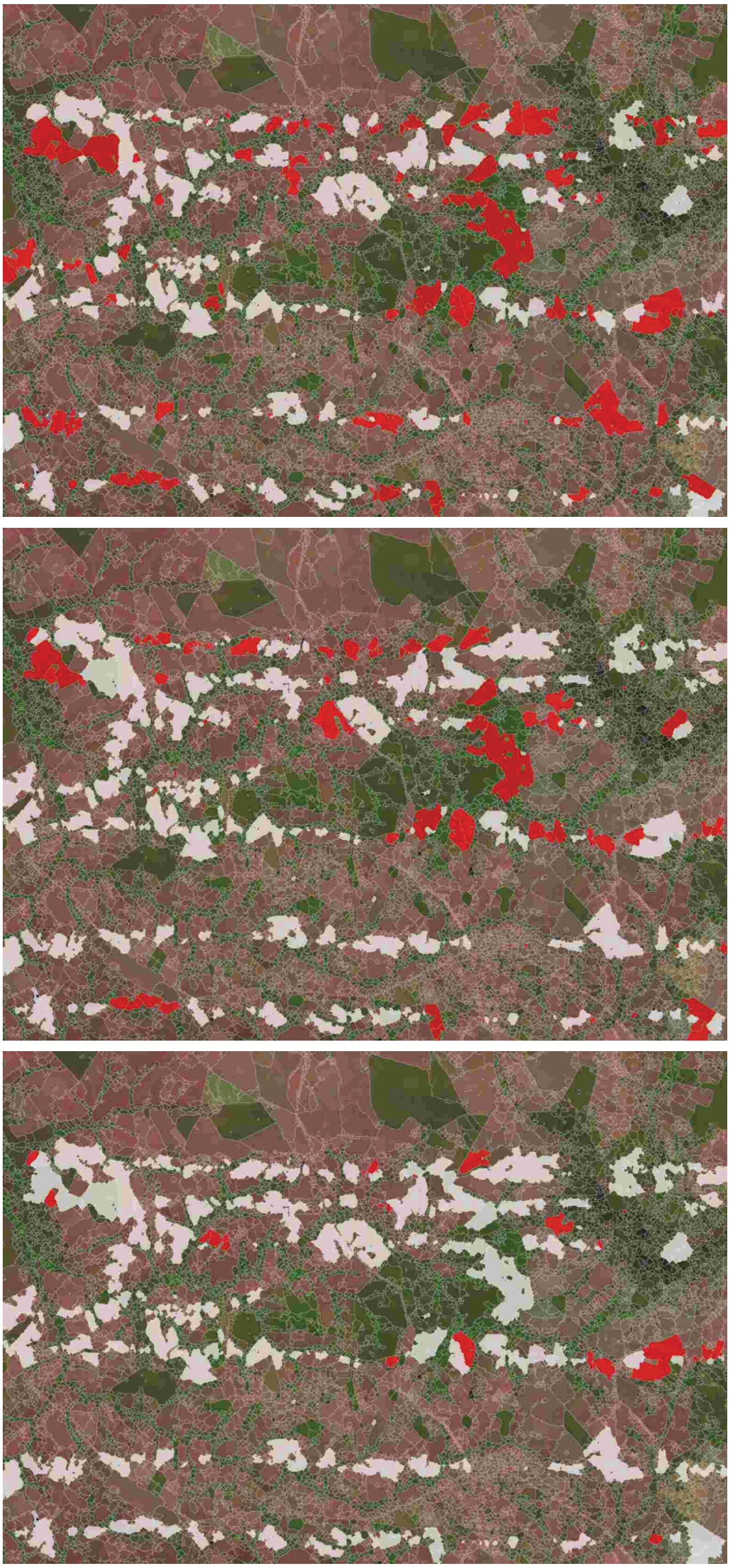

Figure 26. Typical results for the classification of the 2000 image: top, monotemporal; center multitemporal with fuzzy classifier for earlier date; bottom multitemporal with ideal classifier for earlier date (overall accuracies: 62%, 86% and 95%, respectively).

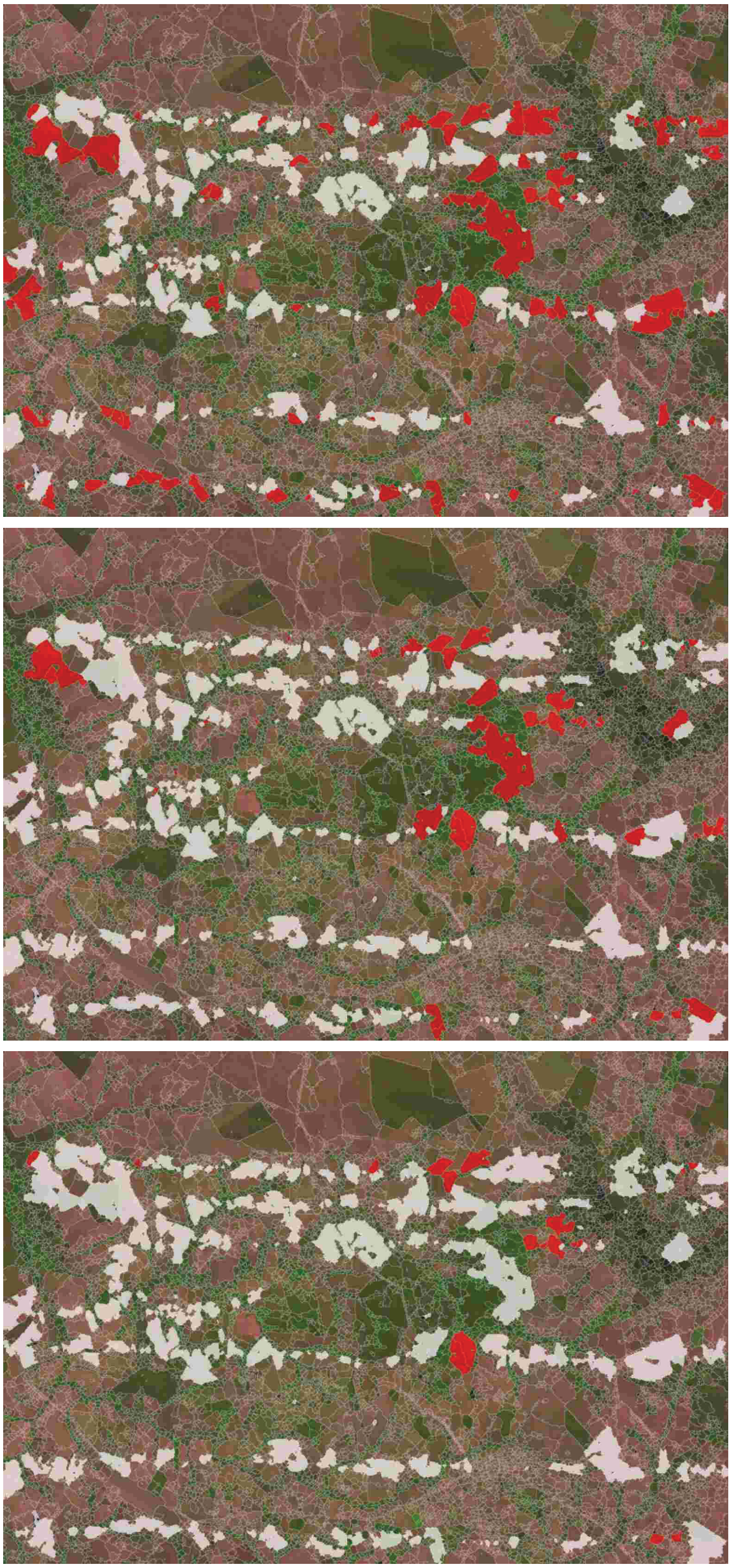

Figure 27. Typical results for the classification of the 2001 image: top, monotemporal; center multitemporal with fuzzy classifier for earlier date; bottom multitemporal with ideal classifier for earlier date (overall accuracies: 59%, 87% and 94%, respectively).

Figures 25 to 27 provide graphical results of this sequence of experiments over the images of 1999, 2000 and 2001, respectively. Misclassified segments are shown in red and the correctly classified ones in white. Transition matrixes were estimated through the least squares-based method (images on the center of the figures), and through the GA-based method (images on the bottom). Classification accuracies of the selected experiments were similar to those shown in tables 12 and 13.

4.5.3. Comparison to alternative approaches

Further experiments were carried out in order to compare the performance of the proposed method to other multitemporal cascade-classification methods found in the literature. Two structurally different decision fusion multitemporal approaches have been selected and implemented for that purpose. The comparison was based on the experimental procedure described in Section 4.5.1.

In the experiments reported hereafter, the true class at the earlier date is not known and average class accuracy is the target performance metric. Although a smaller performance gain with respect to the monotemporal classification was achieved when using that particular accuracy metric (see Sections 4.5.1 and 4.5.2), this is, in the view of the author, the procedure that most accurately assesses the main innovation brought by the proposed method, as it focuses on the benefits of exploring temporal knowledge (as opposed to exploring also class frequency distributions).

The first method selected for comparison was the *maximum jointly decision fusion multitemporal classifier* (TP-LIK) proposed in (Jeon and Landgrebe, 1999). The method is described schematically in Figure 28 for a two-date classification.

Let ${}^{t+1}\mathbf{w}_i$ be the crisp vector label associated to class of class ω_i at date $t+1$, i.e. a vector having "1" in the i-th position and "0" otherwise. For a two-date classification, the multitemporal classification outcome ${}^{t+1}\mathbf{w}$ is determined in the TP-LIK by the decision rule given by the following equation:

$$ {}^{t+1}\mathbf{w} = \max_{{}^{t+1}\mathbf{w}_i} \left\{ P\left({}^{t+1}\mathbf{w}_i\right) P\left(\mathbf{H}\left({}^{t}\boldsymbol{\alpha}\right) \middle| \, {}^{t+1}\mathbf{w}_i\right) P\left(\mathbf{H}\left({}^{t+1}\boldsymbol{\alpha}\right) \middle| \, {}^{t+1}\mathbf{w}_i\right) \right\} \qquad (33) $$

where $P({}^{t+1}\mathbf{w}_i)$ is the prior probability of class ω_i at date $t+1$, $\mathbf{H}({}^{t}\boldsymbol{\alpha})$ and $\mathbf{H}({}^{t+1}\boldsymbol{\alpha})$ are the output of the hardening function defined in Section 3.4 (Equation (15)) applied to the outcome, respectively, of the earlier and of the later monotemporal classifiers; and P=$(\mathbf{H}({}^{t}\boldsymbol{\alpha}) \mid {}^{t+1}\mathbf{w}_i)$ and P=$(\mathbf{H}({}^{t+1}\boldsymbol{\alpha}) \mid {}^{t+1}\mathbf{w}_i)$ are the probability of $\mathbf{H}({}^{t}\boldsymbol{\alpha})$ and $\mathbf{H}({}^{t+1}\boldsymbol{\alpha})$, given that the class at the later date is ω_i. All classifier parameters were estimated upon training samples selected as described in the first paragraph of Section 4.5.1.

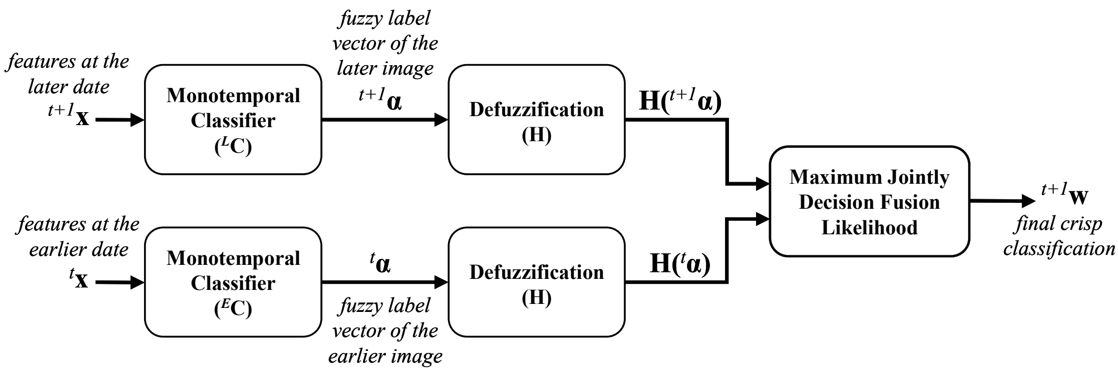

Figure 28. Maximum jointly decision fusion multitemporal classifier for two dates (TP-LIK).

In a number of related works, e.g. (Melgani et al., 2003) and (Bruzzone et al., 1999), decision fusion is performed using artificial neural networks. In order to consider this alternative and compare the proposed approach to a neural network-based method, a two-date multitemporal classifier was implemented, in which the outcomes of both multitemporal classifiers are combined by a neural network (Figure 29).

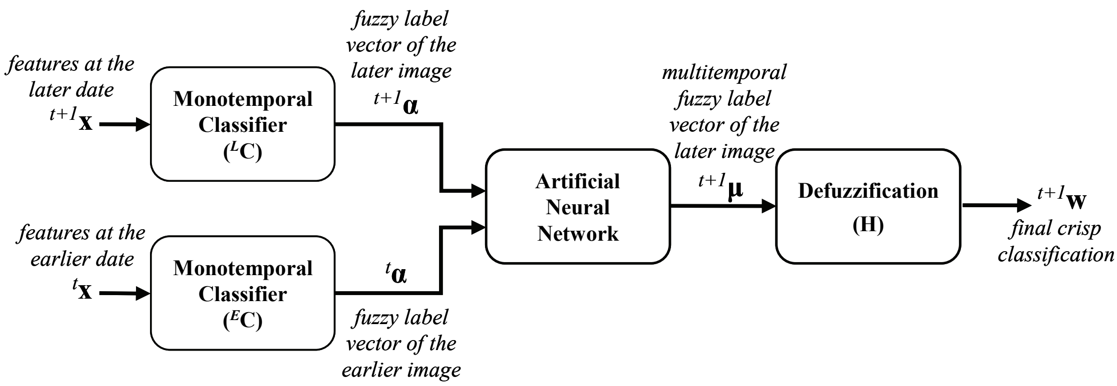

Figure 29. Two date classification fusion using a neural network.

This second alternative method is based on a feed-forward, back-propagation network with 12 inputs (the components of earlier and the later classifier outcomes – ${}^{t+1}\boldsymbol{\alpha}$ and ${}^{t}\boldsymbol{\alpha}$) and six outputs (the elements of multitemporal fuzzy label vector ${}^{t+1}\boldsymbol{\mu}$). The network has a single hidden layer implementing the hyperbolic tangent transfer function and the linear transfer function in the output layer. The Levenberg-Marquardt optimization (Marquardt, 1963) was used for training.

The training samples were replicated so that all classes were evenly represented in the training phase. Alternative configurations of the hidden layer were tested, considering 3 to 20 neurons. The configuration with nine hidden neurons provided the highest performance, and was used in the experiments reported henceforth. The training segments were separated in two groups: 70% to estimate network weights and 30% for validation (Haykin, 1994). The training procedure selected the set of neural network weight values that provided the best average performance in the validation set along 50 epochs. The best estimate was usually attained after 5 to 12 epochs.

The results obtained in these experiments are summarized in Table 16. Again, the performance that is reported are the averages computed over 100 executions of the same experiment, each time with a distinct random selection of training and testing objects. To facilitate comparison, Table 16 also contains the results of the monotemporal classification for the later date and of the proposed fuzzy Markov chain-based method (FCM).

Dates		**Average Class Accuracy (%)**				
Earlier	**Later**	**Monotemporal**	**Multitemporal**			
			TP-LIK	**NN**	**FMC (LS)**	**FMC (GA)**
1999	2000	55	58	56	64	64
2000	2001	54	55	59	67	66
1999	2001	56	56	59	69	65
Average		55	57	58	67	65

Table 16. Performance comparison of the monotemporal and three multitemporal approaches: the proposed method FCM with its two parameter estimation techniques LS and GA; the maximum jointly decision fusion likelihood (TP-LIK), and the neural network method (NN).

All three multitemporal methods were consistently better than the monotemporal counterpart, and the proposed fuzzy Markov chain based method achieved the highest and the TP-LIK the lowest performance among the multitemporal methods.

The superiority of the method proposed in this work may be understood in view of the available data set and of the complexity of each multitemporal method. In terms of number of parameters, the neural network based classifier is the most complex among the methods considered in the comparison. The TP-LIK classifier has as many parameters as the proposed method (11 parameters for the specific target application), and involves a much simpler training step.

The large number of parameters in the neural network based method enables it to learn complex relations, but makes it highly demanding in terms of training samples. The neural network configuration that delivered the highest performance in the experiments contains a total of 177 parameters, – the weights associated to the connections among neurons in adjacent layers. It is reasonable to believe that the neural network based method may eventually outperform our method provided that enough training samples are available. However, this may be impractical in most real applications, especially if segments instead of pixels are the objects being classified. The proposed method, as well as the maximum jointly decision fusion likelihood classifier, exploits prior knowledge about the possible class transitions resulting in a model with significantly less parameters than the neural network.

To close this section, a comparison was made to the work presented in (Feitosa et al., 2008), which introduces part of this research, namely the fuzzy Markov chain formulation, but only for the forward classification model (Equation (5)). As the main concern of that work was on the classification accuracy at the later date, the estimation technique used aimed at obtaining the highest accuracy for the training samples (reference segments) of later image. Hence, the fitness function of the GA-based estimation technique employed there, which has a similar configuration to the one described in Section 4.4, was designed to rate the alternative solutions – transition possibility matrixes – in terms only of the classification accuracy upon training samples from the later date.

It is possible however to obtain the classification of the earlier image from that method. Table 17 shows the results achieved with the same dataset (the one described in Section 4.1) for the monotemporal classifier and for the fuzzy Markov model using the transition matrix estimation technique employed at (Feitosa et al., 2008), denoted as GA-FW, and using the estimation procedures introduced in Section 4.4.

Dates		Average Class Accuracy (%)							
Training/Test		Monotemporal		FMC (GA-FW)		FMC (LS)		FMC (GA)	
Earlier	Later	Earlier	Later	Earlier	Later	Earlier	Later	Earlier	Later
1999	2000	58	55	62	64	63	64	63	64
2000	2001	54	54	62	66	67	67	66	66
1999	2001	54	56	63	64	68	69	65	65
Average		55	55	62	65	66	67	65	65

Table 17. Performance comparison of the monotemporal classification with the proposed FCM method using three different parameter estimation techniques: GA-FW, LS and GA.

Table 17 shows that the performance obtained from the three estimation techniques are practically the same for the later date. However, the procedures described in Section 4.4 produce for the earlier date better results, as it was expected.

The GA-FW estimation technique has one advantage though – it demands no knowledge of the true classes at the earlier date for the reference segments. Such knowledge however will be needed in practical applications for training the monotemporal classifier for the earlier image.