INPE-6636-PRE/2671

EVALUATION OF LABELLING AFTER THE SEGMENTATION PROCESS ON TM-LANDSAT IMAGES OVER THE AMAZON REGION

Elaine Cristina Cardoso Fidalgo
Thelma Krug
João Roberto dos Santos

National Institute for Space Research (INPE)
C.P. 515
12210-012 São José dos Campos, SP, Brazil

ISPRS Commission VII, Working Group WG3

KEY WORDS: Segmentation, Error Matrix, Deforestation.

ABSTRACT

Automatic extraction of information from orbital images constitutes an alternative to reduce the time and cost to map extensive areas. INPE and IBM/Rio are developing an algorithm that combines segmentation of images and classification based on neural network. The segmentation algorithm, based on region growth, was used here to stratify a selected area over the Amazon region. A label was independently assigned to each segment by two interpreters of similar photointerpretative capability. The objective of this work is to contrast the results obtained from these two labellings. The selected area, located in Rondonia State, was chosen due to its "fish bone" pattern of deforestation, which is a common form of land use in the Amazon. The segmentation of the pilot area generated 3048 segments. The following K classes were defined for this work: *forest, non-forest, water, deforestation, cloud,* and *shadow*. To each segment, a concept i (i = 1,2,3,4,5) was assigned to every one of the K classes, according to the probability of the classified segment belonging to that class (1, 0.75, 0.50, 0.25, 0, respectively). The contrast of the labelling by the two interpreters was based on error matrices. The results were first analysed for the concepts associated with each class K_i, and then with each predominant class K_j, where $j = 1$ if $i < 3$; $j = 3$ if $i = 3$; and $j = 5$ if $i > 3$. A high agreement between the interpreters (91%) was achieved at the level of the predominant class K_j in contrast to the poor agreement achieved at the concept level (at most 42%). The influence of the segment size and of the *interference* classes (*cloud and shadow*) on the results was also investigated.

1. INTRODUCTION

The deforestation of the Amazon region is monitored through analysis of the annual fraction of the area of natural vegetation that is affected, and its spatial distribution.

Digital extraction of information from orbital images constitutes an alternative to the conventional methods, reducing the time and the cost involved when mapping extensive areas.

The objective of this study is to evaluate the level of agreement between two interpreters in the labelling of segments in a TM-LANDSAT image over the Amazon region, through analysis of error matrices.

The results from this investigation give support to the application of an algorithm which combines image segmentation and classification based on neural networks, in images from the Amazon region. The algorithm is being developed jointly by IBM Scientific Centre in Rio de Janeiro and the National Institute for Space Research (INPE).

2. METHODS

2.1 Segments Labelling

This study was carried out using a colour composite TM-LANDSAT image in bands 3, 4, and 5, which was previously segmented applying an algorithm of region growing implemented in the image processing software developed at INPE (SPRING).

Image labelling was performed independently by two interpreters, resulting in two different final products.

The selected image corresponds to the path/row 231/68 quadrat X, from July 7, 1991, and comprises an area of 53,084.16 hectares. The area is located in the State of Rondonia, Brazil, between coordinates 11°00' and 12°00' Latitude South and 62°00' and 62°30' Longitude West. The deforestation pattern in the region is commonly known as *fish bone*, characteristic of deforestation along roads of access to agropecuary projects.

The labelling was based on a pre-defined legend of six thematic classes: *forest* (F), *non-forest* (NF), *water* (W), *deforestation* (D), *cloud* (C) and *shadow* (S).

The *deforestation* class corresponds to the areas of primary forest cover that have been subject to activities that promoted their elimination or temporary/permanent conversion to agricultural or pecuary purposes. This class also includes the selective exploitation of wood and minerals (Santos et al., 1993).

The *forest* class comprises all the formations characterized by dense biomass and by a significant and uniform arboreal cover, as defined by the RADAMBRASIL Project (Santos et al., 1993). By their classification scheme, the forest types in the area are the Tropical Open Ombrophilous Forest, and the areas of ecological tension (Savanna/Forest contact).

The *non-forest* class includes all the other natural vegetation formations in the region, which do not present an arboreal pattern, either phisionomically or structurally.

To each segment, a concept i was assigned to each of the K classes, according to the probability of the classified segment belonging to that class. The concepts i (i = 1,2,3,4,5) corresponded respectively to the probabilities 1, 0.75, 0.50, 0.25 and 0. Different probabilities were associated with each segment, for each of the distinct classes. Note that a segment could be associated with more than one class.

Each segment s had associated a notation of the type $B_s = F_\alpha NF_\beta W_\chi D_\delta C_\varepsilon S_\phi$. The indices α, β, χ, δ, ε, ϕ refer to the concept i which is associated with each of the defined classes. Whenever the associated concept was 5, the notation was simplified with deletion of the corresponding class. Hence, a segment with associated notation F_4D_2, had probability 0.25 of being *forest* and 0.75 of being *deforestation*.

The classes *forest*, *non-forest*, *water* and *deforestation* were defined as *basic* classes, since they provide relevant information to monitoring. *Cloud* and *shadow* were defined as *interference* classes (Barbosa et al., 1993). The accumulated probability for the *basic* classes, corresponding to the concepts assigned to each segment, should add 1, i. e., $Pr(F) + Pr(NF) + Pr(W) + Pr(D) = 1$. In the presence of *interference* classes, separate concepts were assigned to them and to the *basic* classes. For instance, in an area of forest, partially affected by the presence of a small but dense cloud, the segment corresponding to cloud would have concept 1 assigned to the *interference* class *cloud*, and concept 1 to the *basic* class *forest*, that is, $B_s = N_1F_1$.

Segment labelling was carried out by two interpreters of similar experience in photointerpretation and ground truth knowledge.

The photointerpretation process considered image characteristics such as colour, texture, shape, size, and location.

2.2 Analysis of the Products Resulting from Labelling

The products obtained by the two interpreters (interpreter I and interpreter II) were compared through the analysis of error matrices, which are usefull for classification accuracy estimation. In this case they show the number of sampling units - i.e., pixels, set of pixels, or polygons - associated with a given class in the classification process and the real class to which they belong (Story and Congalton, 1986; Congalton, 1991).

In the present study the error matrices express the number of pixels associated with each of the defined classes by interpreter I and by interpreter II.

The matrices were constructed to evaluate the level of agreement between the two products taking into account the following:

(a) all the segments in the study area;
(b) the segments comprising less than 100 pixels;
(c) the segments comprising at least 100 pixels;
(d) the segments where the *interference* classes (K = N or S) had associated a concept $i \geq 3$; and
(e) the segments where the *interference* classes (K = N or S) had associated a concept $i < 3$.

In (a), the global level of agreement between the two products was considered, independently of the size of the segment and the presence of *interference* classes.

In (b) and (c), the influence of the size of the segment in the level of agreement between the two products was evaluated. The threshold value of one hundred pixels was empirically chosen.

In (d) and (e), the comparison of products was carried out considering the presence of *interference* classes. These were considered present when either interpreter I or interpreter II assigned the concept $i \geq 3$ to either class *cloud* or *shadow*.

Two analysis were performed for each of the five cases. In the first one, the classes K_i were considered for all concepts i = 1 to 5. In the second one, only the predominant classes K_j were considered, assuming j = 1 if $i < 3$; j = 3 if i = 3; and j = 5 if $i > 3$. In this case, each segment had associated a notation of the type $B_s^* = F_\alpha \cdot NF_\beta \cdot W_\chi \cdot D_\delta \cdot C_\varepsilon \cdot S_\phi \cdot$, where the indices α^*, β^*, χ^*, δ^*, ε^*, ϕ^* refer to the concept j.

The level of agreement was defined as the ratio between the number of pixels which had been assigned the same label by both interpreters, and the total number of pixels in the image. Hence, it is a function of the area of the segments.

3. RESULTS

The segmentation process generated 3048 segments, 6 of which were eliminated from analysis for presenting concepts that represented a cumulative percentage different of one for the *basic* classes.

The error matrix in Table 1 presents the results from the labelling of the 3042 segments performed by the two interpreters.

Table 1 - Error matrix for the analysis of the classes and the concepts assigned (K_i), when the total number of segments (3042) was considered. The results are expressed in number of pixels.

		INTERPRETER II								
		F_1	F_2D_4	F_3D_3	F_4D_2	D_1	NF_1	F_3NF_3	A_1	TOTAL
I	F_1	100040	244820	5843	393	67	0	0	0	351163
N	F_2D_4	705	7694	15320	12051	775	0	0	0	36545
T	F_3D_3	0	0	780	9166	101	0	0	0	10047
E	F_4D_2	0	201	4967	67435	16255	0	0	0	88858
R	D_1	147	113	107	25286	74586	0	0	0	100239
P	NF_1	0	0	0	1735	249	0	0	0	1984
	F_3NF_3	0	0	0	0	27	0	0	0	27
I	A_1	0	0	21	0	31	0	0	0	52
	TOTAL	100892	252828	27038	116066	92091	0	0	0	588915

The level of agreement, computed as the ratio between the sum of the values in the diagonal of the error matrix and the total number of pixels in the image, was 42.48%.

The graphic representation of the results in Table 1 are shown in Figure 1. For simplicity sake, the results for NF_1, F_3NF_3 and A_1 (which corresponded to only 0.35% of the total area) were ommited.

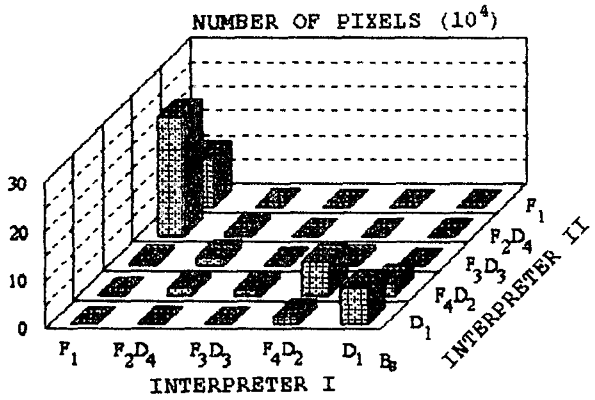


Fig. 1 - Graphic representation of the results (in number of pixels) obtained from the segments labelling by the two interpreters. All the segments, and the classes and concepts (K_i) were considered.

From Table 1 and Figure 1 it can be noted that the disagreements between the labels occurred in the following situations:

- 70% of the pixels labelled as F_1 by interpreter I were labelled as F_2D_4 by interpreter II;
- from the pixels labelled as F_2D_4 by interpreter I, 21% were labelled as F_2D_4, 42% as F_3D_3, and 33% as F_4D_2 by interpreter II;
- from the pixels labelled as F_3D_3 by interpreter I, 8% were labelled as F_3D_3, and 91% as F_4D_2 by interpreter II;
- from the pixels labelled as F_2D_4 by interpreter II, 97% were labelled as F_1, and 3% as F_2D_4 by interpreter I; and
- from the pixels labelled as F_3D_3 by interpreter II, 22% were labelled as F_1, 57% as F_2D_4, 3% as F_3D_3, and 18% as F_4D_2 by interpreter I.

The agreement between the interpreters reflects the importance of a prior definition of the contents of each class. The disagreement, at concept level, reflects the lack of a precise definition of the combination between classes and concepts.

Figure 2 shows a graphic representation of the results obtained in the analysis, considering only the predominant class K_i. In this case, the level of agreement between the interpreters was 91.15%, which is considerably higher than the level when all *basic* classes were considered.

The large number of disagreements observed among neighbouring concepts, within a same class (for instance, F_1 and F_2), and the low level of agreement when K_i was considered (specially when compared to the value obtained considering only K_j), suggest that there was an excessive subdivision of concepts.

Four graphic representations were designed, to visually show the effect of the size of the segment in the level of agreement between products (Figure 3 a, b, and Figure 4 a, b). Figure 3 presents the results from labelling only for those segments comprising less than 100 pixels. Figure 4 is similar, but involves only those segments comprising at least 100 pixels. In the figures indexed by (a), the classes and concepts (K_i) were considered, while in those indexed by (b) only the predominant classes K_j were investigated.

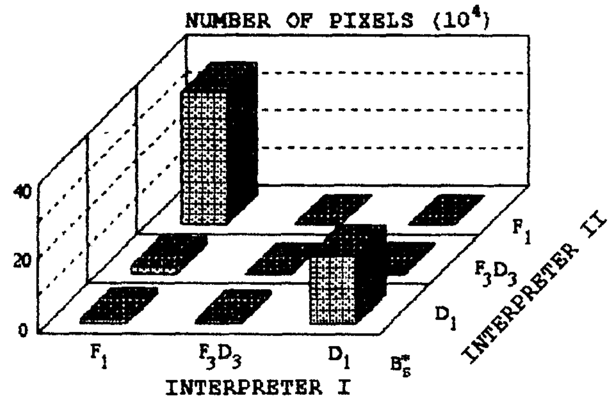
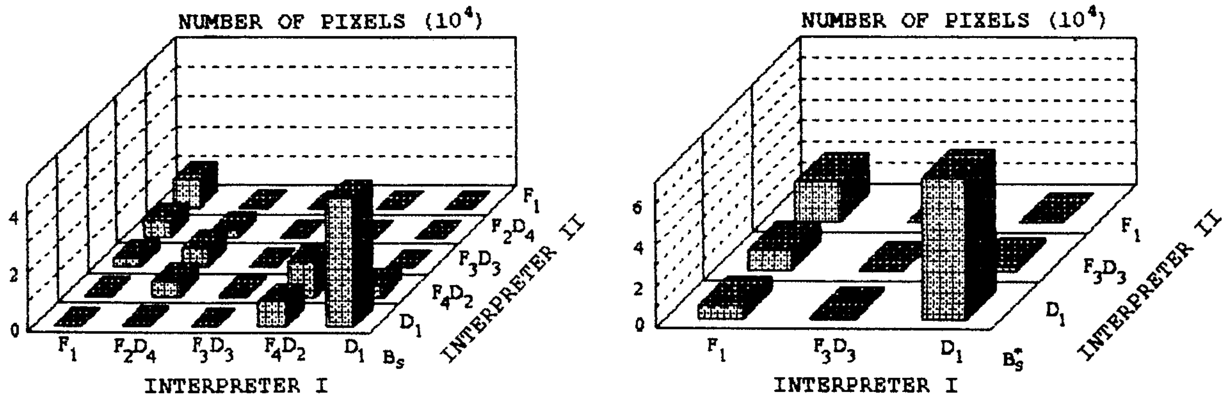
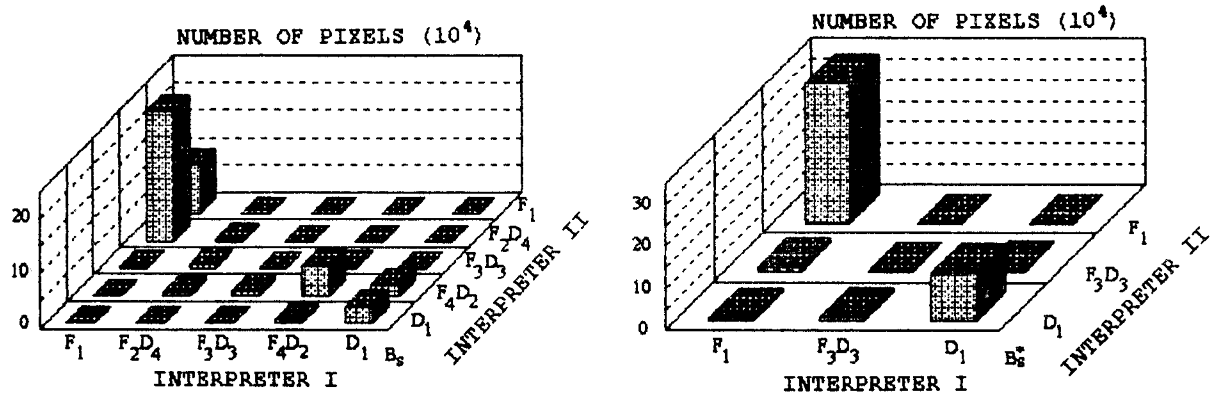


Fig. 2 - Graphic representation of the results (in number of pixels) obtained from the segments labelling by the two interpreters. All the segments, and the predominant classes (K_j) were considered.



(a) (b)

Fig. 3 - Graphic representation of the results (in number of pixels) obtained from the segments labelling by the two interpreters. Only segments comprising less than 100 pixels were considered. In (a), the results refer to the classes and concepts (K_i), whereas in (b) they refer to the predominant classes (K_j).



(a) (b)

Fig. 4 - Graphic representation of the results (in number of pixels) obtained from the segments labelling by the two interpreters. Only segments comprising at least 100 pixels were considered. In (a), the results refer to the classes and concepts (K_i), whereas in (b) they refer to the predominant classes (K_j).

The large number of disagreements between the two products occurred mainly for those segments with less than 100 pixels. The disagreements were of the following type:

- assignment of label F_1 by interpreter I and labels F_4D_2 or D_1 by interpreter II; and
- assignment of label D_1 by interpreter I to the segments labelled F_1 or F_2D_4 by interpreter II.

The levels of agreement computed for those segments that comprised less than 100 pixels, for K_i and K_j, were 64.34% and 81.85%, respectively. For those comprising at least 100 pixels, these levels were 37.73% and 93.37%, respectively.

When analysing the presence of *interference* classes, it was observed that only *shadow* was present. The total number of pixels with probability of at least 0.50 associated with the class *shadow*, in at least one of the products, was were low, corresponding to only 6% of the image area. In this case, the levels of agreement were 53.17% and 80.41%, for K_i and K_j, respectively. For those segments with probability less than 0.5 assigned to the class *shadow*, these results were 41.91% and 91.92%, for K_i and K_j, respectively.

The levels of agreement, considering K_i and K_j, are shown in Figure 5 for:
- all the 3042 segments,
- those segments comprising less than 100 pixels,
- those segments comprising at least 100 pixels,
- those segments with probability of at least 0.5 associated with the *interference* class, and
- those segment with probability less than 0.5 associated with the *interference* classes.

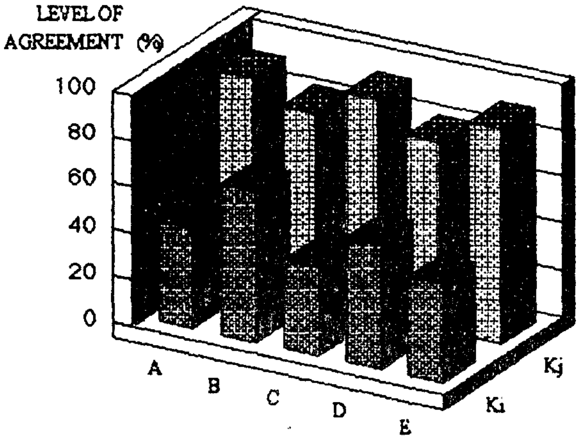


Fig. 5 - Level of agreement for K_i and K_j, considering: (A) all 3042 segments; (B) segments comprising less than 100 pixels; (C) segments comprising at least 100 pixels; (D) segments with probability of at least 0.5 associated with the *interference* classes; and (E) segments with probability less than 0.5 associated with the *interference* classes.

It can be noted from Figure 5 that the smallest level of agreement obtained for K_j was greater than the highest level obtained for K_i. The largest level of agreement was obtained for those segments comprising at least 100 pixels, considering K_j.

4. CONCLUSIONS

The level of agreement between the products obtained by the two interpreters was very low (42.48%) when considering the classes and the concepts assigned to them (K_i). The analysis at the level of the predominant classes (K_j) showed a considerable increase in the level of agreement (91,15%).

The lack of a precise definition of the composition of each combination between classes and concepts, and the excessive subdivision of the concepts can be pointed out as the factors that contributed to the low level of agreement between the two products, when K_i was considered.

The largest level of agreement was 93,37%, associated with the segments comprising at least 100 pixels, considering K_j.

The disagreements between the products predominated amongst concepts which were closer, for instance, F_1 e F_2.

The greater disagreements were observed for the segments comprising less than 100 pixels.

The results for different segment size stress the importance to establish a threshold based on the minimum size of the segment to be considered during the labelling process.

It is worth noting that the results from this study refer to an area in the Amazon region which has a specific land use pattern. The analysis of the results from labelling of other areas is relevant to evaluate the labelling effect from different interpreters in the classification result.

BIBLIOGRAPHIC REFERENCES

Barbosa, V. C.; Machado, R. J.; Liporace, F. S., 1993. **A neural system for deforestation monitoring on LANDSAT images of the Amazon Region**. Rio de Janeiro, IBM. (Technical Report, CCR-157).

Congalton, R. G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. **Remote Sensing of Environment**, 37(1), pp. 35-46.

Projeto RADAMBRASIL, 1978. Mapa fitoecológico, escala 1:1000000. In: _______ **Folha SC. 20 Porto Velho**; geologia, geomorfologia, pedologia, vegetação e uso potencial da terra. Rio de Janeiro. (Levantamento de Recursos Naturais, 16).

Santos, J. R.; Mello, E. M. K.; Moreira, J. C., 1993. **Mapping and estimation of the Amazonia deforestation**: state of the art and perspectives in the uses of remote sensing techniques. (Comunicação pessoal).

Story, M.; Congalton, R. G., 1986. Accuracy assessment: a user's perspective. **Photogrammetric Engineering and Remote Sensing**, 52(3), pp. 397-399.

Acknowledgments: We would like to express our gratitude to Adriano Venturieri for being one of the photointerpreters for this study, and to Leonardo Bins for his invaluable assistance in the use of the segmentation algorithm implemented at SPRING.