

Deforestation Monitoring of the Amazon Region Using Neural Networks - A Comparison Between Different Photo-Interpreters and Networks

Ricardo J. Machado, Valmir C. Barbosa,
Frederico dos S. Liporace, João R. dos Santos
and Adriano Venturieri.

Abstract— This paper deals with the task of evaluating the performance of neural networks designed for classifying satellite digital images with the purpose of monitoring the deforestation on the Amazon region. We are based on the Barbosa's previous work [1] that adopts a combination of image segmentation and classification techniques, the later employing a multi-layer perceptron that works on a fuzzy model of classification. The present study reports on experiments on five significant areas of the Amazon region whose satellite images were interpreted independently by two experienced photo-interpreters as well as by a set of different neural networks. The performances of the photo-interpreters and of the neural networks were evaluated on the basis of several performance indexes calculated over the same test set (including the Area Hit Ratio that is the most important for our practical application). This paper shows that different photo-interpreters do not fully agree when interpreting the same set of images, allowing us to establish reasonable performance target values to be sought by designers of automatic classifiers. Moreover, it is shown that from a practical point of view the neural networks we have built achieved area hit ratios compatible to those exhibited by different photo-interpreters. Suggestions for further improvements in the neural classifiers are discussed.

Keywords— Neural networks, remote sensing, fuzzy classification.

I. INTRODUCTION

The Amazon region is one of the major components of the planet's environment. Covering several million square kilometers, the region has the world's largest rain forest and river

system, playing a very important role in many global processes. Recent decades have witnessed a dramatic increase in human activity in the region, as deforestation, dam and road building, mining and agriculture, with potential effects on environmental stability.

Although of great relevance in the context of several global issues, the understanding of the extent to which human activity in the Amazon region can be harmful is rather superficial, particularly because of the lack of reliable information. The use of satellite imagery for nearly a decade has improved the situation somewhat, but the process of extracting significant information from the remotely sensed images is still rudimentary, representing a serious bottleneck, as it has to be repeated yearly for incremental monitoring.

The essential problem to be tackled is the yearly determination of the fraction of the Amazon region that has undergone deforestation, as well as of the locations where this process has been most pronounced. Currently, the extraction of information from the satellite images is to a large extent achieved manually, which not only is too costly but also renders some images virtually intractable. It is then quite desirable to automate considerably more of the entire process, and to this end various image classification techniques can be borrowed from the field of image analysis. Within the realm of remotely-sensed image classification, several automatic approaches have been suggested, often approaching the image on a pixel-by-pixel basis. Many of these techniques have a statistical nature [2], and some — including ours [3], [4], [5] — have recently employed neural networks [6], [7].

Ideally, we'd like to have available field information to train and measure the performance of automatic supervised classification systems, but unfortunately this is not always possible. Usually, the classification made by a experienced photo-interpreter based solely on the observation of satellite imagery is used for this purpose. Considering that the automatic classifier will never reach a 100% level of agreement with the photo-interpreter, the problem that is posed is to define which level of agreement should be considered as reasonable. We claim that the level of agreement exhibited between

Ricardo J. Machado and Frederico dos S. Liporace are with IBM Rio Scientific Center, Caixa postal 4624, 20001-970, Rio de Janeiro - RJ, Brazil

Valmir C. Barbosa is with Universidade Federal do Rio de Janeiro, Caixa postal 68511, 21945-970, Rio de Janeiro - RJ, Brazil.

João Roberto dos Santos and Adriano Venturieri are with INPE, Caixa Postal 515, 12227-010, São José dos Campos - SP, Brazil

two different photo-interpreters should be considered as a reasonable target to be achieved by an automatic classifier system. In this work we intend to explore how well our neural networks do in relation to this target.

In Section II we describe in details the architecture of the neural classifier system used in this work. In Section III we present experiments based on comparisons of the classifications provided by different photo-interpreters and neural networks on a set of significant images of the Amazon region. The concluding remarks are shown in Section IV.

II. ARCHITECTURE OF THE NEURAL CLASSIFIER

Our approach to image classification [3], [4], [5] proceeds in two phases. The first phase, contrasting with the usual pixel-by-pixel approaches, consists in the segmentation of the image into spectrally homogeneous portions (called *segments*, generated by a region-growing technique [8]). In the second phase, each segment is then classified into one or more of the thematic categories Forest (F), Savanna (S), Water (W), Deforested area (D), Cloud (C), and Shadow (Sh). The categories F, S, W, and D embody all the relevant information to be monitored, and are referred to as *basic categories*. The remaining categories, C and Sh, called *interfering categories*, are needed to account for the interference caused by clouds and shaded areas in the classification process.

The classification of segments into these categories follows a fuzzy-logic approach, that is, a segment may belong to multiple categories with partial degrees of membership, in what we call a fuzzy classification scheme. This fuzzy classification approach allows to represent phenomena like the transition between two basic classes, such as the recovery of forest in a area that was previously deforested but abandoned later, and the presence of interference like shadows and clouds that still permit the identification of the basic category of the segment, such as the presence of a transparent and thin cloud over a region of forest. The membership values of the segments in each category may vary in the interval $[0, 1]$, with 0 indicating empty and 1 indicating full membership of the segment in each category. Table I shows some examples

Situation	F	S	W	D	C	Sh
Deforested area	0	0	0	1	0	0
Deforested area with incipient reforestation	.25	0	0	.75	0	0
Cloud (opaque)	0	0	0	0	1	0
Tenuous shadow over forest	1	0	0	0	0	.5
Anomalous segment (Forest and Savanna)	1	1	0	0	0	0

TABLE I
EXAMPLES OF VALID CLASSIFICATIONS FOR A
SEGMENT

of fuzzy classifications to clarify this idea. It should be noted that there is a possibility of the generation of segments that covers more than one defined category. These *anomalous* segments are allowed to have in their classification full membership in more than one basic category, being that the ones covered by the segment.

The architecture of the neural classifier used in this work is a subset of the one defined in [4]. Our classifier system is centered around a neural network trained by a variation of the backpropagation algorithm [9]. Each segment is presented to the neural network for classification as a collection of features, which describe spectral and textural characteristics of the segment.

The spectral features of a segment are the gray-level averages, taken over all the pixels of the segment, in each of the six Landsat TM bands (excluding the thermal band) [10]. The textural features for each TM band are of two types, one based on the gray-level variance, and the other encompassing two functions of the co-occurrence matrix for the segment, namely the entropy and correlation [8]. These descriptors were selected from a set of about 100 spectral, textural and geometric defined segment's features, based on its contribution to the classifier performance [4], [5].

The neural networks we built for this application have one single layer of hidden neurons. Each network is structured as six independent modules, one for each category, sharing the segment-feature inputs. Each module is specialized in detecting one specific category and rejecting the others. This arrangement into

independent modules appears to be generally preferable for multiple-classification problems [11], and this was in our case demonstrated by early experimentation [12].

The previous mentioned work [4] upon which we are based employed also contextual information as a way of improving the classification when we deal with the defined interference categories *cloud* and *shadow*. This contextual information is represented by a set of *neighborhood* descriptors calculated for each segment, that took into account the fuzzy classification of the segment's neighbors.

III. EXPERIMENTATION AND RESULTS

To achieve the goals of this paper we first built a database of significant images of the Amazon region. Two different photo-interpreters classified independently such images and the comparison of these two interpretations allowed us to determine the target level of performance to be achieved by automatic classifiers. A set of different neural networks was then built and their performances were compared with this target value.

Five representative images of the Amazon region, chosen for containing a great variety of the possible situations, including many of great complexity, were used to build the database. These five images come from the states of Acre, Mato Grosso, Pará, and Rondônia, and from around Tucuruí dam. Approximately seventeen thousand segments were generated by the segmentation procedure, and these were divided into roughly two thirds for training and the remaining third for testing of the neural networks. Special care was taken to guarantee uniform distribution of the larger segments between the training and testing sets to avoid distortions on the area-based performance indexes defined later in this Section.

The segments were classified by the previously mentioned photo-interpreters, named **A** and **B** into the six described categories, using only five possibilities for the degrees of membership in each one of them (0, .25, .5, .75, and 1).¹

We trained two different networks, named **n_A**

¹Note that the use of these discrete degrees of membership accounts for far more than the six categories, allowing in fact for nearly six hundred compound categories.

and **n_B**. The first network was trained using only the classification provided by **A**, and the second with the classification provided by **B**.

Each of the six independent modules that constitute each network was trained separately, using an extension of the training set obtained for that module by the unbiased replication of some segments. This replication intends to uniform the uneven distribution of segments concurring toward and against establishing the membership in the category of the module being trained. Additionally, in the case of the D module, examples of S were replicated with more intensity, as this is the category that generates most of the confusion and it was poorly represented in the set of examples concurring against the category D. This was not necessary conversely in the training of the S module, because D examples were naturally well represented in the unbalanced training set. This procedure is described in details in [12].

A. Definition of performance measures

We define a set of performance indexes to quantify the excellence of the classification provided either by a neural network or a photo-interpreter when compared to the interpretation of one photo-interpreter taken as standard. The used set is composed of five indexes, namely the *Mean Squared Error (MSE)*, the *Area Mean Squared Error (AMSE)*, the *Jacquard index (JQ)*, a *Hit Ratio (HR)* and an *Area Hit Ratio (AHR)*. It's convenient for the definition of each one of these indexes, made in the following subsections, consider the classification of each segment S^μ belonging to the test set $S = \{S^1, S^2, \dots, S^M\}$ as a tuple $\mathbf{V}^\mu = (V_1^\mu, V_2^\mu, \dots, V_6^\mu)$ with each component representing the S^μ membership degree in each of the defined classes F, S, W, D, C and Sh. Let \mathbf{X}^μ be a classification for the segment S^μ provided either by a photo-interpreter or by a neural network whose performance we intend to measure, and \mathbf{Z}^μ be the classification of one photo-interpreter taken as standard. We may then define the following indexes:

A.1. Mean Squared Error (MSE). The Mean Squared Error measured considering one photo-interpreter's classification as standard is given

by

$$MSE = \frac{1}{M} \sum_{\mu=1}^M \frac{1}{2} \sum_{i=1}^6 [X_i^\mu - Z_i^\mu]^2. \quad (1)$$

This is the value that the training process of the neural network tries to minimize, being for that reason a natural performance index. This index is however very sensitive to the distribution of the classes in \mathcal{S} and gives the same importance to larger and smaller segments, since the areas of the segments are not considered.

A.2. Area Mean Squared Error (AMSE). This index tries to take into account the areas of the segments by implementing a weighted average of the MSE, giving a higher importance to segments of larger areas. It is expressed as

$$AMSE = \frac{1}{\sum_{\mu=1}^M A^\mu} \sum_{\mu=1}^M \frac{A^\mu}{2} \sum_{i=1}^6 [X_i^\mu - Z_i^\mu]^2 \quad (2)$$

where A^μ denotes the area of a segment S^μ . We may verify if the neural network errors are more significant in large or small segments by the comparison of the MSE and AMSE indexes.

A.3. Jacquard index. The Jacquard index between two classifications for a segment S^μ is given by

$$\mathcal{J}(S^\mu) = \frac{\sum_{i=1}^4 \text{smaller}(X_i^\mu, Z_i^\mu)}{\sum_{i=1}^4 \text{greater}(X_i^\mu, Z_i^\mu)} \quad (3)$$

where **smaller**(x, y) returns the smaller number between x and y and **greater**(x, y) works conversely. The Jacquard index is measured considering only the basic categories, as it may be observed by the limits of the summations in its definition.

The Jacquard index for a set \mathcal{S} of segments is defined as the average of the indexes calculated for each segment $S^\mu \in \mathcal{S}$. This index has the advantage of being bounded in the 0–1 range, with 1 representing maximum agreement.

A.4. Hit ratio (HR). The hit ratio depends on the definition of what a *correct* classification is, which is far from consensual in view of the fuzzy character of our classification scheme. In this paper we take a classification \mathbf{X}^μ of a segment S^μ to be correct or not, relative to a

threshold τ , $0 \leq \tau < 1$, according to the outcome of the following simple steps. First mark the categories corresponding to those components of \mathbf{Z}^μ which are in excess of (i.e., greater than) τ . Let N be the number of marked categories. If $N = 0$, then say that \mathbf{X}^μ is correct if none of its components is in excess of τ either. If $N > 0$, then check whether the categories corresponding to the N largest components of \mathbf{X}^μ are the same as the marked categories. Say that \mathbf{X}^μ is correct in the affirmative case. The classification \mathbf{X}^μ is said not to be correct in all other situations. The hit ratio for a set of segments \mathcal{S} is then defined to be the fraction of \mathcal{S} corresponding to segments $S^\mu \in \mathcal{S}$ such that \mathbf{X}^μ is correct.

The definition of a correct classification can be changed slightly so that only the components belonging to a certain group of categories is looked at during the steps we just outlined. This allows us to treat, in the context of this paper, the correctness of a classification with respect to the group of basic categories, regardless of the interfering categories.

A.5. Area Hit Ratio (AHR). The AHR is defined as the fraction of the area from \mathcal{S} corresponding to segments $S^\mu \in \mathcal{S}$ such that \mathbf{X}^μ is correct using the same criteria described in the previous subsection. It should be noted again that only the basic categories are considered.

B. Levels of agreement between different photo-interpreters

Table II summarizes the performances of the photo-interpreters **A** and **B** as well as the trained networks **n_A** and **n_B** using both photo-interpreters as standards. Each line displays the indexes obtained by photo-interpreters and networks when compared to the classification given by the column's photo-interpreter. For instance, the first two lines of Table II show the degree of agreement between the different photo-interpreters **A** and **B**.

From Table II we conclude that in fact there is no full agreement between different photo-interpreters. Hence, it is not reasonable to expect from an automatic classifier a full agreement with neither of the photo-interpreters, as that would mean building a classifier that would be able to reproduce even the photo-

	A					B				
	MSE	AMSE	JQ.	HR	AHR	MSE	AMSE	JQ.	HR	AHR
A	xxx	xxx	xxx	xxx	xxx	0.104	0.050	0.883	87.42	95.64
B	0.104	0.050	0.883	90.95	96.36	xxx	xxx	xxx	xxx	xxx
n_A	0.155	0.152	0.783	86.90	94.07	0.158	0.153	0.781	83.58	91.83
n_B	0.183	0.166	0.774	86.74	94.24	0.158	0.145	0.784	85.23	95.38

TABLE II
EVALUATION OF PHOTO-INTERPRETERS AND NEURAL NETWORKS CLASSIFICATIONS AGAINST **A** AND **B**

interpreter's errors.

Reasonable performance targets to be sought when building automatic classifiers are then:

- MSE: 0.104
- AMSE: 0.050
- Jacquard index: 0.883
- Hit Ratio: 87.42%
- Area Hit Ratio: 95.84%

We may observe that the indexes that takes into account the areas of the segments (AHR and AMSE) present better performance values than its counterparts based solely on the number of segments. This indicates that most mismatches have occurred among segments of small area, an expected behavior since among small segments we frequently find transitions between the defined categories and occurrence of categories that are more difficult to discriminate, such as S and D.

The hit ratios indexes measured between photo-interpreters differ as we take **A** or **B** as standard. This phenomenon is explained by the manner which the hit ratio indexes were defined, as well as by the presence of anomalous segments (Section II) in the database.

Concerning the neural networks, it is not a surprise to verify that both **n_A** and **n_B** exhibit better performance when tested against the same photo-interpreter that labelled its training data. We may also observe that, similarly to what happened with the photo-interpreters, the performance indexes weighted by the area present better values than the corresponding indexes based solely on segments. We believe that the same explanation given before may be used for this fact.

A major result concerns to the Area Hit Ratio — the most important index in our deforestation monitoring application, since it is

related to the overall correctly classified area. The neural networks achieved values very close to the target (95.64%). The network **n_A** achieved 94.07% and **n_B** achieved 95.38% when assessed against the photo-interpreters that labelled its training data. Even when tested against alien photo-interpreters, the networks exhibited good results: 91.30% for **n_A** and 94.24% for **n_B**.

For the other indexes, including the AMSE, the performance indexes achieved by the neural networks were not so brilliant, but it is convenient to remind that these indexes are less important for the practical application that we propose to tackle.

IV. CONCLUDING REMARKS

The performance of different photo-interpreters and neural networks was evaluated on the basis of different performance indexes calculated over the same test set. Some of these indexes (MSE, Jacquard and Hit Ratio) use solely segments as their basic units, while the others (AMSE and Area Hit Ratio) takes into account the number of pixels of each segment (its area) in their definition. In the later case, the performance figures relate more directly to the classifiers' ultimate goal (i.e., to assess the evolution of the deforested area in the Amazon region), as they refer to the same units as end users do, namely areas (as numbers of pixels).

This paper showed that different photo-interpreters do not fully agree when interpreting the same set of images of the Amazon region, and allowed us to establish reasonable performance targets to be sought by designers of automatic classifiers. Moreover it showed that from a practical point of view the trained

neural networks achieved area hit ratios comparable to those exhibited by the different photo-interpreters. Indeed, the classifier performance was rated as very good by photo-interpreters who used the system to classify images that were never seen by the network, which shows a satisfactory generalization ability.

Overall, our results appear to be comparable to the results typically reported on the performance of other classifiers employing traditional pixel-based techniques. However, apparently these other techniques have a very hard time dealing with the transition and interference phenomena we mentioned earlier, which lie as a fundamental motivation to our fuzzy approach. For example, results recently reported on the same images that we employed in our evaluation are comparable to ours but only look at those portions of the image to which the photo-interpreter assigned crisp degrees of membership [13]. What this means is that approximately one fifth of the entire area of the images had to be left untouched, owing to that classifier's inability to deal with fuzzy information. In addition to the ease with which our approach deals with the inherent fuzziness of the domain, there is also the intrinsic ability that it has to cope with the geometric, textural, and contextual (in the form of immediate neighborhoods) characteristics of spectrally homogeneous regions of an image. Pixel-oriented approaches seem to lack this ability.

However, if we look the indexes that don't consider the area of the segments as well as the AMSE we observe that there is much room for improvements in the neural classifier we have built. Some of the improvements may be easily undertaken by further training of our neural networks, that were not trained at their best, as well as by the use of the contextual descriptors referred in [4], which in addition with a recurrent architecture has the ability to deal in a convenient way with segments that present interference like shadows and clouds.

One important technical issue that we feel needs to be addressed has to do with the current complete separation between the segmentation and classification phases in our approach. Perhaps a higher degree of integration between the two phases might be able to produce less fragmented segments in the situations

in which this fragmentation leads to the loss of most of the geometric information of interest. Although we are as yet uncertain as to how this might be achieved, one possibility seems to be the development of a segmentation system that somehow takes into account the categories of interest at later stages, or a system that does not merely follow a region-growing technique but also tries to preserve significant rectilinear boundaries untouched. This later possibility has already received some attention elsewhere [14], [15].

Other possible beneficial ideas are the construction of a neural network trained with segments labelled by photo-interpreter **A** as well as by photo-interpreter **B**, the construction of a specific neural network for small segments and another neural network for larger segments as was proposed in [16]. An evaluation of photo-interpreters and neural networks using data obtained from the field is planned within our research agenda.

V. ACKNOWLEDGEMENTS

We are thankful to the following colleagues, who in various ways contributed to the research described in this paper. Guaraci Erthal and Leonardo Bins, of Brazil's Institute for Space Research (INPE), implemented the software for the segmentation of the images, and produced all the segments we used in our experiments and Cláudio A. de Almeida carried out one of the photointerpretations of the satellite images.

REFERENCES

- [1] V. C. Barbosa, R. J. Machado, and F. dos S. Liporace. A neural system for deforestation monitoring on Landsat images of the Amazon region. Technical Report CCR-157, IBM, Centro Científico Rio, October 1993.
- [2] R. Nelson and B. Holben. Identifying deforestation in Brazil using multiresolution satellite data. *International Journal of Remote Sensing*, 24(1):130-138, 1986.
- [3] R. J. Machado, V. C. Barbosa, F. dos S. Liporace, and C. Ferlin. Monitoring the deforestation of the Amazon region with neural networks. In *Proceedings of the International Joint Conference on Neural Networks*, pages 1239-1242, Nagoya, Japan, 1993.
- [4] V. C. Barbosa, R. J. Machado, and F. dos S. Liporace. A neural system for deforestation monitoring on Landsat images of the Amazon region. *International Journal of Abstract Reasoning*. To appear.
- [5] F. dos S. Liporace. Um sistema neural para monitoração do desflorestamento na região Amazônica utilizando imagens do Landsat. Master's thesis, Universidade Federal do Rio de Janeiro, 1994. In Portuguese.
- [6] J. A. Benediktsson, P. H. Swain, and O. K. Ersoy. Neural networks approaches versus statistical methods

- in classification of multisource remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 28(4):540-552, 1990.
- [7] P. D. Heermann and N. Khazenie. Classification of multispectral remote sensing data using a back-propagation neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 30(1):81-88, 1992.
 - [8] R. C. Gonzales and R. E. Woods. *Digital Image Processing*. Addison-Wesley, 1992.
 - [9] M. Riedmiller and H. Braun. A direct adaptive method for faster backpropagation learning: The RPROP algorithm. In *Proceedings of the IEEE International Conference on Neural Networks (ICNN)*, pages 586-591, San Francisco, 1993.
 - [10] T. M. Lillesand and R. W. Kieffer. *Remote Sensing and Image Interpretation*. John Wiley & Sons, 1987.
 - [11] D. Bailey and D. Thompson. How to develop neural network applications. *AI Expert*, (5):38-47, 1990.
 - [12] F. dos S. Liporace, R. J. Machado, and V. C. Barbosa. Equalization of the training set for backpropagation networks applied to classification problems. Submitted to the First Brazilian Congress on Artificial Neural Networks, 1994.
 - [13] A. J. F. Machado e Silva. An evaluation of the impact of compression techniques on digital image classification. Technical Report CCR-153, IBM, Centro Científico Rio, June 1993.
 - [14] J. F. Haddon and J. F. Boyce. Image segmentation by unifying region growing and boundary information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(10):929-948, 1990.
 - [15] T. Pavlidis and Y. T. Liow. Integrating region growing and edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(3):225-233, 1990.
 - [16] Peter Dreyer. Classification of land cover using optimized neural nets on SPOT data. *Photogrammetric Engineering & Remote Sensing*, 59(5):617-621, May 1993.