Y. E. Shimabukuro
P. Hernandez Filho
N. F. Koffler
S. C. Chen
Departamento de Sensoriamento Remoto
Instituto de Pesquisas Espaciais (INPE)
12.200-São José dos Campos, S.P., Brazil

Automatic Classification of Reforested Pine and Eucalyptus Using Landsat Data

Pixel classification accuracy ranged from 56.61 percent to 93.36 percent, while acreage estimates differed from field data by 6.31 percent.

Abstract: *Single date Landsat* CCTS *(computer compatible tapes) were processed using the G.E. Image-100 to classify Pinus species and Eucalyptus age groups. The study area Mogi-Guaçu is located in the humid sub-tropical climate zone of São Paulo State, Brazil. Prominent Pinus species are P. elliottii and P. taeda. Eucalyptus species are E. alba and E. saligna. The ages of these tree types range from 8 months to 20 years. Four classes were classified by using the Image-100: (1) class PE—P. elliottii, (2) class PO—Pinus species other than P. elliottii, (3) class EY—Eucalyptus spp. from 8 months to 2 years, and (4) class EO—Eucalyptus spp. over two years. The percentage of pixel classification accuracy ranged from 56.61 percent to 93.36 percent. Small lot sizes and the spatial discontinuity of class PO caused difficulties in pixel classification. Comparison of acreage estimated from the Image-100 with field data showed relative differences of −8.38 percent, −12.20 percent, −10.11 percent, and +3.30 percent for the above four classes, respectively. An overall acreage estimation difference of 6.31 percent was achieved. Greater accuracy of area estimation could have been obtained if the field data based on the seedling area at planting time had been updated.*

Introduction

During the past few years, various researchers have tested the possibility of identifying forest plant communities using Landsat data by conventional or automatic classification techniques. Sayn-Wittgenstein (1972) used a color composite of MSS bands 4, 5, and 6 to separate coniferous forest (mostly spruce), hardwood (mostly poplars), and areas covered by willow and alder. Lopez-Cuervo (1973) visually analyzed Landsat enlarged images and identified *Pinus* and *Eucalyptus* at the generic level according to their tonal differentiations. Gimbarzevsky (1974) also confirmed the separability of coniferous and deciduous forest using Landsat imagery. Lee *et al.* (1974) observed that Landsat imagery could be used to distinguish forested and non-forested lands. Difficulties were encountered in separating coniferous species, even though age difference was recognized. Kalensky and Scherk (1975) reported that overall General Electric Image-100 classification accuracies for co-

Photogrammetric Engineering and Remote Sensing,
Vol. 46, No. 2, February 1980, pp. 209-216.

niferous forest, deciduous forest, and nonforested land ranged from 67 percent to 81 percent for single date imagery, while accuracy for multidate imagery analysis was consistently above 80 percent.

The objective of this study was to classify *Pinus* and *Eucalyptus* species and their age groups, using the General Electric Image-100 automatic classification system, and to assess the accuracy of classification.

Study Area and Data Sources

São Paulo State contributes more than 50 percent of the paper and cellulose produced in Brazil. The state has an artificially reforested area of 6,111 km² predominantly in pine (22 percent) and eucalypt (76 percent). The study area, Mogi-Guacu, is located at 22°15′S and 47°10′W and is representative of pine and eucalypt plantations common to the region. This area, approximately 200 km², includes the Campininha pine experimental station of the Forestry Institute of São Paulo State (IFSP) and the Santa Terezinha eucalypt plantation of the Champion Cellulose & Paper Company (CCP). The major *Pinus* species in Campininha are *P. elliottii* and *P. taeda*. Other species such as *P. caribaea*, *P. bahamensis*, *P. oocarpa*, and *P. palustris* are also planted in small areas. The prominent *Eucalyptus* species in Santa Terezinha are *E. alba* and *E. saligna*. The trees range in age from 8 months to 20 years.

Landsat multispectral scanner CCTs acquired on 13 September 1975 with path/row annotation of 236/75 (Figure 1) were processed with the Image-100 system to classify *Pinus* and *Eucalyptus* species and their age groups.

Field data and forest cover maps were provided by IFSP and CCP and used for supervised classification and accuracy analysis. Spot field checks were also carried out.

Image-100 Analysis Procedures

Feature Selection

Prior to the supervised signature extraction and classification, the CCTs were corrected radiometrically. A false color composite image was displayed on the image monitor at an approximate scale of 1:82,000, and a preliminary analysis of the study area was carried out. Owing to the morphological similarities among *Eucalyptus* spp. and their mixed stands in the study area, no species separation was attempted. The following ten classes with sufficiently large areas for testing were selected for the Image-100 analysis:

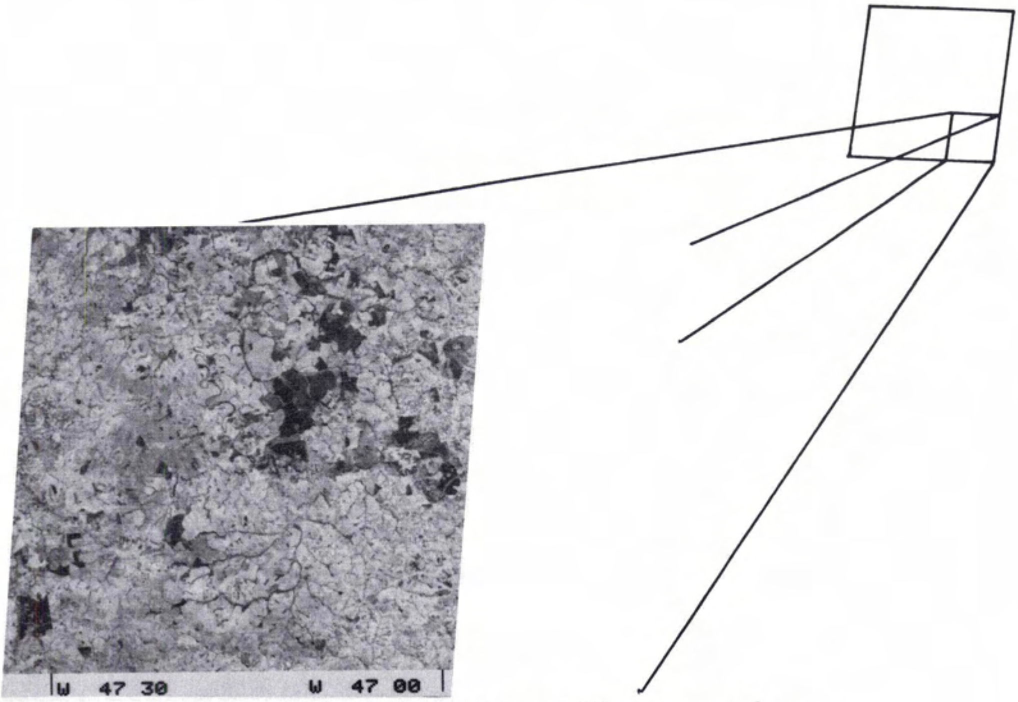

Fig. 1. Landsat MSS band 5 image of Mogi-Guaçu study area.

PE_a—*P. elliottii* over 13 years.
PE_b—*P. elliottii* under 13 years.
PT—*P. taeda*.
OP—*Pinus* species other than *P. elliottii* and *P. taeda*.
$E_{8\ mo.\ \sim\ 2\ yrs.}$—*Eucalyptus* spp. from 8 months to 2 years.
$E_{2\ yrs.\ \sim\ 4\ yrs.}$—*Eucalyptus* spp. from 2 to 4 years.
$E_{4\ yrs.\ \sim\ 7\ yrs.}$—*Eucalyptus* spp. from 4 to 7 years.
$E_{7\ yrs.\ \sim\ 20\ yrs.}$—*Eucalyptus* spp. from 7 to 20 years.
$RE_{<2\ yrs.}$—Regrowth *Eucalyptus* spp. under 2 years.
$RE_{2\ yrs.\ \sim\ 9\ yrs.}$—Regrowth *Eucalyptus* spp. from 2 to 9 years.

Bhattacharyya distances (B-distance, Marill and Green, 1963) between these ten preliminary classes were calculated using INPE's feature selection algorithm. This algorithm assumed that each class has a multivariate Gaussian distribution. In this operation, 40 to 50 training samples (4 pixels in size) of each class were delineated on the image monitor with the electronic cursor. Then the B-distance value was calculated in the four channels. The results of the B-distance values in four channels for the preliminary *Pinus* and *Eucalyptus* classes are presented in Tables 1 and 2.

The separability of the preliminary classes was defined according to the values of "B-distance". Using the curve of Swain and King (1973), the estimated probability of correct classification is >0.85 for the value of B-distance >1.30. Thus, those classes having B-distance values less than 1.30 were grouped together to form four classes:

PE —*P. elliottii*.
PO —*Pinus* species other than *P. elliottii*.
EY —*Eucalyptus* spp. from eight months to two years.
EO —*Eucalyptus* spp. over two years.

Figure 2 shows the distribution of the four classified classes in the study area. Table 3 lists the "B-distance" values of classes PE, PO, EY, and EO in four spectral channels.

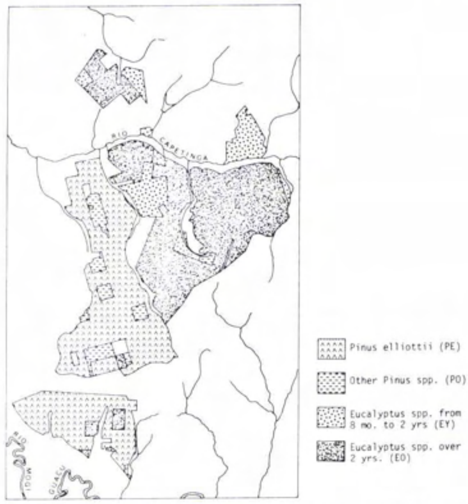

FIG. 2. Forest cover types of Campininha and Santa Terezinha, grouped to match forest types being classified using the G.E. Image-100.

TABLE 1. B-DISTANCE VALUES OF PRELIMINARY CLASSES OF *Pinus* IN FOUR SPECTRAL CHANNELS

CLASS \ CLASS	PE_b	OP	PT
PE_a	0.22	1.92	1.70
PE_b		1.79	1.37
OP			0.67

SIGNATURE EXTRACTION AND CLASSIFICATION

Single cell signature acquisition of the G.E. Image-100 was used for classification. This option creates a four-dimensional parallelepiped, each of the sides of which corresponds to the signature limits of the training areas in each channel. Training samples, independent from the test area, were selected

TABLE 2. B-DISTANCE VALUES OF PRELIMINARY CLASSES OF *Eucalyptus* IN FOUR SPECTRAL CHANNELS

	E 2 yrs. ~ 4 yrs.	E 4 yrs. ~ 7 yrs.	E 7 yrs. ~ 20 yrs.	RE <2 yrs.	RE 2 yrs. ~ 9 yrs.
E 8 mo. ~ 2 yrs.	1.99	1.99	1.98	1.94	1.98
E 2 yrs. ~ 4 yrs.		0.41	0.23	0.51	0.31
E 4 yrs. ~ 7 yrs.			0.63	0.70	0.49
E 7 yrs. ~ 20 yrs.				0.56	0.55
RE <2 yrs.					0.23

TABLE 3. B-DISTANCE VALUES OF CLASSES PE, PO, EY AND EO IN FOUR CHANNELS

CLASS \ CLASS	PO	EY	EO
PE	1.78	2.00	1.99
PO		2.00	2.00
EY			1.90

from each of the four classes (PE, PO, EY, and EO) using the electronic cursor. The limits of the histograms of the training areas (spectral bounds of the classes) were modified in the four spectral channels until a satisfactory classification was obtained. The spectral statistics of the histograms from the training samples of four classes are given in Tables 4 through 7, respectively. The classification results using sample information are presented in Figure 3.

TABLE 4. SPECTRAL STATISTICS FROM THE TRAINING SAMPLES OF CLASS PE

	Spectral Bounds				
Channel	LB	UB	Delta	Mean	Variance
4	17	21	5	19.0	0.8
5	11	15	5	13.3	1.0
6	18	26	9	22.2	2.5
7	27	32	6	30.1	2.5

TABLE 5. SPECTRAL STATISTICS FROM THE TRAINING SAMPLES OF CLASS PO

	Spectral Bounds				
Channel	LB	UB	Delta	Mean	Variance
4	18	19	2	18.5	0.2
5	11	15	5	13.3	1.0
6	16	21	6	19.1	2.3
7	24	26	3	25.4	0.7

TABLE 6. SPECTRAL STATISTICS FROM THE TRAINING SAMPLES OF CLASS EY

	Spectral Bounds				
Channel	LB	UB	Delta	Mean	Variance
4	18	21	4	19.6	1.0
5	12	19	8	14.5	2.5
6	35	42	8	36.7	3.8
7	42	56	15	48.1	7.9

TABLE 7. SPECTRAL STATISTICS FROM THE TRAINING SAMPLES OF CLASS EO

	Spectral Bounds				
Channel	LB	UB	Delta	Mean	Variance
4	18	21	4	19.8	1.0
5	13	18	6	15.2	2.0
6	24	34	11	27.6	4.6
7	34	44	11	37.6	7.7

ACCURACY OF IMAGE-100 CLASSIFICATION

PIXEL CLASSIFICATION ACCURACY

After classification, the electronic cursor was again positioned on the image monitor. The "alphanumeric theme print" was executed for a cursored area of approximately 50 km^2. The IFSP and CCP forest lot boundaries were delimited on the printout to make pixel by pixel comparisons (Figure 4). The correctly classified and misclassified pixels were counted and are shown in a confusion table (Table 8). This table shows the number of pixels classified according to their "true" class and G.E. Image-100's "chosen" class in the cursored area. To verify the G.E. Image-100's pixel classification accuracy, the following formulas were used to calculate the percentages of correct classification and proportion of misclassification.

$$\text{Correct Classification (\%)} = \frac{\text{No. of correctly classified pixels in class A}}{\text{No. of pixels of true class-A in the test area}} \times 100$$

$$\text{Proportion of misclassification (\%)} = \frac{\text{No. of pixels classified as class A but actually belonging to another class}}{\text{Total no. of pixels classified as A in the test area}} \times 100$$

AREA ESTIMATION ACCURACY

Area estimation accuracy is as important as class identification. A close approximation of estimated class area to field data depends on percent correct pixel identification and proportion of misclassification. In this study the G.E. Image-100 estimated class areas were compared to corresponding field data and their percentages of relative difference are presented in Table 9. The areal fractions

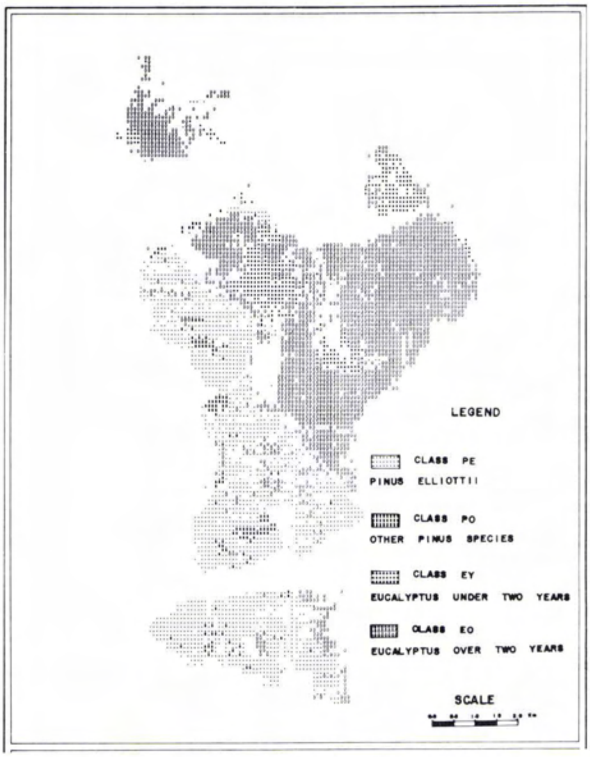

FIG. 3. Alphanumeric printout of the classified Mogi-Guaçu study area.

of each of the four classes calculated from field data were used as weighing factors to determine the overall acreage estimation difference.

DISCUSSION AND CONCLUSIONS

A "B-distance" value of 0.22 indicated that no satisfactory separation between PE_a and PE_b was possible. *P. taeda* and the class OP were readily separable from *P. elliottii* but unseparable from each other (Table 1). The morphological similarity among the *Eucalyptus* species made their differentiation difficult, although the G.E. Image-100 successfully classified *Eucalyptus* of 8 months to 2 years from the other preliminary age groups (Table 2). This separability was due to the distinct spectral characteristics of the homogeneous young *Eucalyptus* which aided the G.E. Image-100 identification. The greater "B-distance" values between *Pinus* and *Eucalyptus* genera confirmed that genera were more separable than species and age classes (Table 3).

For classification, a hyperparallelepiped signature file was constructed for each of the classes PE, PO, EY, and EO using a "single cell signature acquisition" program. The spectral statistics of the signature files are shown in Tables 4 through 7. The percentage of correct classification of G.E. Image-100 for the classes PE, PO, EY, and EO were found to be 89.14 percent, 56.61 percent, 88.30 percent, and 93.36 percent, respectively. The lowest percentage of correct classification was obtained from the class PO (56.61 percent). This may be explained by the fact that class PO only occupies 3.5 percent of the study area and is distributed in eight different lots with areas from 5 to 50 hectares. The spatial discontinuity and small lot sizes caused a higher border effect; thus, a higher proportion of pixels being classified as their surrounding class (PE).

The G.E. Image-100 estimated class areas when compared to the field data resulted in relative differences of −8.38 percent, −12.20 percent, −10.11 percent, and +3.30 percent. The overestimation of class EO (+3.30 percent) was due to a high proportion of misclassified pixels (13.61 percent) which had produced a spectral signature similar to class EO. A total acreage estimation difference of 6.31 percent was achieved. The results of the G.E. Image-100 areas estimations were en-

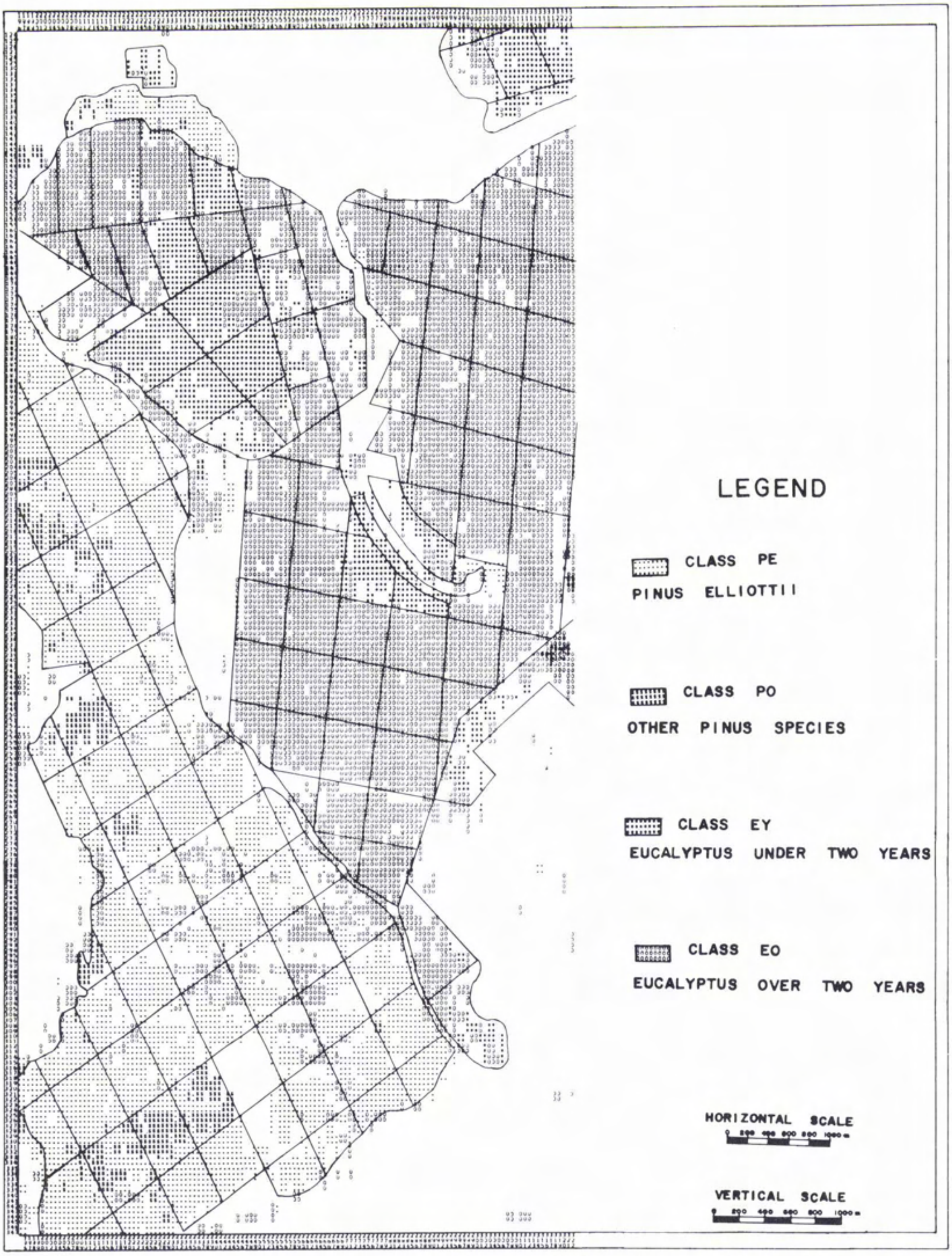

FIG. 4. Alphanumeric theme print with IFSP and CCP forest lot boundaries used for pixel classification accuracy measurement.

couraging, considering that the machine classification was carried out at the species level for pine, and based on age groups for eucalypt.

The field data provided by the IFSP and CCP agencies were based on the seedling area at planting time. These data represented "theoretically reforested areas" under "optimum conditions" without damage and mortality. In the field check, various plant-

TABLE 8. CONFUSION TABLE

True Class \ Chosen Class	PE	PO	EY	EO	Blank	Pixel Total	Correct Classification (%)
PE	5730	30	0	449	219	6428	89.14
PO	359	535	0	40	11	945	56.61
EY	0	0	1713	195	32	1940	88.30
EO	24	0	155	6272	267	6718	93.36
Other Classes	98	13	10	304			
Pixel Total	6211	578	1878	7260			
Proportion of Misclassified Pixels (%)	7.74	7.44	8.79	13.61			

TABLE 9. COMPARISON OF G. E. IMAGE-100 ESTIMATED AREAS WITH FIELD DATA

	Area (KM²)		(WF)	(RD)
Class	(A) Field Data	(B) Image-100 Estimates	Weighing Factor $(A_i)/(\Sigma A_i)$	Relative Difference $(B - A)/(A) \times 100$
PE	18.86	17.28	0.404	− 8.38
PO	1.64	1.44	0.035	−12.20
EY	4.45	4.00	0.095	−10.11
EO	21.79	22.51	0.466	+ 3.30

Overall Acreage Estimation Difference $= \sum_{i=1}^{4} (WFi \times |R\,Di|) = 6.31\%$

failure areas were observed. The area estimation accuracy could have been greater if the field data had been up-dated annually, thus accounting for seedling mortality.

The following significant results were obtained from automatic classification of reforested areas of pine and eucalypt:

Based on "B-distance" values the following spectral separations were achieved:

- *P. elliottii* was separated from the other pine species,
- *Eucalyptus* spp. from 8 months to 2 years old were separated from *Eucalyptus* spp. more than 2 years old, and
- Clear spectral distinctions were made between pine and eucalypt.

Based on Image-100 area estimations, the relative differences ranged from −12.20 percent to +3.30 percent for the four classes defined (PE, PO, EY, and EO) when compared to the same information provided by forestry agencies.

These results point to the potentiality of computer aided processing of Landsat data to recognize tree species, especially in remote regions, where ground and aerial surveys are often costly. Before the methodology used in this study can be applied on an operational scale, more consideration should be given to signature extention to other geographical regions, and to spatial differences between areas of natural and artificial reforestation.

ACKNOWLEDGMENT

The authors are grateful to the Director of the Instituto de Pesquisas Espaciais, Dr. Nelson de Jesus Parada, for his support and to Dal A. Cottrell for assistance on the manuscript.

REFERENCES

Gimbarzevsky, P., 1974. ERTS-1 imagery in biophysical studies, *Proceedings, Second Canadian Symposium on Remote Sensing*, Ottawa, V. 2, P. 391.

Kalensky, Z., and L. R. Scherk, 1975. Accuracy of forest mapping from Landsat computer compatible tapes, *Proceedings, Tenth International Symposium on Remote Sensing of Environment*, Environment Research Institute of Michigan, Ann Arbor, Michigan, V. 2, p. 1159.

López-Cuervo, S., 1973. A study of agroforestry natural resources, *Proceedings, First Panamerican Symposium on Remote Sensing*, Panama City, EROS Data Center, Sioux Falls, P. 86.

Lee, Y. J., E. T. Oswald, and J. W. E. Harris, 1974. A preliminary evaluation of ERTS imagery for forest-land management in British Columbia, *Proceedings, Second Canadian Symposium on Remote Sensing*, Ottawa, V. 1, P. 87.

Marill, T., and D. M. Green, 1963. On the effectiveness of receptor in recognition systems, *IEEE Trans. on Information Theory*, V. IT-9, P. 11.

Sayn-Wittgenstein, L., 1972. A first look at Canadian ERTS experiments in forestry, *Proceedings, Earth Resources Technology Satellite-1 Symposium*, Goddard Space Flight Center, Beltsville, Maryland, P. 48.

Swain, P. H., and R. C. King, 1973. *Two effective feature selection criteria for multispectral remote sensing*, LARS Information Note 042673, Laboratory for Applications of Remote Sensing, Purdue Univ., W. Lafayette, IN.

(Received 29 September 1978; revised and accepted 13 September 1979)

New Sustaining Members

Diretoria de Eletrônica e Proteção ao Vôo

Divisão de Cartografia e Informações Aeronauticas
Aeroporto Santos Dumont, 4º And. 20021, Rio De Janeiro, Brasil

THE DIRETORIA de Eletrônica e Proteção ao Vôo, Divisão de Cartografia e Informações Aeronáuticas, is the branch of the Brazilian Air Ministry responsible for the production of Aeronautical Charts and related publications.

Division products include the World Aeronautical Chart (WAC) and the Navigation Set, produced and distributed every 112 days.

The Navigation Set consists of textual and graphic information required to plan and conduct a flight under varying weather conditions. Information within these documents are navigational facilities, aerodrome data, airway system, etc.

The Division employs military and civil engineers and technicians qualified in the fields of geodesy, photogrammetry, and flight information.

Equipment consist of one Wild B8S, one CMA-751 Satellite Doppler Survey System with a mini/computer HP 21 MX, one CALCOMP Plotter, and related mapping and charting equipment.

Dunn Instruments, Inc.

544 Second Street, P.O. Box 77172, San Francisco, CA 94107

DUNN INSTRUMENTS entered the field of computer graphics and digital image processing in the spring of 1979 with an inexpensive device for high resolution, color hard copy. This device, the 631 Color Camera System, is a compact, stand-alone unit which connects to the RGB outputs of any raster scan graphics terminal and produces photographic hard copy in a variety of formats, including 8 × 10 instant color prints, 35 mm slides and 8 × 10 overhead transparencies. The 631 is fully automated, fast (it makes up to thirty 8 × 10 prints or sixty 35 mm slides per hour), costs a fraction of what other available color hard copy systems cost, and produces extremely high resolution results with accurate rendition of color hue, saturation, and brightness.

Color hard copy from the 631 System is a basic communications tool for business, government, and science. Its applications include the reproduction of digitally processed images (such as those obtained by remote sensing), cartography and the presentation graphics used in management information systems, for sales and marketing support, and for internal reporting and group meetings.

Dunn Instruments' 631 System was first exhibited at the 1979 National Computer Conference and has been used extensively at the Harvard Laboratory for Computer Graphics and Spatial Analysis, the Goddard Space Science Center, Lawrence Livermore and Lawrence Berkeley Laboratories, and the NASA-Ames Research Center where it made the first photographs for media distribution of Pioneer II's Saturn encounter.

(Continued on page 242)