E7.4-10571
CR-138646

N74-28815

(E74-10571) A MULTILEVEL, MULTISPECTRAL DATA SET ANALYSIS IN THE VISIBLE AND INFRARED WAVELENGTH REGIONS (Purdue Univ.) 45 p HC $5.25 CSCL 05B

Unclas G3/13 00571

A MULTILEVEL, MULTISPECTRAL DATA SET ANALYSIS IN THE VISIBLE AND INFRARED WAVELENGTH REGIONS*

L. L. Biehl and L. F. Silva

Laboratory for Applications of Remote Sensing
Purdue University
West Lafayette, Indiana 47907

ABSTRACT

SKYLAB multispectral scanner data, digitized SKYLAB color IR photography, digitized SKYLAB black and white multiband photography, and Earth Resources Technology Satellite (ERTS) multispectral scanner data collected within a twenty-four hour time period over an area in south-central Indiana near Bloomington on June 9 and 10, 1973, were compared in a machine-aided land use analysis of the area.

The overall classification performance results, obtained with nine land use classes were 87% correct classification using the "best" 4 channels of the SKYLAB multispectral scanner, 80% for the channels on the SKYLAB multispectral scanner which are spectrally comparable to the ERTS multispectral scanner, 88% for the ERTS multispectral scanner, 83% for the digitized color IR photography, and 76% for the digitized black and white multiband photography.

The results indicate that the SKYLAB multispectral scanner data was degraded by noise problems and that the digitized black

*This work was supported by the National Aeronautics and Space Administration under Contract No. NAS 9-13301.

and white multiband photography was degraded by "flare" around a lake in the infrared frames.

INTRODUCTION

The infrared region has long been recognized as an important region of the spectrum to be utilized for remote sensing of the environment by a large variety of disciplines as geology, forestry, water resources, crop studies, and land use studies. The energy reflected and emitted from the earth in the infrared and visible region are collected by two primary techniques--photographic emulsions and multispectral scanners. The resulting imagery can be interpretted by visual techniques or if in a digital format by a computer using pattern recognition techniques.

Multispectral scanners sample the energy in different regions (termed channels) of the optical spectrum as a mirror scans the scene. The optics and detectors determine the spectral windows of the individual channels. The energy levels for each channel are stored in either an analog or a digital format on a magnetic tape to form a multispectral data set. Multiemulsion photography (as color IR film) and multiband photography (combination of different cameras, films, and filters) are also used to form multispectral data sets.

This particular study compares four multispectral data sets acquired within a 24 hour time period on June 9, and June 10, 1973, over an area in south central Indiana in a land use analysis of the study area (see fig. 1). The data sets were acquired by the 4 channel multispectral scanner (MSS) on the Earth Resources

Technology Satellite (ERTS), the Earth Resource Environmental Package (EREP) 13-channel MSS on Skylab, the color infrared photograhy from the EREP camera system, and the black & white multiband photography from the EREP camera system.

The study area was centered around Lawrence County and parts of the surrounding counties in south central Indiana, an area covering approximately 147,000 hectares. The land is rolling with around 140 meters of local relief and a variety of land use categories--a lake, a river, forests, agricultural fields, residential, and commercial areas.

This study used the IBM 360 model 67 computer at the Laboratory for Applications of Remote Sensing (LARS) at Purdue, the MSS data sets and the digitized photographic data sets to classify the study area into twelve land use classes which correspond to those suggested by Hardy, Anderson, and Roach [1]: new residential, old residential, commercial-industrial, extractive, light soil, dark soil, grass, sparse wood, deciduous forest, coniferous forest, river, and lake. For analyzing the results, the new and old residential classes were considered as one class, as were the light and dark soil classes and the sparse wood and deciduous forest classes. These pairs of classes were combined because of the difficulty in deciding where one class ended and the other one began.

The classification process used the pattern recognition algorithms that have been implemented on the computer at LARS in a software package called LARSYS [Ref. 2,3]. The procedure consisted of choosing fields or areas to represent each of the 12 classes listed above. These training fields representing around 3% of the study area were scattered throughout the study area. The same ones were used for each of the four data sets. The possibilities for training fields were somewhat limited by clouds in the ERTS MSS data set. The training fields were evaluated using a clustering routine to find if further division of the 12 classes were necessary so that the resulting spectral class represented unimodal distributions. Under the assumption that the samples for each spectral class were normally distributed (Gaussian), the n-dimensional mean vector and the n x n covariance matrix were computed for each class where n represents the number of channels of spectral information in each data set. The mean vectors and covariance matrixes were used in the actual classification routine which uses the Bayes maximum likelihood classification rule with the *a priori* probability of occurance for each class being equal.

A modified divergence was used to select the "best" set of 4 out of the 11 usable channels in the SKYLAB MSS (S192) data set (a feature selection routine) and to obtain a measure for the separability of the classes in each of the data sets. Divergence, a measure of separability between two density functions which represent two land use classes, for n spectral channels $C_1, C_2, \ldots C_n$ assuming normal variables is given [4] by Eq. (1)

$$D(i,j/C_1, C_2, \ldots C_n) = \mathrm{tr}\,[(K_i-K_j)\,(K_j^{-1} - K_j^{-1})] + 1/2\,\mathrm{tr}\,[(K_i^{-1} + K_j^{-1})\,(M_i-M_j)\,(M_i-M_j)^T] \quad (1)$$

where M and K represent the mean vector and covariance matrix respectively; tr [A] (trace A) is the sum of the diagonal elements of A. A modified form of divergence D_T, termed transformed divergence, was used in this study because it behaves more like the probability of correct classification [3]. See Eq. (2).

$$D_T = 2000\ [1-\exp(-D/8)] \tag{2}$$

Transformed divergence was extended to a multiclass case to choose the "best" 4 channels in two separate ways. The average taken over all possible class pairs was maximized and the minimum transformed divergence of all possible class pairs was maximized. There is no guarantee, however, that either of these methods are optimal. The selection of 4 channels for the classification was done to reduce computer costs and to find if the channels selected included any spectral bands not available in the other data sets.

The products obtained in the study include the transformed divergence measures for the separability of each pair of classes in each data set, the classification maps, and the classification performance results. The classification performance results were obtained by selecting six test areas scattered throughout the study area representing the classes under consideration. The underflight photography was projected onto the classification map of these test areas and a point by point check of the classification was done. The total test area represented 3.5% of the study area or approximately 5,150 hectares. (The test areas are outlined in fig. 13.)

INSTRUMENTS (PLATFORMS) USED IN STUDY

Ground observations consisted of spectral data acquired over Lake Monroe at the time of the Skylab overpass by a field spectroradiometer system (described elsewhere in this issue). Spectral data was also acquired over the Purdue Farm near Oolitic, Indiana following the overpass. A pyroheliometer system monitored the solar irradiance for several hours throughout the day.

At 14:39 GMT on June 10, 1973, an aircraft (flown by the Environmental Research Institute of Michigan, ERIM) acquired 12 channel MSS data and photography over a north-south strip over Lake Monroe 1520 m. above the ground. The 12 channels cover a wavelength range from .41-11.7 μm, 8 visible channels, 3 infrared channels, and 1 thermal or far infrared channel. The photography from the aircraft consisted of 24.1 cm. (9.5 inch) black and white transparencies and prints (see fig. 2), 70mm color and color IR positive transparencies.

At 16:30 on June 10, 1973, a WB-57 high altitude reconnaissance aircraft (flown by the National Aeronautics and Space Administration) acquired photography along the track of the Skylab pass across Indiana and a north-south strip across Bloomington, Lake Monroe, Bedford, and Mitchell at an altitude of 18,300 m. (60,000 ft). The photography consisted of 24.1 cm (9.5 inch) color and color IR positive transparencies (see fig. 3), 70 mm color, color spectral IR, and 4 bands of black & white positive transparencies.

On October 24, 1973, a Cessna 310 (operated by LARS) flew the portion of the study area that is in Monroe County at an altitude of 2,120 m. (7,000 ft.) acquiring 70 mm color and color IR positive transparencies.

At 14:26 GMT on June 10, the Skylab space station flew-over the test area at an altitude of 440 km. The weather conditions were excellent; the sky was clear and the air was calm. Data was utilized from 3 of the 5 Earth Resources Experiment on Skylab: photographic, spectroradiometric, and MSS data [5].

The photography was acquired by the S190A and S190B experiments on EREP. The S190A experiment is an array of 6 cameras with high precision f/2.8, 21.2° field of view, 153 mm (6 inch) focal length lenses obtaining 1:3,000,000 scale photography. The film products (see fig. 4 and 5) include 70 mm color, color IR, and 4 spectral bands of black and white transparencies (.5-.6, .6-.7, .7-.8 & .8-.9μm). The film types and ground resolutions of the S190A photography used for this study are given in Table 1. The S190B photographic camera (Earth Terrain Camera) has a f/4 lens with a focal length of 458 mm (18 inches) obtaining 1:80,000 scale photography. The film products were 5" color positive transparencies.

The S191 Experiment of EREP is the spectroradiometer. The spectroradiometer scans the ground target spectrally from .39-2.5 μm and 5.82 to 15.99 μm. The field of view of the instrument is one milliradian or approximately 564 meters on the ground. The radiation is collected by a 254 mm (10 inch) primary mirror.

Table 1

S190A Film Characteristics

Wavelength (μm)	Film	Filter*	Dynamic Resolution On the Ground (m)
0.5-0.6	PAN-X B&W (SO-022)	AA	30
0.6-0.7	PAN-X B&W (SO-022)	BB	28
0.7-0.8	IR B&W (EK 2424)	CC	68
0.8-0.9	IR B&W (EK 2424)	DD	68
0.5-0.88	IR Color (EK 2443)	EE	57

*As designated by the EREP INVESTIGATORS' INFORMATION BOOK

The multispectral scanner (the S192 experiment) on Skylab has 13 channels which cover the wavelength range of .41 μm to 12.5 μm, 5 visible channels, 7 infrared channels, and 1 thermal infrared channel (see Table 2). The incoming radiant energy is collected by a 432 mm (17 inch) spherical collecting aperature. The radiation in all 13 bands are focused onto HG:CdTe detectors. The multispectral scanner has a conical line scan with an instantaneous field of view (IFOV) of .182 milliradians or 79 m sq. ground coverage. The total field of view is 68.5 km. on the ground.

At 15:59 GMT on June 9, the Earth Resources Technology Satellite (ERTS) acquired multispectral scanner data over the test site at an altitude of 933 km. The multispectral scanner on ERTS consists of 4 channels, 2 visible, and 2 infrared which cover a wavelength from .5 μm to 1.1 μm (see Table 3). The multispectral scanner on ERTS has a rectilinear line scan with a total field of view of 185 km. and an IFOV of 79 meters sq. The detectors for channels 1, 2, and 3 are photo multiplier tubes (PMT) and the detectors for channel 4 is a silicon photo diode.

Table 2

SKYLAB MSS (S192) Spectral Bandwiths

Channel	Spectral Bandwidths (μm)
1	.41- .46
2	.46- .51
3	.52- .56
4	.56- .61
5	.62- .67
6	.68- .76
7	.78- .88
8	.98- 1.08
9	1.09- 1.19
10	1.20- 1.30
11	1.55- 1.75
12	2.10- 2.35
13	10.20-12.50

Table 3

ERTS MSS Spectral Bandwidths

Channel	Spectral Bandwidths (μm)
1	.5- .6
2	.6- .7
3	.7- .8
4	.8-1.1

ANALYSIS OF MULTISPECTRAL DATA SETS

Photographic Data Analysis

The EREP's S190A film products from Skylab used for the analysis were 2nd generation contact positive transparencies made from the 70 mm original. The original color IR transparency was duplicated onto Kodak Ektachrome Aerographic Duplicating Film SO-360 by Kodak Rainbow or Colorado Continuous Contact Printers;

Versamat 1811/Ea-5 Process. The 70 mm original Kodak SO-022 and Kodak 2424 were duplicated onto Kodak Fine Grain Aerial Duplicating Film 2430 and Kodak Aerographic Duplicating Film 2420 respectively. These duplications were done on a Kodak Niagra Continuous Contact Printer by the Fultron or Versamat/MX 641 process.

These 2nd generation 70 mm color IR, and the 4 black and white positive transparencies were digitized by Mead Technology Laboratories at Dayton, Ohio. The instrument used was a modified Fairchild Scan-A-Color Model 4 drum scanner. The instrument has been modified so that its spot size could be reduced to 12.5 μm and 25 μm from the original 50 μm.

The 70 mm color IR transparency was separated and digitized using a 25 μm aperature and a 20 μm sampling interval. The method used was to scan a line with a Kodak 92 filter to separate the red layer, repeat the scan with a Kodak 93 filter for the green layer, and scan again with a Kodak 94 filter for the blue layer. The scanner was then stepped down a line and the process was then repeated. The 25 μm aperature corresponds to approximately 68 m. on the ground, the sampling interval to 55 m. The same aperature size and sampling interval were used to digitize the 4 70 mm black and white transparencies; filters were not used.

The binary calibration of the scanner was done such that the binary was set to 0 when there was no attenuation of light on the glass drum and the binary was set to 255 using the dark area between the frames of the imagery. The calibration for the color IR image was performed through each of the three filters.

The digital tapes from Mead were then reformatted at LARS to be compatible with the LARSYS software for the classification routines used in the analysis. The digitization of the color IR film produced a 3 channel data set (see Table 4 and fig. 6) to be used the same as the MSS data sets. The three channels were registered quite closely, because of the technique used to digitize the frame.

Table 4

Spectral Bandwidths for the Digitized
S190A Color IR Data Set

Channel	Spectral Bandwidths (μm)
1	.52-.58
2	.58-.68
3	.68-.88

*There is actually some overlap between the channels because of the dye characteristics of the film.

The 4 black and white digitized frames had to be registered. A 1100 line by 1100 column area (test site included) of the original 3,000 lines by 3,000 columns for each of the 4 frames were registered. Around sixty points evenly distributed across the area for each of the frames were chosen as common points. The 4 areas were then registered with the help of the computer by a LARS registration program which uses a 2nd order 6 coefficient polynomial least square fit so that a 4 channel data set was produced (see Table 5 and fig. 7). The registration of the two visible channels and the registration of the two infrared channels are within a pixel, approximately 40 m. The registration between the visible and infrared channels is within 2 pixels. The registration between the visible and infrared channels was difficult because of the differences in the resolutions of the two types of film--approximately 30 m. on the ground for the visible and 70 m. on the ground for the infrared.

Table 5

Spectral Bandwidths for the Digitized
S190A B&W Multiband Data Set

Channel	Spectral Bandwidth (μm)
1	.5-.6
2	.6-.7
3	.7-.8
4	.8-.9

The training fields were selected as mentioned before for both the 4 channel and 3 channel data sets for each of the 12 classes. The areas were first clustered and refined as explained previously. In clustering the individual classes for the 4 channel data set, three separate spectral classes of deciduous forest were recognized, two spectral classes of old residential and one spectral class for each of the other classes. For the 3 channel data set one spectral class was recognized for each land use class. One of the three spectral classes for deciduous forest in the 4 channel data set occured near Lake Monroe because of a gradual change in response from the lake to the forest in the two infrared bands rather than a sharp change as indicated by the infrared band in the color IR data set. The infrared response of this spectral class of deciduous forest was lower than the spectral class for coniferous forest. This phenomenon will be discussed later. Another spectral class represented a transition between the dense forest (the third spectral class) and the sparse wood class. The training field for this spectral class was from a forested area near Bloomington. The two spectral classes for old residential were actually close to each other; however, one was very close to the spectral class for river so that it was decided best to keep the two classes separate rather than combine them to help separate the river from old residential.

The classification performance results (see Table 6 and fig. 8 and 9) show that for the classes considered the digitized color IR data set did better overall than the digitized black and white multiband data set. This was in part due to an area around the lake which was assigned to the coniferous class instead of the deciduous class and to some of the points around the edge of the lake which were delineated as commercial-industrial instead of lake. Also many data points in the bare soil and the residential areas were incorrectly classified as river.

The misclassifications around the lake are due to the gradual change in response or fuzzy boundary between the lake and the forest as mentioned earlier. The indistinct boundary or "flare" is noticeable in the two original infrared photographic frames (see fig. 10). This anomaly is similar to the well known phenomenon in photography when one photographs bright objects on a dark background. "Flare" occurs at the edge of the object in the photograph because of the high contrast. The brightness level of the object is reduced along the edge and that of the background is raised.

In this particular case, there is a high contrast between the lake (low response) and the forest (high response) in the two infrared bands which have been filtered so that the film was not exposed to the energy in the visible spectrum. In the color IR frame the dye layer most sensitive to infrared was also exposed to energy in the visible spectrums so that the contrast between the lake and the forest was lower; therefore none or little "flare" occured. It is suspected that the "flare" was caused by internal reflections in the lens system of the camera.

Because of the small aperature used in digitizing the film (compariable to the resolution of the infrared film), the drum scanner measured this "flare" around the lake, giving a false indication of the actual infrared response of the lake (higher) and the forest (lower) near the boundary of the two. The spectral class of deciduous forest trained near the lake did represent the deciduous forest data points nearest the lake; however, there was a band of data points outside of this one which was nearest the spectral class for coniferous forest. Those data points representing water near the edge of the lake were nearest the spectral class for commercial-industrial areas because of the higher infrared response.

The delineation of those points in the bare soil and residential areas as river was probably due to a poor training class for river. Because the four frames making up the four channel data set were digitized separately, the centers of the resolution elements are probably not the same. When registered to the nearest point, the centers may still be off a maximum of 40 meters. Since the river is rather narrow, one to two data points wide, the training class for river probably didn't represent the true spectral nature of the river.

If the test area which includes the lake is left out, the overall performance of the two data sets for the five remaining test areas are much closer-77% correct classification for the digitized multiband photography and 80% correct classification for the digitized color IR photography.

Table 6

Classification Performance Results
(Percent Correct)

Class	SKYLAB MSS 3,7,8,11	SKYLAB MSS 3,5,6,8	ERTS MSS	Color IR	4 Band B&W
Residential	97	81	97	91	84
Commercial	73	33	61	76	46
Extractive	51	59	61	32	34
Soil	87	78	83	67	78
Grass	95	86	93	82	69
Wood-Decid.	81	80	86	84	77
Coniferous	99	68	95	85	43
River	87	27	77	16	64
Lake	89	86	86	98	93
Overall (by points)	87	80	88	83	76
Class Average	84	66	82	70	65

SKYLAB Multispectral Scanner (S192) Data Analysis

The quality of the MSS data from Skylab from a visual observation was good to poor (see fig. 11). Channel 1, .41-.44 μm, was of no use. The atmospheric scattering was so great that the scanner saw just the atmosphere in this band. The noise varies from channel to channel with channels 3, 7, & 11 appearing to be the best from a visual point of view. The thermal band (channel 13) was not received in this data set from the Johnson Space Center.

The same training fields as used in the digitized photography described earlier were selected from this data set to represent the 12 classes. Clustering indicated 13 spectral classes, two spectral classes of lake. The 2nd lake class represented those points in Channel 12 which were anomalies, saturated data values, probably occuring in the processing of the original data.

A separability measure as described earlier was used to choose the best 4 of the eleven usable channels. The set of 4 channels having the highest minimum transform divergence for all pairwise combinations was chosen. The average transformed divergence for this combination was only one less than the combination with the highest average transformed divergence. The optimum channels as given by the transformed divergence measure included the three channels which appear to be the cleanest visually. A classification was also done using the channels of the SKYLAB MSS data set which correspond as near as possible to the channels on the ERTS MSS.

The test areas for computing the classification performance results were selected to match those used in the digitized photography. The results are given in Table 6 (also see fig. 12). The individual class performances and hence the overall performance of the channels 3,5,6,8 classification were generally lower than the channels 3,7,8,11 classification. The difference is probably due to two reasons--the difference in information content of the channels and the difference in the noise of the channels. Channels 5 and 6 appear visually to be noisier than channels 7 and 11. In the 3,7,8,11 classification all classes except the extractive and the commercial-industrial class were separated quite well. The extractive class includes the limestone quarries in the study area. There is a large variance in the quarries, some with no vegetation and some older ones with trees scattered around the water in the pits. The trees and water made some of the quarries hard to delineate with the training fields that were used. The commercial-industrial class was confused with the dark soil class. This pair of classes had the lowest transformed divergence measure separating them.

In the 3,5,6,8 classification, the commercial-industrial and the extractive classes were again not well separated. In addition to those classes, the river class was confused with the residential, commercial-industrial, and the dark soil classes. The coniferous class was confused with the sparse woods class. In all of the cases, the divergence measurements between the confused classes were significantly less than in the 3,7,8,11 combination (see Tables 9 & 10).

To obtain a handle on the noise in the data, the mean and standard deviation of the data values for each of the eleven channels were computed for four separate areas and the total of the four areas (189 data points) in the deepest part of Lake Monroe (greater than 6 meters) and well away from the shore (at least 150 meters). The means and standard deviations of the 4 separate areas were in close agreement. The lake was chosen because it would have the most uniform spectral response of any scene in the study area. It should be noted that the spectral response of water does vary with such factors as turbidity; however, the areas were selected to minimize these effects. Two measures were used to obtain a representation of the noise level in the individual channels--the standard deviation and the standard deviation divided by the mean (see Table 7). Three (3,7,11) of the four channels (3,7,8,11) selected as the optimum set of four by the transformed divergence measure were ranked in the top 3 by the data spread measure and in the top four if one considers the standard deviation measure. This indicates that the noise level of the data may have had a bearing on the channels selected for the optimum classification performance.

Also the two "quality" measures, the standard deviation and data spread measures, seem to indicate that channel 8 is an important spectral band to complement channels 3,7, and 11 for this set of data and classes. Even though channel 8 ranks 6th and 8th in the two quality measures, the channel is still chosen

Table 7

Data Quality Measures

Obtained from a Portion of Lake Monroe

SKYLAB MSS (S192)

Channel	Mean	Std. Dev.	Data Spread*($x10^{-3}$)
2	98.93	6.41	65.
3	53.34	3.39	64.
4	44.85	3.95	88.
5	36.35	3.47	95.
6	34.73	4.63	133.
7	42.86	1.90	44.
8	38.68	3.64	94.
9	33.21	2.79	84.
10	49.43	4.11	83.
11	44.57	1.84	41.
12	3.48	3.92	1,111.
		ERTS MSS	
1	66.00	1.58	24.
2	43.26	1.54	36.
3	32.16	2.58	80.
4	19.80	2.56	130.

*Std. Dev./Mean

to be one of the best 4 bands to use. The transformed divergence measure of the separability of classes using just one channel also indicates that channel 8 is important. (See Table 8). Channel 8 ranks 3rd after channels 11 and 10 for the highest average transformed divergence of all class combination.

ERTS MULTISPECTRAL SCANNER DATA ANALYSIS

The ERTS MSS data collected the day before was also classified using the same training areas to represent the classes used. The quality of the data for the study area was good except for one bad

Table 8

Separability of Classes
Using One Channel

Channel	Transformed Divergence Average of 77 Class Combinations*
11	1588
10	1476
8	1441
9	1394
7	1375
6	1344
12	1289
3	1273
5	990
4	947
2	868

*2000 is maximum value of separability.

data line and clouds in the south-eastern part of the area (see fig. 13). Because of the clouds, it was not possible to obtain a good representative commercial-industrial class from Bedford, so the training field for this class was taken from Bloomington 45 km. to the north. A class for clouds and a class for cloud shadows were used in this classification; however, these classes were not considered in the evaluation. The clustering procedure indicated that none of these classes needed to be divided to obtain unimodal distributions for each spectral class. All four channels were used for this classification.

The same six test areas were selected again and the same procedure as stated before was used to obtain the classification performance results (see Table 6 and fig. 14). The overall

Table 9

INTERCLASS TRANSFORMED DIVERGENCE MATRIX
SKYLAB MSS (S192)
Channels 3,7,8,11

Average Transformed Divergence-1984 Minimum Transformed Divergence-1776

	Old Res	New Res	Com Ind	Extrac	Lt Soil	Dk Soil	Grass	Sprs Wd	Decid.	Conif.	River
New Res	1831										
Com Ind	1982	1999									
Extrac	2000	2000	2000								
Lt Soil	1999	1966	1983	1997							
Dk Soil	1916	1997	1776	2000	2000						
Grass	1927	1999	2000	2000	1995	2000					
Sprs Wd	1820	2000	2000	2000	2000	2000	1929				
Decid	1999	2000	2000	2000	2000	2000	1867	1995			
Conif	1978	2000	2000	2000	2000	2000	1987	1978	1849		
River	1998	2000	1998	2000	2000	2000	2000	2000	2000	1982	
Lake	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000

Table 10

INTERCLASS TRANSFORMED DIVERGENCE MATRIX
SKYLAB MSS (S192)
Channels 3,5,6,8

Average Transformed Divergence-1937

Minimum Transformed Divergence-1160

	Old Res	New Res	Com Ind	Extrac	Lt Soil	Dk Soil	Grass	Sprs Wd	Decid	Conif	River
New Res	1921										
Com Ind	1961	1998									
Extrac	2000	2000	2000								
Lt Soil	1880	1160	1839	1989							
Dk Soil	1780	1996	1309	2000	1978						
Grass	1893	1997	2000	2000	1930	2000					
Sprs Wd	1913	2000	2000	2000	2000	1999	1894				
Decid	1987	2000	2000	2000	1999	2000	1264	1865			
Conif	1985	2000	1999	2000	1998	1986	1974	1746	1759		
River	1984	2000	1367	2000	2000	1779	2000	2000	2000	1999	
Lake	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000

Table 11

INTERCLASS TRANSFORMED DIVERGENCE MATRIX
ERTS MSS
Channels 1,2,3,4

Average Transformed Divergence-1989 Minimum Transformed Divergence-1759

	Old Res	New Res	Com Ind	Extrac	Lt Soil	Dk Soil	Grass	Sprs Wd	Decid	Conif	River
New Res	1759										
Com Ind	2000	2000									
Extrac	2000	2000	2000								
Lt Soil	2000	2000	1997	1911							
Dk Soil	2000	1999	1788	2000	1992						
Grass	1999	2000	2000	2000	2000	2000					
Sprs Wd	2000	2000	2000	2000	2000	2000	1826				
Decid	2000	2000	2000	2000	2000	2000	2000	2000			
Conif	2000	2000	2000	2000	2000	2000	2000	1999	1896		
River	2000	2000	1888	2000	2000	1977	2000	2000	2000	2000	
Lake	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000

Table 12

INTERCLASS TRANSFORMED DIVERGENCE MATRIX
SKYLAB DIGITIZED COLOR IR (S190A)
Channels 1,2,3

Average Transformed Divergence-1924

Minimum Transformed Divergence-352

	Old Res	New Res	Com Ind	Extrac	Lt Soil	Dk Soil	Grass	Sprs Wd	Decid	Conif	River
New Res	1663										
Com Ind	1916	1843									
Extrac	2000	2000	2000								
Lt Soil	1997	1837	1781	1867							
Dk Soil	1797	1498	352	2000	1910						
Grass	1746	1962	2000	2000	2000	1997					
Sprs Wd	2000	2000	2000	2000	2000	2000	1699				
Decid	2000	2000	2000	2000	2000	2000	1986	1870			
Conif	2000	2000	2000	2000	2000	2000	2000	1989	1871		
River	1461	1999	1972	2000	2000	1985	1985	2000	2000	2000	
Lake	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000

Table 13

INTERCLASS TRANSFORMED DIVERGENCE
SKYLAB DIGITIZED MULTIBAND PHOTOGRAPHY (S190A)

Channels 1,2,3,4

Average Transformed Divergence-1972 Minimum Transformed Divergence-1028

	Old Res		NewRes	ComInd	Extrac	LtSoil	DkSoil	Grass	SprsWd	Decid			Conif	River
	1	2								1	2	3		
NewRes	1881	1998												
ComInd	2000	1948	1998											
Extrac	2000	2000	2000	2000										
LtSoil	2000	2000	1958	1982	1974									
DkSoil	2000	1993	1999	1028	2000	1853								
Grass	2000	2000	1971	2000	2000	2000	2000							
SprsWd	2000	2000	2000	2000	2000	2000	2000	1858						
Decid1	2000	2000	2000	2000	2000	2000	2000	2000	1935					
2	2000	2000	2000	2000	2000	2000	2000	2000	2000					
3	2000	2000	2000	2000	2000	2000	2000	2000	2000					
Conif	2000	2000	2000	2000	2000	2000	2000	2000	1988	2000	2000	1992		
River	2000	1891	1995	1388	2000	1998	1556	1999	2000	2000	2000	2000	2000	
Lake	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000

classification performance was good-88%. The worst cases were the commercial-industrial, the extractive, and the river classes. The commercial-industrial class was confused with the light soil class. The river class was confused with the commercial-industrial class. The transformed divergence between these pairs of classes were some of the lower ones in the sets.

The same measures used for the SKYLAB MSS data set to obtain a quantitative handle on the noise quality of the data were used for the ERTS MSS (see Table 7). The four areas were selected to match those used in the SKYLAB MSS as near as possible; the means and standard deviations have been corrected to match the quantization level used by the SKYLAB MSS. (The quantization level of the SKYLAB MSS is 256, that of the ERTS is 128 for channels 1,2, & 3 and 64 for channel 4). The two visible channels in the ERTS MSS by these measures are not as noisy as the infrared channels.

SUMMARY

In comparison of the classification performances for the four data sets, the overall performance for the MSS data sets were better than those for the digitized photographic data sets. The classification performance of the digitized color IR data set was better overall than the digitized black and white multiband data set. Also the overall performance for the "optimum" four channels in the SKYLAB MSS data set was essentially the same as that for the ERTS MSS data set. However, when the four channels in the SKYLAB MSS which most nearly correspond to those in the ERTS MSS were used to classify the study area, the overall performance for the SKYLAB MSS data set was significantly lower than that for the ERTS MSS.

The difference in the performances for the ERTS MSS and the ERTS simulated SKYLAB MSS may be due to the different noise levels of the two MSS sets as well as the possible difference in information content of the two sets since the spectral windows of the four channels don't match entirely. In comparison of the data quality measures for the two MSS sets (see table 7), the standard deviations for each of the channels in the ERTS MSS are lower than those for the corresponding channels in the SKYLAB MSS. If the data spread measure is used, three of four ERTS MSS channels are better than the corresponding SKYLAB MSS channels. These measures indicate that the SKYLAB MSS is noisier than the ERTS MSS and they also suggest that one must be careful in concluding that one spectral band is better than another for delineating different classes since the decision may be biased by the noise content of channels introduced by the optics and electronics of the scanner. More work needs to be done to obtain a better quantitative comparison of the noise in the two MSS.

For this study the expected gains from the better spectral resolution of the multiband photography were apparently offset by the problem of "flare" around the lake in the infrared spectral bands and the inability to spatially register the four bands exactly. In other areas and/or different uses of multispectral data these problems may not be significant. As has been reported in other studies [6,7], there seems to be little difference in digitized multiemulsion and multiband photography. A useable digitized multiemulsion photographic data set however is much easier to obtain than a digitized multiband data set since the registration process is not needed.

In comparing the individual class results, there is one case where the addition of a middle infrared channel helped significantly in the classification, in delineating river from commercial-industrial and dark soil. The classification results for the river class was significantly higher in the channels 3,7,8,11 combination than for any other data set. The interclass transformed divergence measures for the separability of these two combinations were the highest in the channels 3,7,8,11 combination (see Tables 9-13). In the ERTS and SKYLAB MSS data sets, the transformed divergence measurements using just one channel indicates that the middle infrared spectral bands (channels 11 & 12) for the SKYLAB MSS are the only channels to give a separation better than 1552 for the two combinations being considered. This was born out of the classifications; if a middle infrared channel was not used, there was much confusion separating river from commercial-industrial and dark soils. Caution must be taken when comparing the MSS data sets with the digitized photographic data sets because of the difference in the ground resolution of the two and the rather narrow river. The digitized color infrared photography separated the river from commercial-industrial and dark soil better than the ERTS MSS and the digitized multiband photography but did a much poorer job separating river and old residential.

Many of the classifications errors in all the data sets are due to boundaries where the resolution elements consisted of the energy from two different classes. Around Lake Monroe the averaging of the water and the forest results in a resolution element most similar to the coniferous forest or the commercial-industrial classes. Also, around the light soil areas adjacent to grass, the averaging of the two classes results in resolution elements most similar to the new residential class and light soil areas adjacent to forest cover results in resolution elements most similar to the old residential. The boundary classes may be separable from the classes they were assigned to; however, this point wasn't examined in this study.

It is also believed that the difference in the size of the resolution element for the MSS's on ERTS and SKYLAB and the digitized photography had had a bearing on the classification performances. Because of the smaller resolution element for the digitized photography (see Table 14) each field is represented by more data points. The ratio of the number of interior data points to the number of boundary points is greater for the digitized photography than for the MSS data. This may have been a help in the overall classification performance. As an example, for the lake class the digitized photography did better than the MSS data sets. This is misleading because the reason for the difference in performance is the greater number of interior points compared to the number of boundary points. To put this point in another perspective, it is felt that if the photography had been digitized for the same ground resolution as that of the MSS's on ERTS and SKYLAB, there would be a larger difference in the classification performances. This was not examined however.

Table 14

Ground Representation of Sample Elements

Data Set	Sample Size (m)	Area (hectares)
SKYLAB MSS (S192)	68x72	.49
ERTS MSS	59x79	.47
Color IR	57x57	.32
B&W Multiband	57x57	.32

This was not an optimum study in the sense that the best classification performances were obtained. Using more of the non-supervised classification approach [3] and creating each set separately (not using the same training areas from set to set) better performances may be obtained for some or all of the multispectral data sets. Other spectral classes may probably be recognized that were not in the training areas used in this study.

At this time, the SKYLAB MSS (S192) data is being "filtered" at the Johnson Space Center to reduce the noise. When this has been completed the analysis will be done again using the filtered data set to find if there are any significant changes from the original data set. Also, the thermal infrared channel will be available in the filtered data set so that another important spectral region can be examined. Further work should also be done using other channel combinations.

REFERENCES

[1] J. R. Anderson, E. E. Hardy and J. T. Roach, "A Land-Use Classification System for Use with Remote Sensor Data," U.S. Geological Survey Circular 671, 1972.

[2] T. L. Phillips, ed., LARSYS Version III User's Manual, Vol. 2, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana, 1973.

[3] P. H. Swain, "Pattern Recognition: A Basis for Remote Sensing Data Analysis," LARS Information Note 111572, Purdue University, West Lafayette, Indiana, 1972.

[4] T. Marill and D. M. Green, "On the Effectiveness of Receptors in Recognition Systems," I.E.E.E. Trans. on Information Theory, Vol. IT-9, pp. 11-17, 1963.

[5] A. E. Potter, et al "Summary of Flight Performance on the Earth Resources Experiment Package (EREP)," Proc. of the Ninth International Symposium of Remote Sensing of Environment, Ann Arbor, Michigan, April 15-19, 1974.

[6] P. E. Anuta, et al, "Crop, Soil, and Geological Mapping from Digitized Multispectral Satellite Photography, "Proc. of the Seventh International Symposium on Remote Sensing of Environment, Ann Arbor, Michigan, Vol. 3, pp. 1983-2016, May 17-21, 1971.

[7] W. G. Brooner, R. M. Haralick and I. Dinstein, "Spectral Parameters Affecting Automated Image Interpretation Using Bayesian Probability Techniques, Proc. of the Seventh International Symposium on Remote Sensing of Environment, Ann Arbor, Michigan, Vol. 3, pp. 1929-1940, May 17-21, 1971.

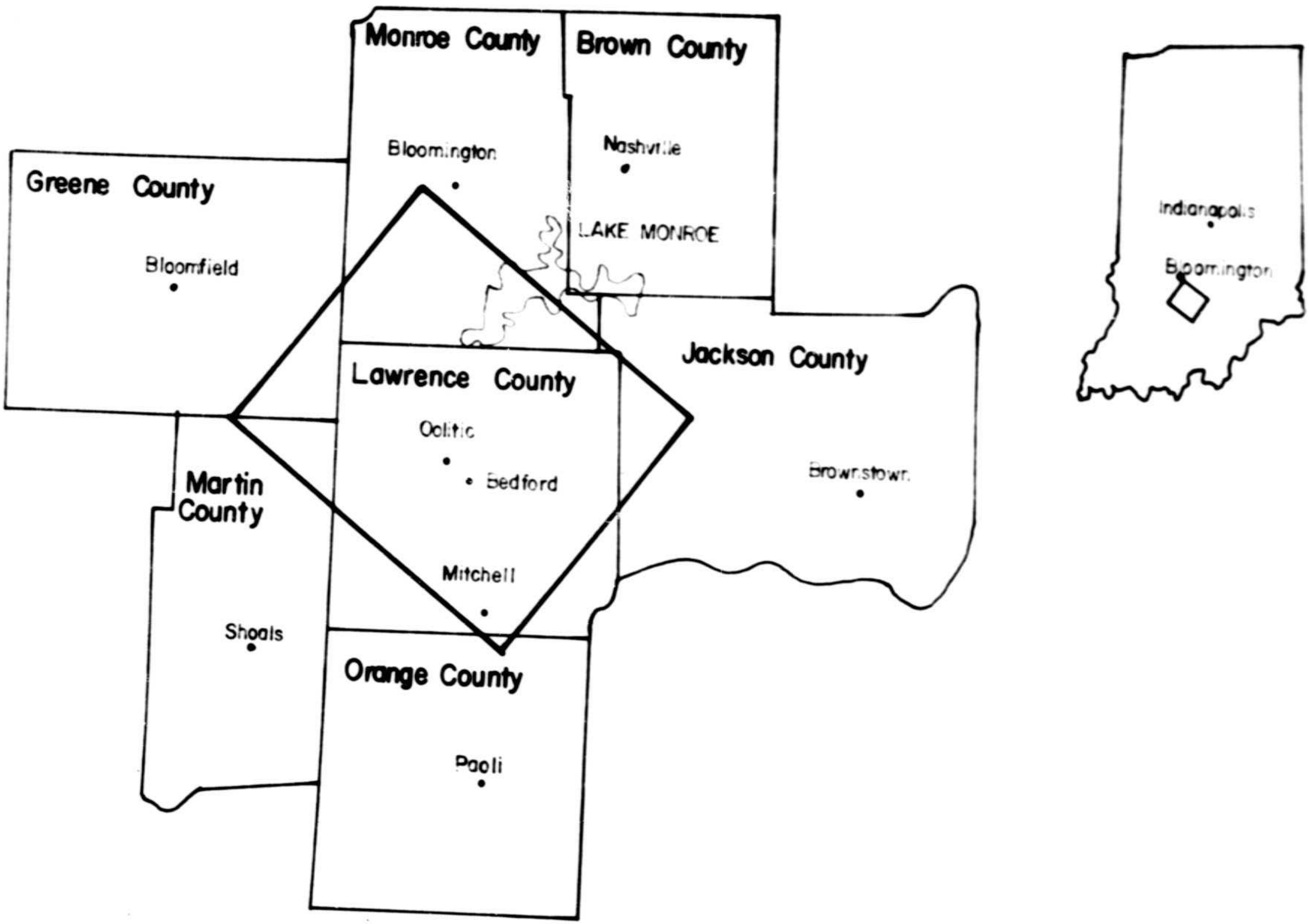


Fig. 1. Position of study area in south central Indiana. Study area outlined by bold black rectangle.

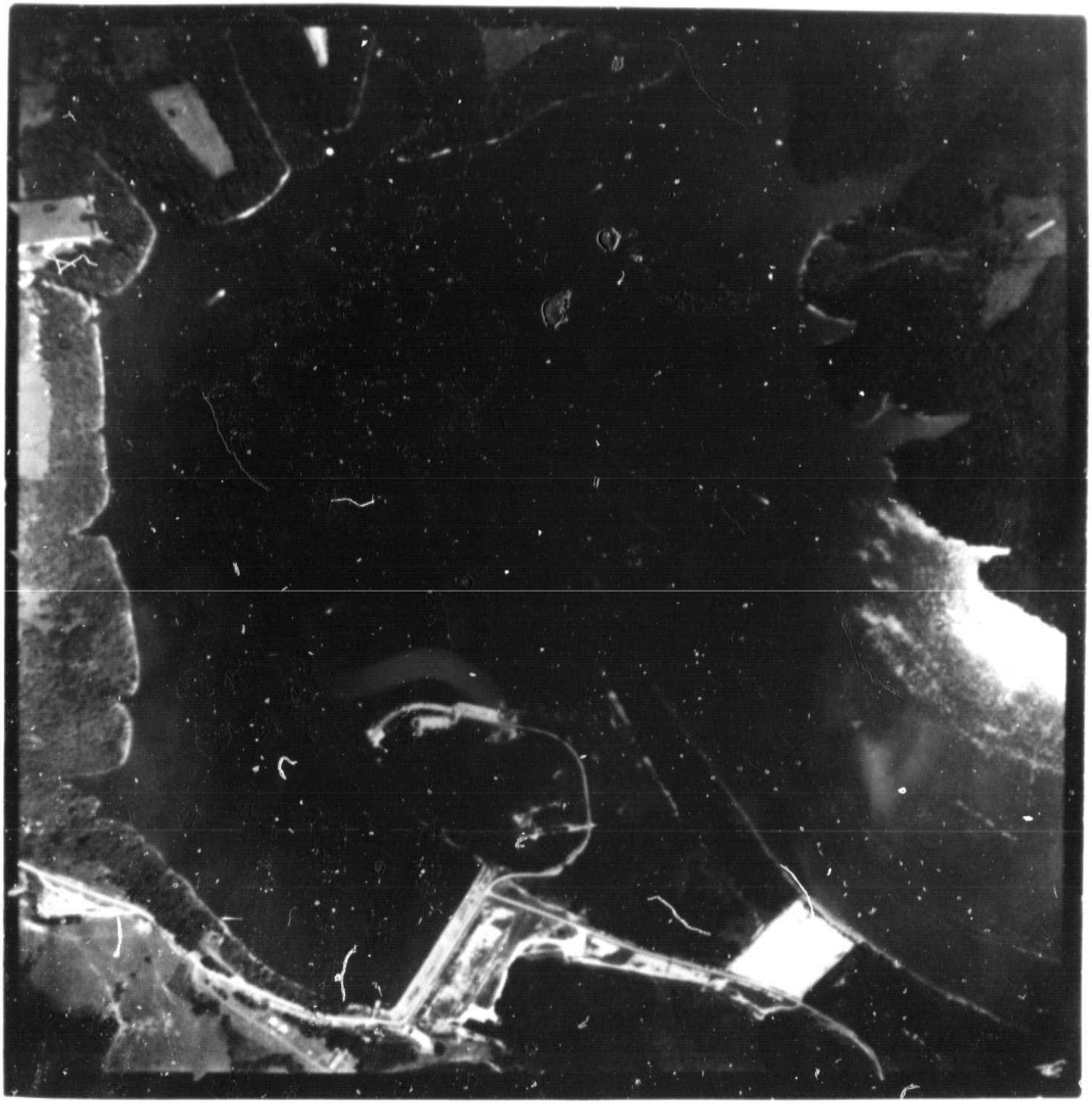


Fig. 2. Frame from the Michigan (ERIM) underflight photography over the Monroe Lake dam. The field spectroradiometer system was set up on the ramp (visible in this frame) going into the lake just a few hundred feet upstream from the dam. Altitude-1520 meters.

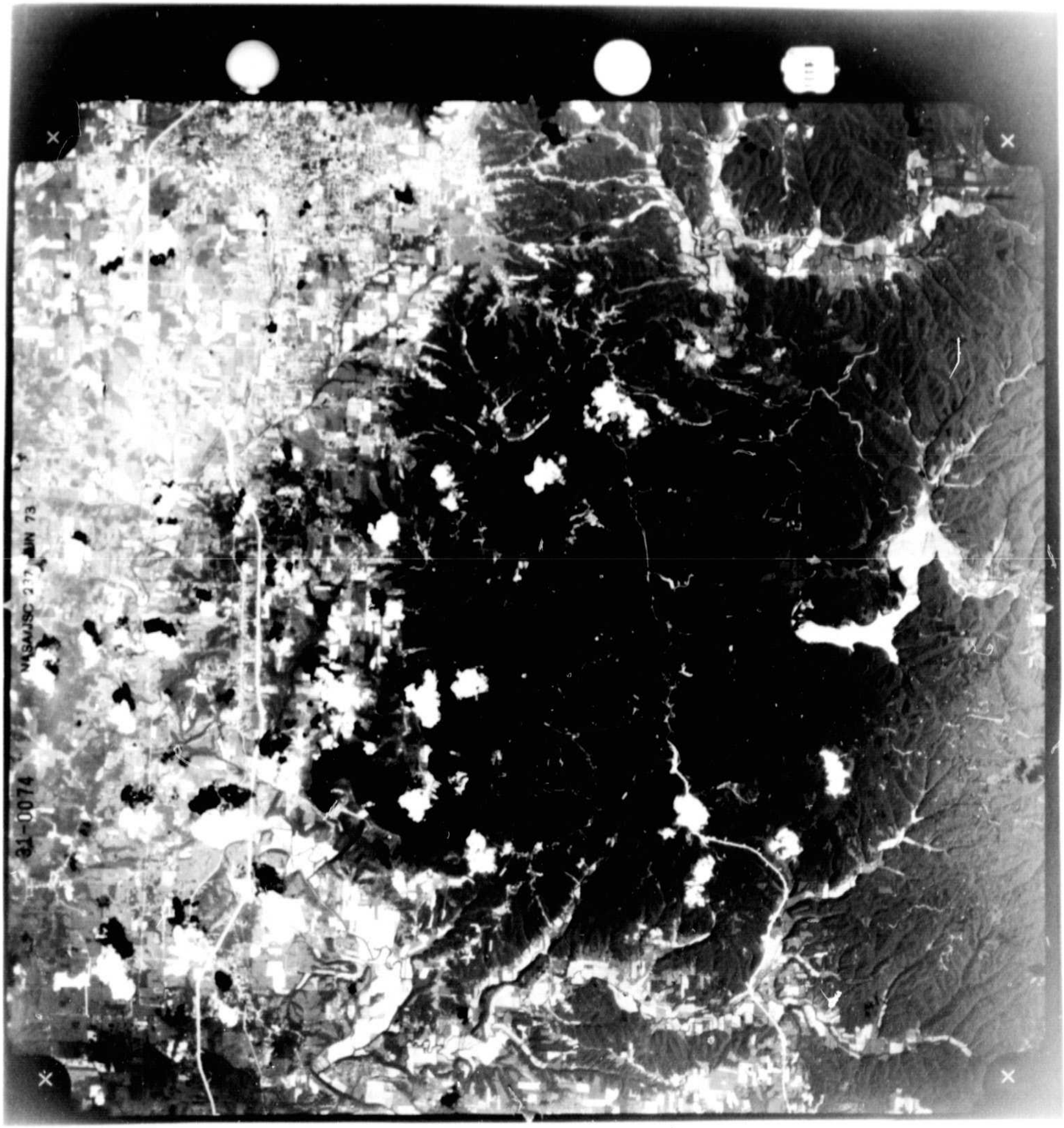


Fig. 3. Black and White reproduction of color IR frame from WB57 underflight photography. Lake Monroe and U.S. Highway 37 are easily seen. The town in the upper left is Bloomington, Ind. Altitude-18,300 meters.

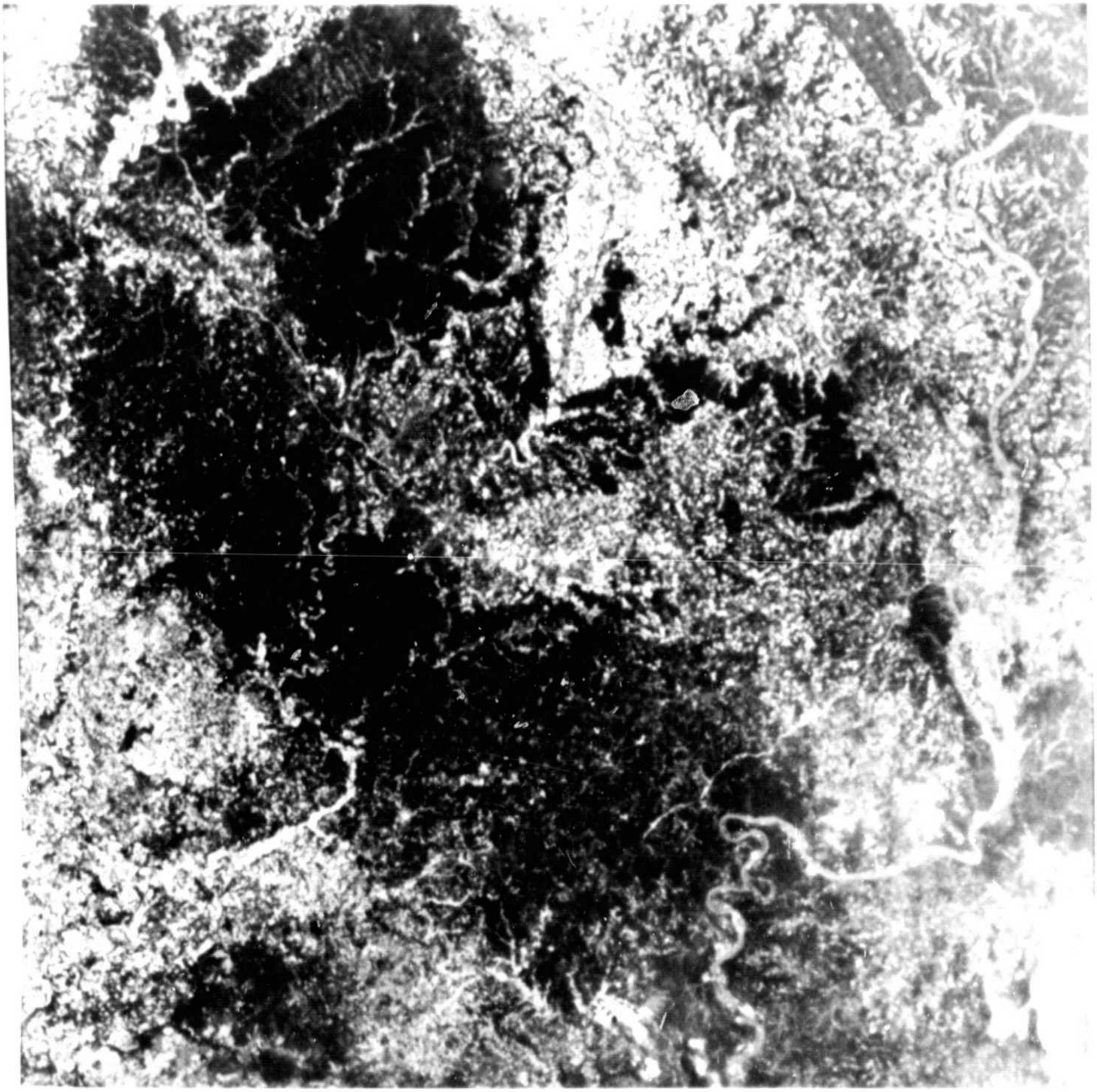


Fig. 4. Black & white reproduction of SKYLAB S190A color IR frame that was digitized. Study area is in upper left quarter of frame. The Ohio River runs along right edge of frame. Altitude-440 km.

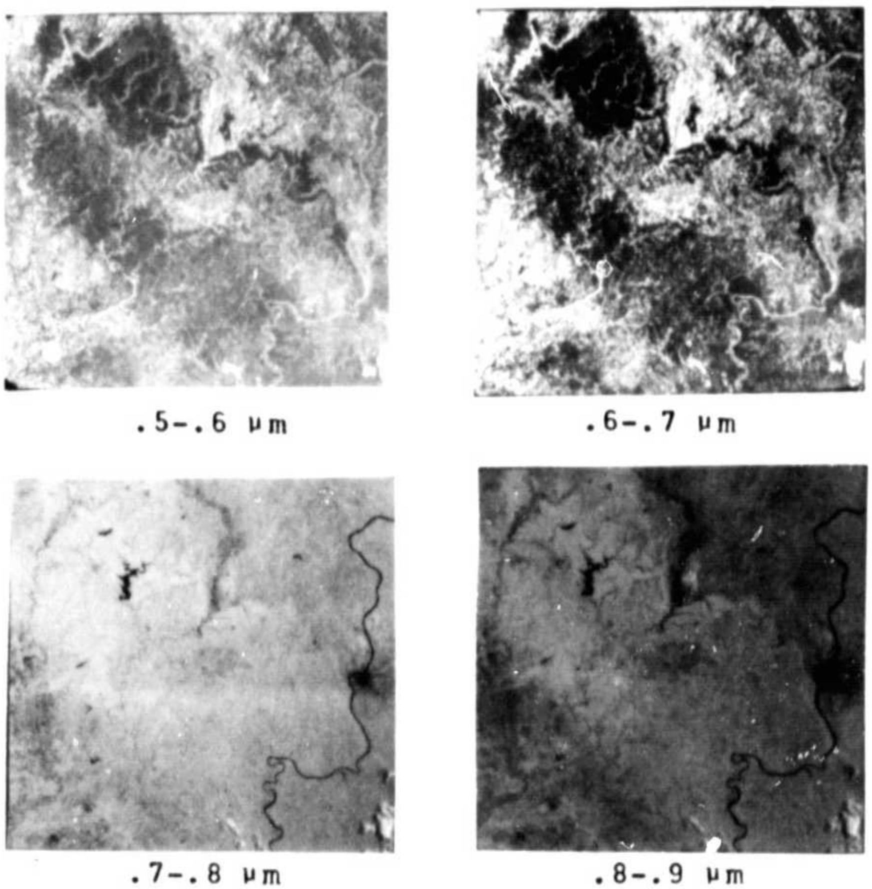
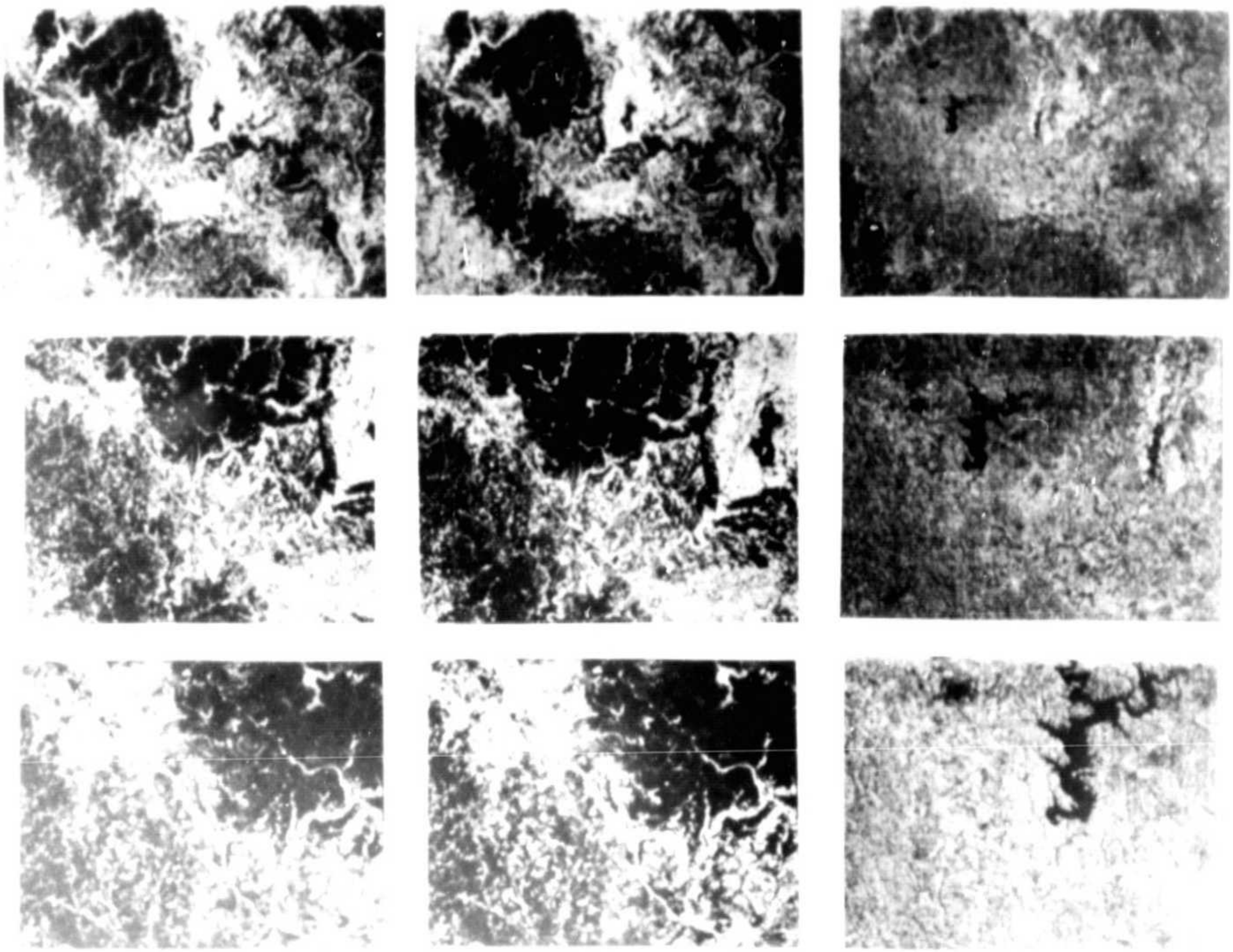


Fig. 5. SKYLAB S190A black & white multiband frames that were digitized. Study area is in upper left quarter of frame. Altitude-440 km.



.52-.58 μm* .58-.68 μm* .68-.88 μm*

Fig. 6. Digitized SKYLAB S190A color IR multiemulsion data set obtained from digital display screen using 16 different gray levels. Top row-complete digitized frame. Middle row - enlargement of study area (See Fig. 7 for exact boundary of study area). Bottom row - enlargements within study area.
*See note in Table 4.

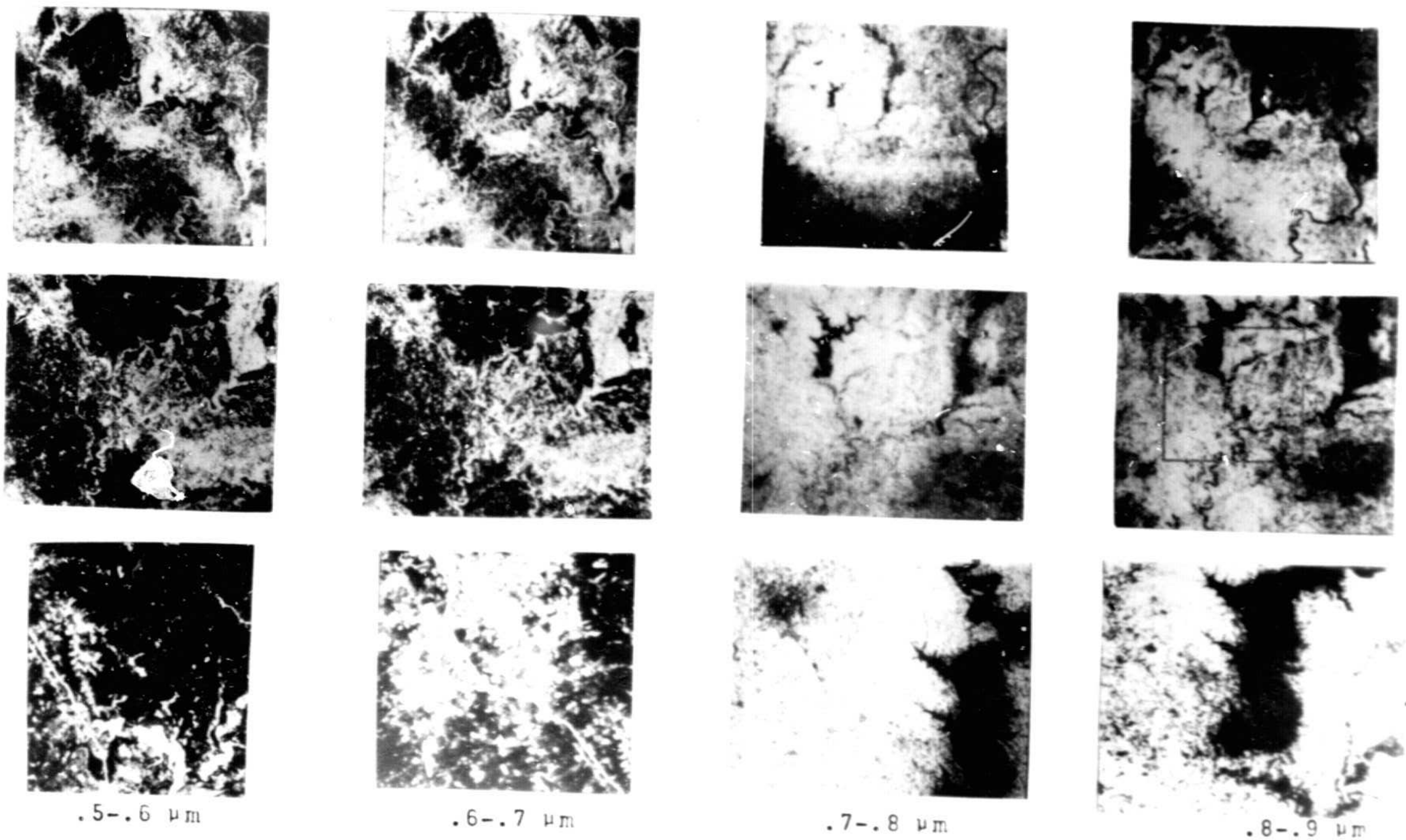


Fig. 7. Digitized SKYLAB S190A black & white multiband data set obtained from digital display screen using 16 different gray levels. Top row-complete digitized frames. Middle row-enlargement of study area. Study area is outlined in the .8-.9 μm band. Bottom row-further enlargements of study area. Note difference in resolution between visible and infrared frames.

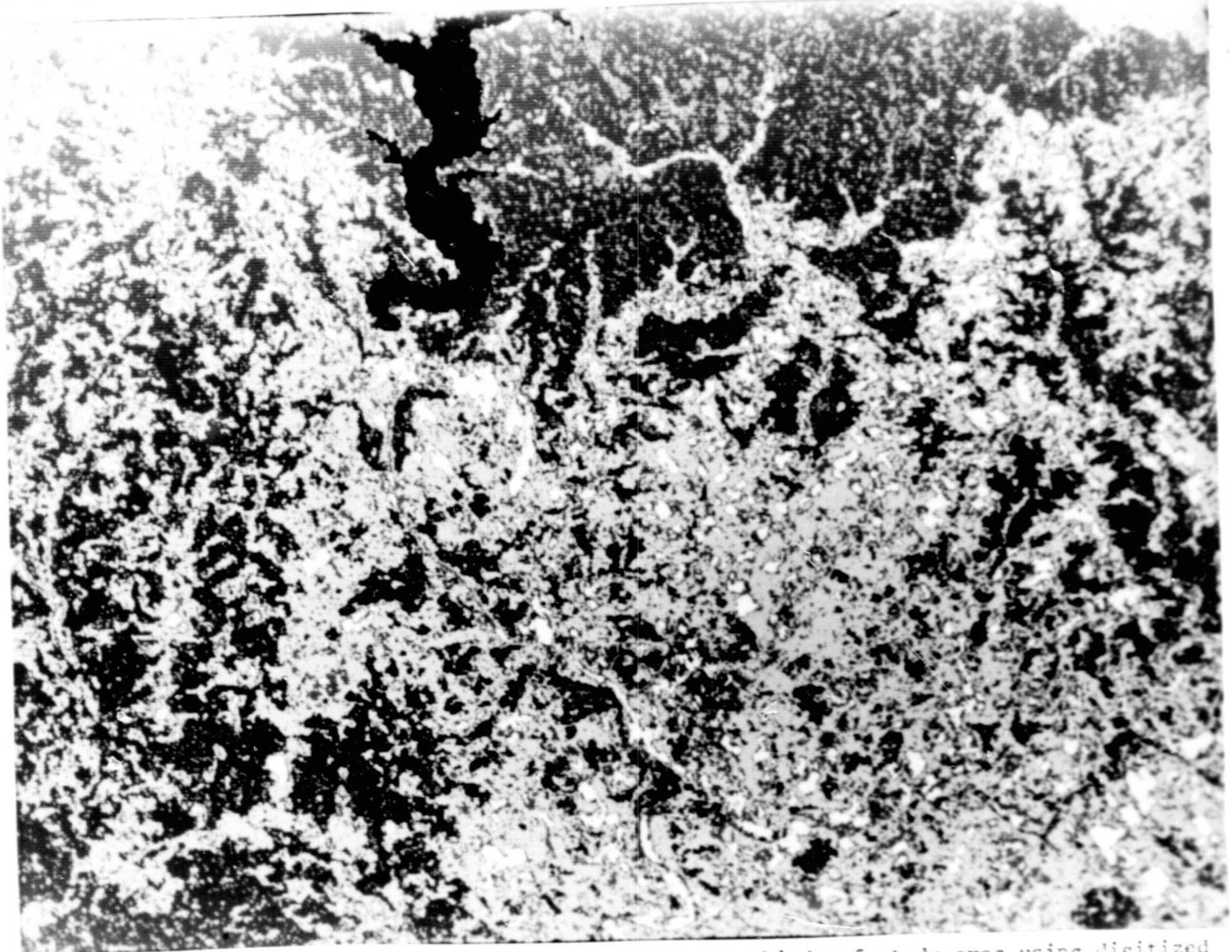


Fig. 8. Grey level coded classification map of two-thirds of study area using digitized S190A color IR data set. Nine land use classes represented: in order of increasing brightness-Lake, River, Commercial-Industrial, Deciduous Forest, Residential, Coniferous Forest, Grass, Bare Soil, and Extractive.

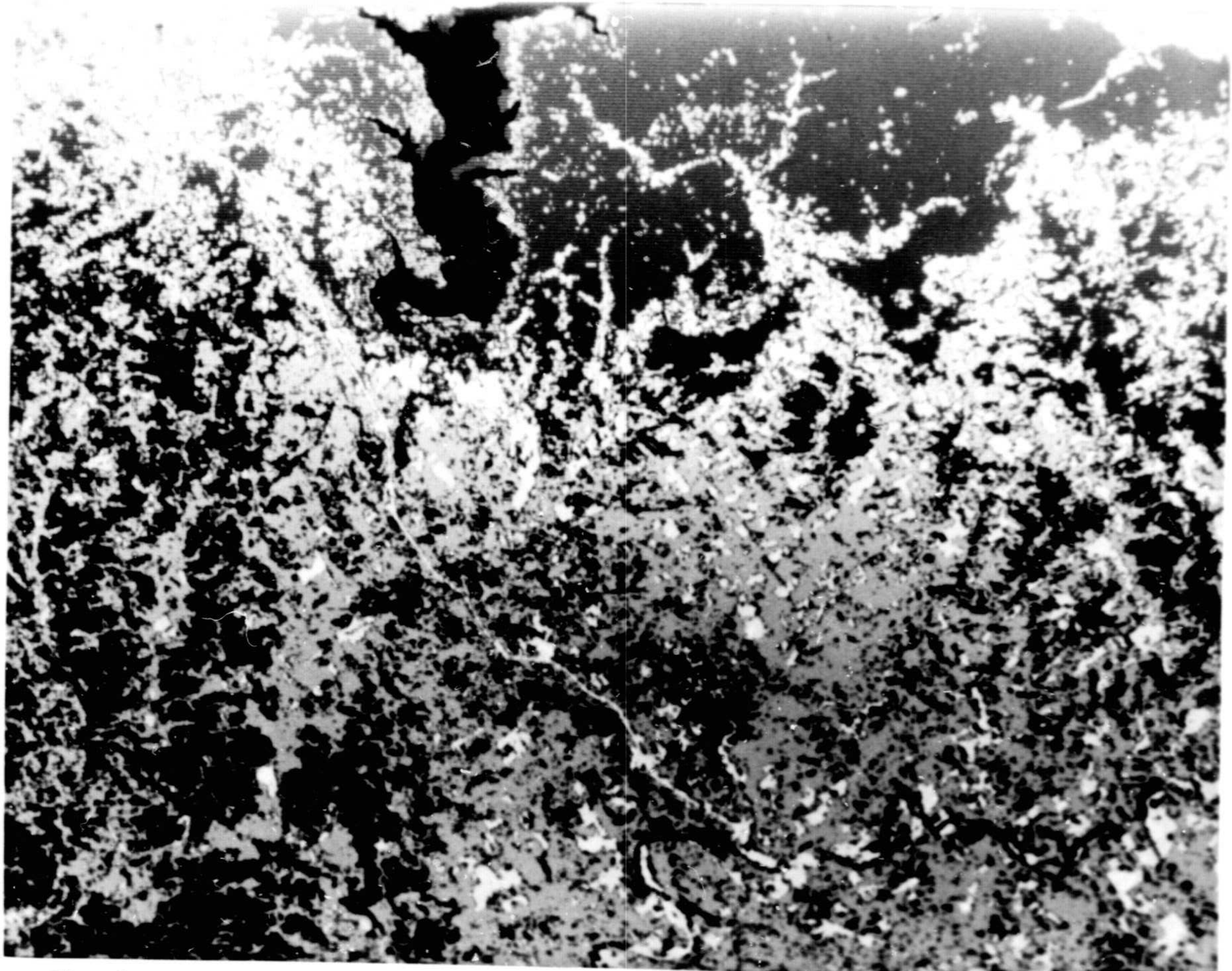


Fig. 9. Grey level coded classification of two-thirds of study area using digitized S190A black and white multiLand data set. Nine land use classes represented: in order of increasing brightness-Lake, River, Commercial-Industrial, Deciduous Forest, Residential, Coniferous Forest, Grass, Bare Soil, and Extractive.

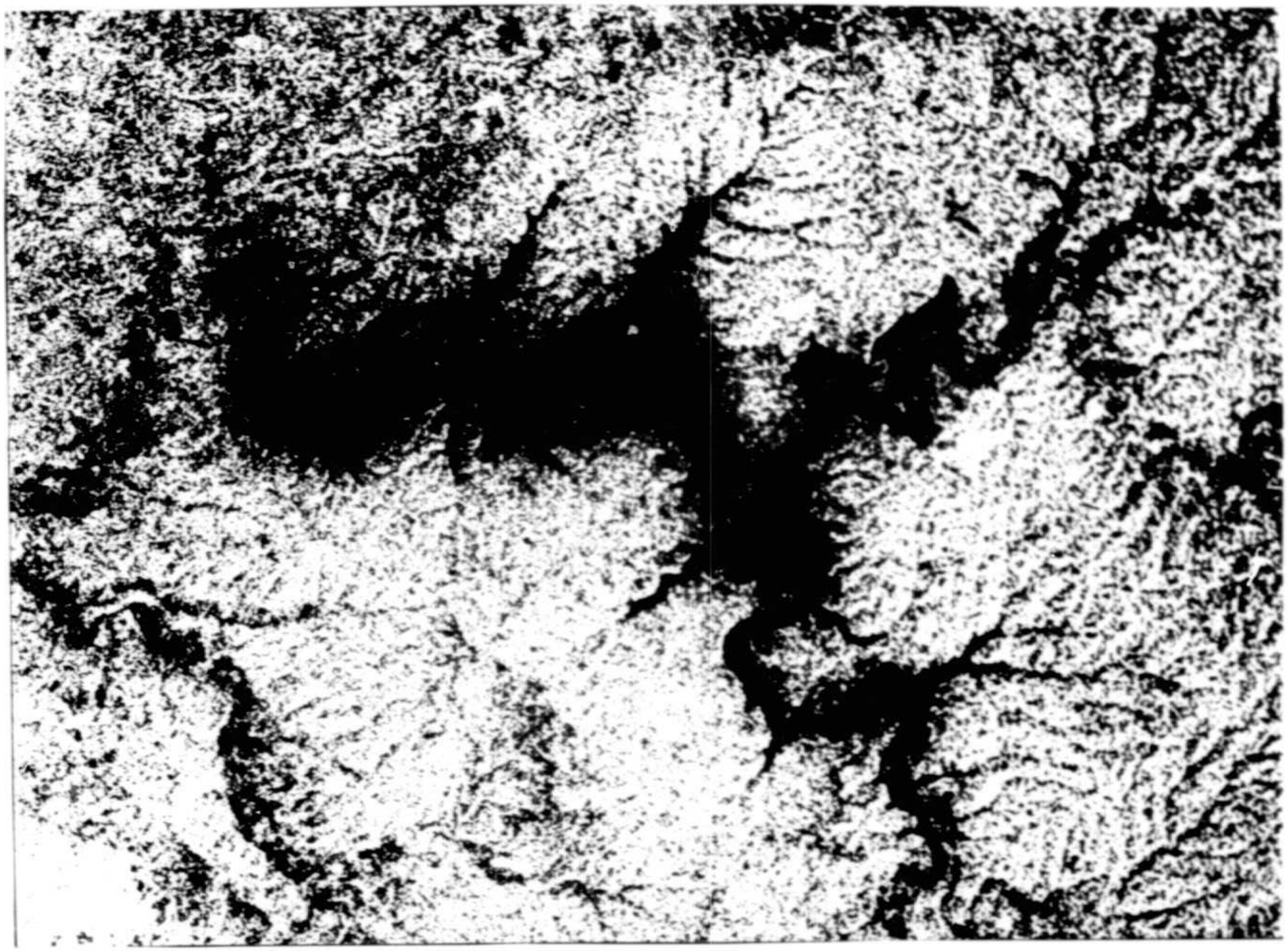


Fig. 10. Enlargement of the .8 - .9 μm black & white frame that was digitized, illustrating the "flare" around Lake Monroe.

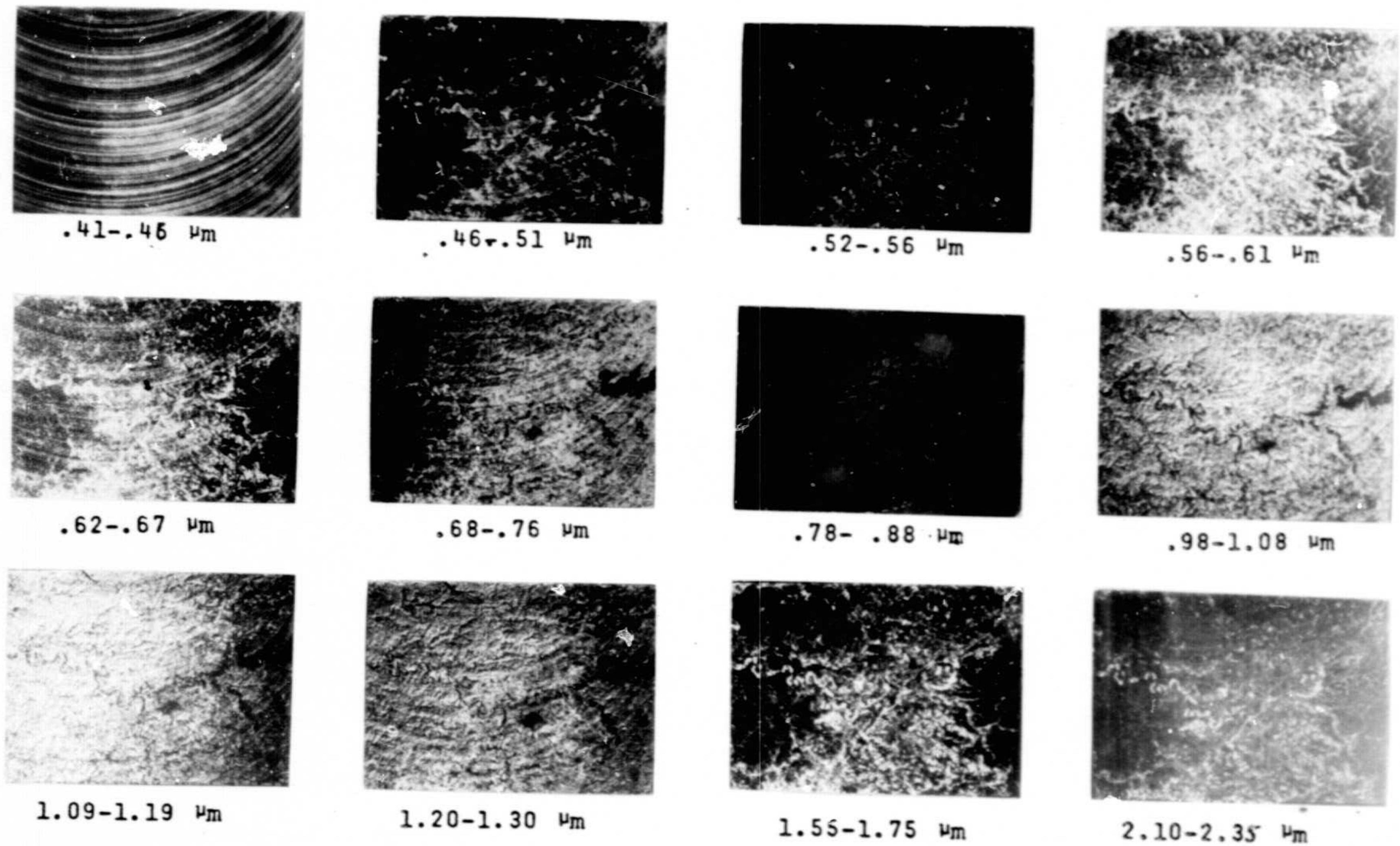


Fig. 11. SKYLAB MSS (S192) 12 channel data set obtained from digital display screen using 16 different gray levels. Lake Monroe, U.S. Highway 37, and the East Fork of the White River show up clearly. Altitude-440 km.

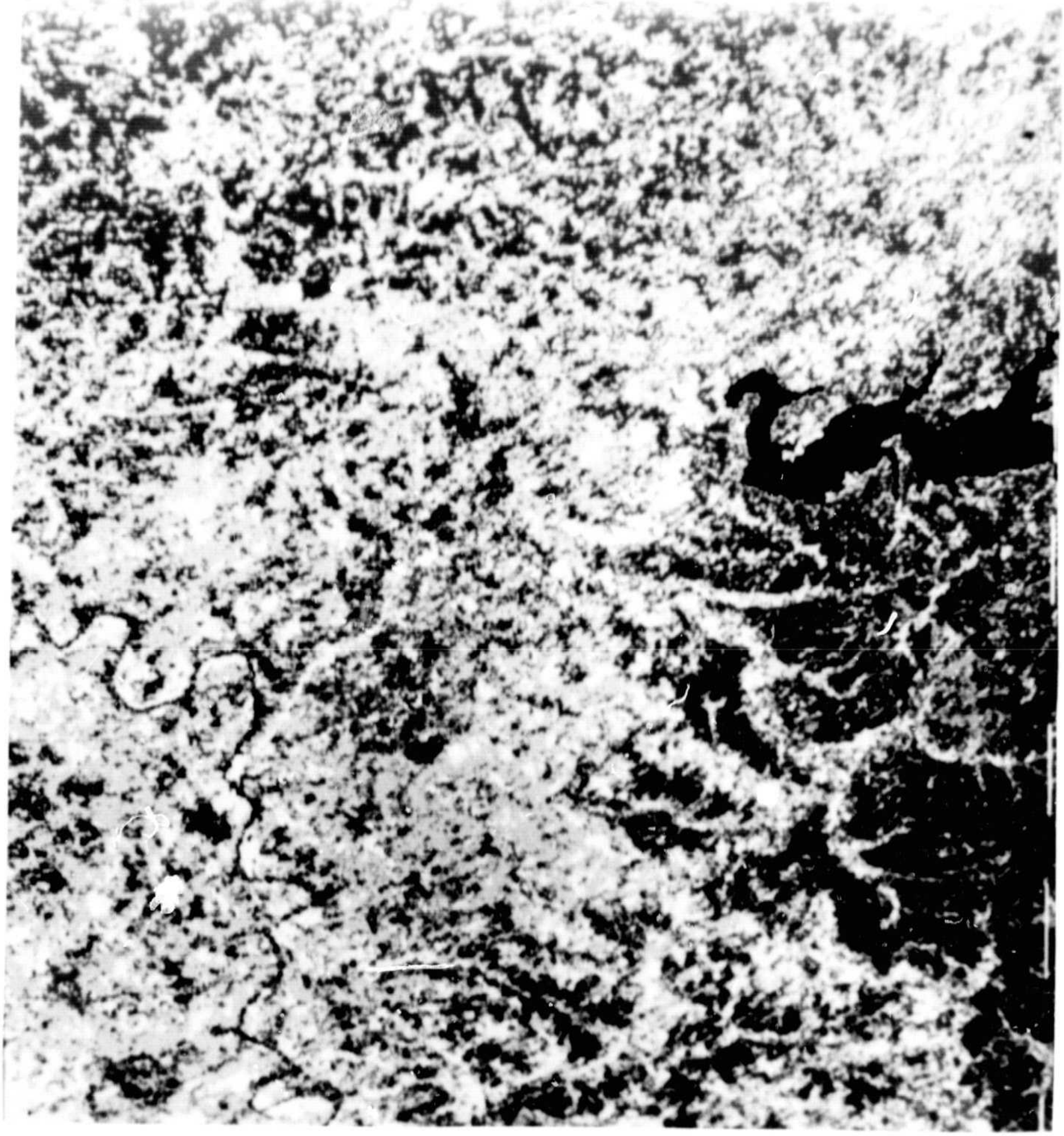


Fig. 12. Gray level coded classification of study area using SKYLAB MSS data set. Nine land use classes represented: in order of increasing brightness-Lake, River, Commercial-Industrial, Deciduous Forest, Residential, Coniferous Forest, Grass, Bare Soil, and Extractive. This is the Channels 3,7,8,11 classification.

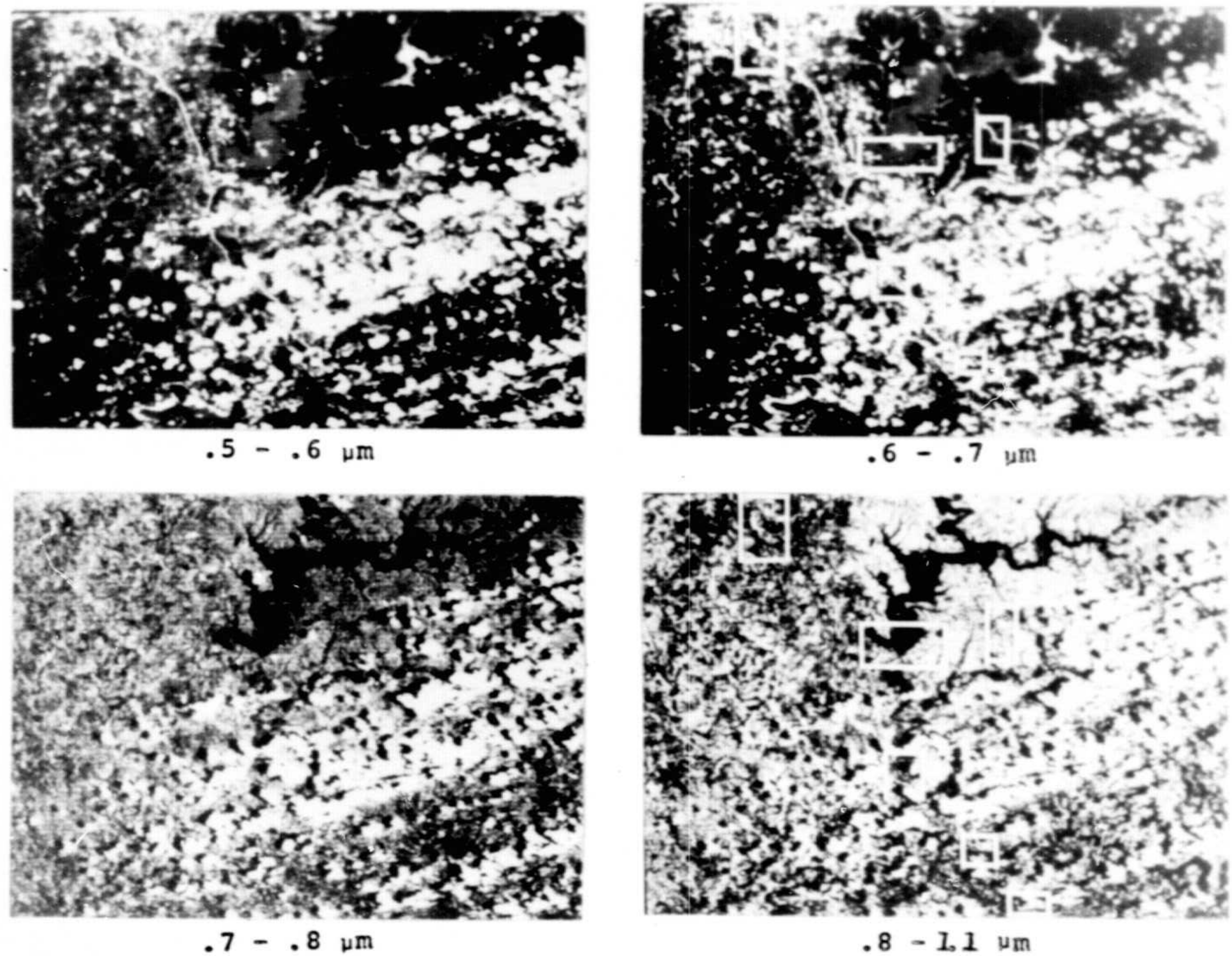


Fig. 13. ERTS MSS 4 channel data set obtained from digital display screen using 16 different grey levels. The six test areas used to obtain the classification performance results are outlined in the .6 - .7 μm and .8 - .9 μm bands. Altitude-933 km.

Fig. 14. Gray level coded classification of study area using ERTS MSS data set. Nine land use classes plus a cloud & cloud shadow class represented: in order of increasing brightness-Lake, River, Commercial-Industrial, Deciduous Forest, Residential, Coniferous Forest, Grass, Bare Soil, Cloud & Cloud Shadow, and Extractive. Horizontal line near top was caused by a bad data line in the original ERTS data.