Conference on

Machine Processing of

Remotely Sensed Data

October 16 - 18, 1973

The Laboratory for Applications of Remote Sensing

Purdue University West Lafayette Indiana

Copyright © 1973 Purdue Research Foundation

This paper is provided for personal educational use only, under permission from Purdue Research Foundation.

LAND USE CLASSIFICATION OF MARION COUNTY, INDIANA BY SPECTRAL ANALYSIS OF DIGITIZED SATELLITE DATA*

William J. Todd and Marion F. Baumgardner

Laboratory for Applications of Remote Sensing, Purdue University,
West Lafayette, Indiana and Indiana State University,
Terre Haute, Indiana; Laboratory for Applications of Remote Sensing,
Purdue University, West Lafayette, Indiana

I. ABSTRACT

Multispectral scanner data, obtained over Marion County (Indianapolis), Indiana at an altitude of 915 kilometers, were analyzed by computer-implemented techniques to evaluate the utility of satellite data for urban land use classification. Several land use classes, such as commerce/industry, single-family (newer) residential, trees, and water exhibited spectrally separable characteristics and were identified with greater than 90 per cent accuracy. Difficulties were encountered in the spectral separation of grassy (open, agricultural) areas and multi-family (older) housing. The confusion between these two classes was largely eliminated, however when spectral characteristics of samples (instead of individual data points) were considered. Another solution to the problem consisted of spatially dividing the data into urban and rural land uses prior to classification. Over 95 per cent accuracy of recognition may be achieved by this "pre-processing" step in an analysis.

II. INTRODUCTORY STATEMENTS

It would be valuable to a metropolitan area if land use information derived from computer analysis of data from the Earth Resources Technology Satellite (ERTS) could be stored on magnetic tape and be periodically updated with subsequent passes of the satellite. To date, land use classifications have been reasonably successful in urban areas (Ellefsen, Swain, and Wray, 1973; Todd, Mausel, and Wenner, 1973), but certain land use classes have been largely elusive to existing methods of classification. While many urban land uses exhibit spectrally separable characteristics, permitting accurate identification through application of a Gaussian maximum likelihood classifier, certain important land uses do not. The objectives in this analysis were to investigate further the spectral characteristics of these "problem" areas. The hypothesis tested, therefore, was whether areas of misclassification could be identified by numerical (spectral) characteristics other than single spectral class. Parameters such as mean, range, standard deviation, and correlation coefficients were the key components of the investigation

The Marion County (Indianapolis), Indiana subframe was selected as the study area. Four bands of digitized, multispectral data from the ERTS pass of 30 September 1972 (Observation ID 106915585) were analyzed by computer processing to test the hypothesis.

^{*}This work was supported by NASA Grant NAS 5-21773.

III. DATA PROCESSING

The Marion County data were initially viewed on a digital imaging display (Figure 1) for purposes of orientation. At that time, small areas were defined for two subsequent analyses. Three areas were defined for histogramming (Wacker and Landgrebe, 1971). Those areas were submitted to a histogramming processor to obtain a deck of histogram cards for future, more controlled, viewing on the digital display. Next, several small areas were defined for the clustering algorithm (Wacker and Landgrebe, 1971). The clustering algorithm was asked to find fourteen clusters in the data. Those cluster classes were used to classify all of the data points in the county, and the results were displayed by a line printer using different alpha-numeric symbols for each class.

The resulting cluster map of Marion County gave important clues to the spectral classes of urban land cover phenomena, but a better definition of the land use classes was needed. Rectangular training samples of various land uses were located on the map, and their line and column coordinates recorded. Samples were located for the following land uses: (1) single family residential, (2) multifamily residential, (3) grassy (open) areas, (4) trees, (5) commercial/industrial, (6) cloud, (7) cloud shadow and (8) water. The county was classified again, and the results displayed by a line printer with different alpha-numeric symbols for each class.

IV. AREAL DISTRIBUTION OF CLASSES

To illustrate the classification results with the map produced from the line printer is not practical because of its unwieldy size. Consequently, the results were displayed on the digital imaging display (Figure 2). Graylevels used for the spectral classes are as follows:

Commercial/industrial areas medium gray Multi-family (older) residential black Single-family (newer) residential white Wooded areas light gray Grassy (open, agricultural) areas dark gray Water black Clouds white Cloud shadows black

Several pairs or trios of classes have been given the same graylevel, but consideration of their areal distribution permits visual separation. Single-family (newer) residential and clouds are both white, but the clouds are all small (0.75 kilometers in diameter), of the cumulus variety, and have an associated shadow located approximately one kilometer to the northwest. Multi-family (older) housing, water, and cloud shadow are all displayed as black, but visual separation is also possible. Cloud shadows are associated with the white cumulus clouds. Water is largely limited to two large reservoirs, Eagle Creek (west central portion of image) and Geist (northeast corner), and to several large ponds. The area of multi-family (older) housing is located in the center of the county, surrounding the Central Business District (classified as commercial/industrial).

Single-family (newer) housing is a class consisting of residential areas developed primarily after World War II. Housing density is relatively low, and family incomes are moderate. Three large areas were classified as single-family (newer) housing:

			Approximate Bound	aries	
		North	South	<u>East</u>	West
1. 2. 3.	WEST EAST SOUTH	46th St. 62nd St. Edgewood Rd.	10th St. Washington St. County Line Rd.	Tibbs Ave. Church Rd. McFarland Rd.	I-465 Arlington Ave. Bluff Rd.

Roads (concrete) and lawns (grass) are the two principal types of ground cover responsible for the spectrally separable nature of this class. Not unusually, therefore, interstate highways, boulevards, and airport runways were classified into this class.

The spectral class called commercial/industry (displayed as medium gray) is characterized by the occurrence of rooftops and streets/parking lots. Any large building with an associated parking lot or a cluster of buildings along a thorough-fare will fall into this spectral class. The largest area classified as commerce/industry is the Central Business District of Indianapolis (central portion of image) and adjacent industrial areas. This area extends from approximately 20th Street on the north to Morris Street on the south, and from West Street on the west to College Avenue on the east. Other, smaller areas in the outer parts of the city were also classified as such; they include larger industrial establishments and shopping centers. All areas in this class are typified by a lack of green vegetation.

Multi-family (older) housing (shown as black) in Indianapolis occurs as a ring of land use surrounding the Central Business District. Older housing is bounded by 56th Street on the north, Troy Avenue on the south, Tibbs Avenue on the west, and Arlington Avenue on the east. At least 75 per cent of the structures in this area were built prior to World War II. Mature tree cover is a primary influence in the spectral responses from these areas, as are the closely spaced rooftops.

Grassy (open, agricultural) areas are found in the outer part of the county. This class includes cropland, pasture, and idle land in rural areas, as well as grassy features in urban areas, such as parks, golf courses, and cemeteries. Areas classified as trees are closely associated with the drainage pattern of the county, i.e., ponds and streams. The most extensive stands of trees are located around Geist and Eagle Creek Reservoirs. The areal distribution of water, clouds, and cloud shadows was discussed above.

V. THEORY OF LAND USE IDENTIFICATION

The land use classes developed in this study are identifiable only because they consist of a unique type of ground cover or a unique combination of ground cover types. Five of the eight classes--trees, grassy (open, agricultural) areas, water, cloud, and cloud shadow--are relatively homogeneous in character, i.e., they consist of a single land cover type. The other three classes--commercial/industrial areas, multi-family (older) residential, single-family (newer) residential--are mixtures of land cover types in various proportions. Rooftops and concrete are the primary constituents of the commercial/industrial class. The class multi-family (older) residential consists primarily of trees and rooftops. Finally, single-family (newer) residential has lawns (grass) and roads (concrete) as the principal components influencing its spectral nature.

Guidelines have been established for extracting land use information from remotely sensed data collected from space platforms (Anderson, Hardy, and Roach, 1972). Table 1 compares that system with the one developed in this study. Level I of the Anderson et al. system represents the land use categories which may be extracted from satellite data; information for Levels II, III, and IV must be acquired from air photos and other sources. A generally close correspondence exists between the Level I categories and the ERTS spectral classes. Indeed, more information was extracted from the satellite data than anticipated. Three sub-categories of Urban and Built-up Land were developed. One deviation from their scheme was the combination of the Urban land use "Open and Other" (to include parks, cemeteries, golf courses, and open areas) with Agricultural Land. It was not possible in this study to obtain a spectral differentiation between these two categories.

VI. ACCURACY OF CLASSIFICATION

It is difficult to obtain a quantitative expression of the accuracy of the Indianapolis classification, for several reasons. The most evident of these is the great number of data points which would have to be tested. In the classification image (Figure 2) there are approximately 441,000 data points, of which 60 per cent are in Marion County. Even if an overlay of existing land use were to be employed for test purposes, a large number of data points would overlap different land uses. Another problem is the "training field accuracy" sample statistic which may be reported for each spectral class. Different land uses comprise

various proportions of the total area from one metropolitan area to another. A sample size of 200 data points for the class commercial/industrial in two different counties may have very different weight, in consideration of overall classification accuracy, from one county to another. Thirdly, there exists the problem of testing sub-resolution features (features smaller than one resolution element). At the resolution of the ERTS scanner, only half of a river flowing through a county may be wide enough to be classified accurately. Were 50 per cent of the water data points "misclassified" in this example, or can be statistic be reported as 100 per cent correct, owing to the sub-resolution characteristic of half of the river?

With the above problems in mind, an attempt was made to assess the classification accuracy by a sampling method. Several rectangular areas, termed test fields, were located for each of the spectral classes and the accuracy determined (Table 2). Four of the eight classes--commerce/industry, single-family (newer) residential, woodland, and water--were identified with over 90 per cent accuracy. Cloud, cloud shadow, and multi-family (older) residential had correct recognitions in the 80 to 90 per cent range. Grassy (open, agricultural) areas were the most poorly identified--only 64.5 per cent correct recognition. Overall classification accuracy (mean of eight values) was 87.1 per cent. Elimination of error due to weather conditions at the time of data collection (cloud and cloud shadow classes) raises the accuracy slightly, to 87.5 per cent.

The classification accuracy was achieved utilizing only spectral information. No attention was given to areal information in the data, i.e., theoretical considerations of urban geography, growth, and planning. Many other scientists who utilize remote sensor data must deal with features that have near-random areal distribution, such as crop types, water pollution, rangeland qualities, and atmospheric conditions. The urban geographer, on the other hand, studies a surprisingly predictable type of areal phenomenon. Referring back to Table 2, it is very difficult for an urban specialist to accept the fact that 9.7 per cent of the older residential area of Marion County should have been confused with grassy or agricultural area. Conversely, he would find it just as difficult to believe that 25.4 per cent of the outlying, agricultural areas of the county should be classified as older, densely-populated residential areas.

Areal information could be introduced into the scheme of classifying urban land use. For purposes of simplification, the spectral classes may be divided into two general categories--urban and rural. The urban category would include the classes commerce/industry, multi-family (older) residential, and single-family (newer) residential; rural would include wooded areas, grassy (open, agricultural) areas, and water. Boundaries could be stored on the computer, delineating the urban-rural boundary in Marion County. Data points within an urban area, for example, could only be classified into one of three classes, commerce/industry, multi-family (older) residential, or single-family (newer) residential. Applying this theory to the test results in Table 2 gives the values in the extreme right column. Accuracy for each class is greater than 90 per cent, and the overall classification accuracy has been increased to 96.4 per cent.

VII. SPECIFIC INVESTIGATIONS

Most of the error in classification was attributed to the confusion between grassy (open, agricultural) areas and multi-family (older) residential areas (Table 2). Another problem of classification arose in two types of residential areas, neither of which could be separated as single spectral classes. One of these types may be referred to as a transitional residential area. It is located between areas classified as multi-family (older) residential, with 75 per cent or more if its structures having been built prior to World War II, and single-family (newer) residential, with 25 per cent or less of its structures having been built prior to the second world war. The transitional areas have housing of mixed age, 25 to 75 per cent of its structures having been built prior to World War II. The second type of residential area is found in the north-central part of the county, from County Line Road south to 56th Street and from Northwestern Avenue east to Interstate 465. Within this area are scattered residential developments, built after World War II, and consisting of upper-income families. Such areas are termed "vegetative residential".

Special investigations were made into the spectral nature of four types of land use--multi-family (older) residential, transitional residential, vegetative residential, and agricultural (grassy, open). A number of rectangular samples were chosen for each type and tested for classification accuracy. The results for multi-family (older) residential and grassy (open, agricultural) were listed in Table 2. Test results for the other two types of land use are listed in Table 3. Quantitative information for a representative number of the samples is listed in Table 4.

Classification results were not satisfactory for the above types of land use, using the Gaussian maximum likelihood classifier (Fu and Landgrebe, 1969), but evaluation of certain parameters (Table 4) did allow separation of land uses by sample. The means and standard deviations in the visible bands of the spectrum presented no evidence of separability between the land uses. In the infrared bands, however, certain of the land uses are separable. Vegetative suburban is readily separable from the other two residential land uses, because of the higher reflectance of the former. The highest sample mean (relative reflectance) in Band 6 (0.7-0.8 \mum) for either transitional residential or multi-family (older) residential is 35.14, while the lowest sample mean (relative reflectance) for vegetative residential in that band is 40.38. Similarly, the highest sample mean in Band 7 (0.8-1.1 μ) for either transitional residential or multi-family (older) residential is 20.97, while the lowest sample mean for vegetative residential in that band is 23.31. Multi-family (older) residential and transitional residential are not separable by application of these parameters.

Although sample means do not indicate separability of grassy (open, agricultural) from the other land uses, consideration of the sample standard deviations in the infrared bands does result in separability. Standard deviations of grassy (open, agricultural) samples are typically twice as large as standard deviations of the other land use samples in either Band 6 or Band 7. The largest sample standard deviation of a residential land use in Band 6 is 3.56, while the lowest of the grassy (open, agricultural) class is 6.13. Likewise, the largest sample standard deviation of the residential land uses in Band 7 is 2.26, while the lowest of the grassy (open, agricultural) samples is 4.48.

Coefficients of correlation were also investigated to determine if they could aid in the spectral separation of land uses. Only one, r_{67} , of the six correlations proved to be helpful. Reference to Table 4 indicates that the correlation between the two infrared bands was always +0.95 or greater (highly significant statistically) for the grassy (open, agricultural) samples. Conversely, the r_{67} for samples from other land uses was +0.83 or less.

VIII. CONCLUSIONS AND APPLICATIONS

Application of the Gaussian maximum likelihood classifier to ERTS multispectral data produced satisfactory results in Marion County, Indiana. Certain land use classes, such as single-family (newer) residential, commerce/industry, water, and wooded areas were identified with greater than 90 per cent accuracy. Three other land use classes, multi-family (older) residential, cloud, and cloud shadow, had classification accuracies in the 80 to 90 per cent range. Difficulties were encountered in classification of grassy (open, agricultural) areas, where accuracy attained was 65 per cent. The mean classification accuracy by class was 87.1 per cent.

Simple, spectral identification of land use may be adequate for gross land use inventories, but the urban-regional planner will probably require better recognition. Two supplemental suggestions for the present software were presented in this manuscript. One of these was the "zone" concept, whereby a differentiation is made between urban and rural land uses. It was shown that classification accuracies of over 95 per cent may be attained by this consideration. The other supplement to the existing classifier is the sample method of identification. By evaluation of sample means, standard deviations, and correlations, recognition of samples' land uses was virtually 100 per cent correct.

The purpose of ERTS investigations in urban areas is to (1) map land use with accuracy, and (2) develop a system of monitoring land use. It can be concluded from this study that the state-of-the-science is very near to reaching that first

goal. If the spectral and areal information can be utilized adequately, then the monitoring of land use should not be too difficult. Digitized ERTS data appears to be readily temporally overlayed based on results from a test overlay of August 9 and October 2, 1972 data (Anuta, 1970). Future analysis of ERTS data may indicate that a certain season or month yields the best results in a given metropolitan area. It may be possible, then, to update yearly land use inventories in those urban areas using overlayed ERTS data.

A final note must be added relative to the cost effectiveness of computer analysis of satellite data. The authors foresee a day, perhaps before the present decade is spent, when data tapes will arrive at computer centers and detailed, accurate maps of urbanized areas will be produced in a matter of days. Actual estimates of economic benefits cannot be made at this time, but one need only consider the man-hours which will have been saved by use of a computer. Presently, the largest U.S. metropolitan areas are growing at the rate of thousands of persons per month. Planning officials must have timely and accurate data to cope with such influxes.

IX. REFERENCES

- Anderson, James R.; Hardy, Ernest E.; and Roach, John T. A Land Use Classification System for Use with Remote Sensor Data. U.S. Geological Survey Circular 671. Washington, D.C.: U.S. Geological Survey, 1972.
- Anuta, Paul E. "Spatial Registration of Multispectral and Multi-temporal Digital Imagery Using Fast Fourier Transform Techniques". <u>IEEE Transactions on Geoscience Electronics</u>. Volume GE-8, No. 4, October, 1970, pp. 353-368.
- Ellefsen, R.; Swain, P. H.; and Wray, J. R. <u>Urban Land Use Mapping by Machine Processing of ERTS-1 Multispectral Data: A San Francisco Bay Area Example.</u>
 To be presented at the Conference on Machine Processing of Remotely Sensed Data, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana, October 16-18, 1973.
- Fu, K. S.; Landgrebe, D. A.; and Phillips, T. L. "Information Processing of Remotely Sensed Agricultural Data". <u>Proceedings IEEE</u>. Vol. 57, April, 1969, pp. 639-653.
- Todd, W. J.; Mausel, P.; and Wenner, K. A. "Preparation of Urban Land Use Inventories by Machine Processing of ERTS MSS Data". Proceedings of Symposium on Significant Results Obtained from ERTS-1 (forthcoming). Greenbelt, Maryland: NASA/Goddard Space Flight Center, 1973.
- Wacker, A. G. and Landgrebe, D. A.

 LARS Information Note 100771.

 Applications of Remote Sensing, Purdue University, 1971.

 The Minimum Distance Approach to Classification, West Lafayette, Indiana: The Laboratory for Purdue University, 1971.

Table 1. Comparison Between Classification System Proposed by Anderson, Hardy, and Roach and System Used in Present Study

Anderson, Hardy and Roach System 1	Present Study			
Level I	Level II	Spectral Class		
02. Agricultural Land 03. Rangeland 04. Forest Land 05. Water 06. Nonforested Wetland	01. Residential 02. Commercial & Services 03. Industrial 04. Extractive 05. Transportation, Communications, Utilities 06. Institutional 07. Strip and Clustered Settlement 08. Mixed 09. Open and Other	1. Old 2. New }Yes No No No No² No³ No }Yes NA⁴ Yes Yes NA NA		
07. Barren Land 08. Permanent Snow & Icefields	3	NA		

¹James R. Anderson, Ernest E. Hardy, and John T. Roach, <u>A Land Use Classification System for Use with Remote Sensor Data (U.S. Geological Survey Circular 671; Washington, D.C.: U.S. Geological Survey, 1972), p. 6.</u>

²Larger areas were classified as commercial/industrial.

³Classified as new residential.

^{*}Not applicable to Marion County.

Table 2. Accuracy of Classification

		Percentage of data points classified into:								1
Spectral Class		C/I¹	OHg ²	NHg ³	Wood	Grsy	C1d	CdSh	Watr	% with ⁵
1.	Commerce/Industry	96.0	1.1	1.3	-	1.6	-		-	97.6
2.	Older housing ²	-	81.0	8.2	1.0	9.7	-	-	-	91.7
3.	Newer housing ³	0.2	2.0	91.2	-	6.0	-	-	-	97.2
4.	Wooded	-	-	0.3	99.4	0.3	-	-	-	99.7
5.	Grassy ⁴	7.9	25.4	2.5	6.6	64.5	-	0.1	_	93.2
6.	Cloud .	-	-	14.4	-	-	85.6	_	-	
7.	Cloud shadow	11.5	0.7	-	-	-	-	86.3	1.4	
8.	Water	3.3	2.9	-	0.9	-	-	-	92.9	99.1
	Ponds	2.6	-	-	0.3	-	-	-	97.3	99.8
Streams		16.7	47.9		10.4	-		-	25.0	89.6

Overall classification accuracy = 87.1%
Accuracy minus weather conditions = 87.5%
(minus cloud and shadow)
Accuracy with areal information = 96.4%
(minus weather conditions)

Table 3. Special Investigations - Classification Accuracy

C		Percentage of data points classified into:										
	ecial Class	_ C/I¹	OHg ²	NHg ³	Wood	Gsy⁴	Cld	CdSh	Watr			
1.	Transitional residential	0.0	13.4	7.9	23.6	55.1	0.0	0.0	0.0			
2.	Vegetative residential	0.0	0.0	29.3	3.2	67.5	0.0	0.0	0.0			

¹Commerce/Industry

¹Commerce/Industry

²Multi-family (older) residential

³Single-family (newer) residential

Grassy (open, agricultural) areas

⁵Percentage with areal information (urban-rural differentiation)

²Multi-family (older) residential

³Single-family (newer) residential

Grassy (open, agricultural) areas

Table 4. Quantitative Information for Samples from Four Selected Land Uses for All Four ERTS Bands 1

Land ²	Τ		Mean	s and	Stand	ard De	viati	ons		Con	relat	ion C			
Use		4 X 3	4σ ⁴	5 $\overline{\mathrm{X}}$	5σ	$6\overline{X}$	6 o	$7\overline{X}$	7 o	r ₄₅	r ₄₆				
Tr. Res.	1 2 3	24.24	2.60	17.09 16.44 16.73	2.89	32.38 35.14	2.25	18.94 20.97 19.06	1.14	+.88	3 +.49 7 +.43 7 +.42	+.00	+.31	06 14	+.79
OHg	1. 2! 3! 4! 5 6	26.57 25.85 27.11 26.27	1.75 1.23 1.81 2.14	20.29 19.75 20.83 19.58	2.22 1.29 1.92 2.60	29.33 31.80 31.06 30.09	2.03 1.79 1.76 1.96		1.48 0.83 0.90 0.83	+.79 +.52 +.77 +.88	5 +.30 +.63 +.28 +.59 +.62 +.22	+.60 +.31 +.51 +.22	+.42 +.27 +.39 +.62	+.34 21 +.40 +.14	+.83 +.49 +.63 +.42
Veg. Hs.	1 2 3 4 5 6	25.17 27.29 25.72 27.75	1.62 1.16 1.18 1.96	17.73 20.17 18.39 21.75	2.07 1.63 1.42 1.86	41.10 42.17 41.89 43.58	2.59 2.58 2.61 1.51	23.31 24.33 25.04 25.11 24.75 25.17	1.95 1.68 1.32	+.70 +.55 +.49 +.83	+.33 +.15 +.30 +.09 +.36 +.06	01 05 +.25 +.23	19 +.09 +.06 +.38	51 16 15 +.20	+.76 +.68 +.72 +.21
Grassy	1 2 3 4 5 6	24.18 24.82 22.78 23.87	1.63 2.44 1.72 1.37	18.45 19.36 16.73 18.64	2.15 4.41 2.12 2.36	30.60 31.87 34.58 31.64 32.76 33.63	6.89 6.53 6.13 8.00	_	4.86 5.16 4.48 6.05	+.44 +.83 +.62 +.43	+.38 +.41 24 +.27 16 +.50	+.34 38 +.17 19	36 51 25 77	41 67 34 78	+.97 +.95 +.96 +.97

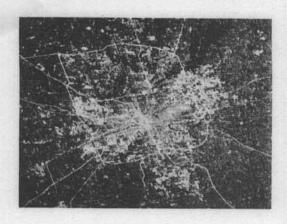
¹Band 4, 0.5-0.6 μ m; Band 5, 0.6-0.7 μ m; Band 6, 0.7-0.8 μ m; Band 7, 0.8-1.1 μ m.

²Tr. Res. = transitional residential; OHg = multi-family (older) residential; Veg. Hs. = vegetative residential; Grassy = grassy (open, agricultural) areas.

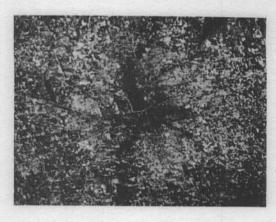
³Relative mean spectral response for Band 4.

^{*}Standard deviation for Band 4.

⁵Correlation between Band 4 and Band 5.



A



B

Figure 1. Photos of Marion County imagery from digital display. Figure 1-A is from the visible portion of the spectrum (Band 4, 0.5-0.6µm); B is from the reflective infrared (Band 6, 0.7-0.8µm). Area shown represents approximately 2524 square kilometers (966 square miles), approximately 60 per cent of which is Marion County. Horizontal length of image is 54.4 kilometers (34 miles); vertical length is 46.4 kilometers (29 miles). The true north-south line is rotated about 18 degrees counterclockwise to vertical. Horizontal scale is approximately three-quarters that of the vertical scale.



Figure 2. Photo of computer-implemented land use classification image of Marion County from digital display.