USING HCMM THERMAL DATA TO IMPROVE CLASSIFICATION OF MSS DATA

R. G. WITT
NASA/Goddard Space Flight Center

R. S. SEKHON
Computer Sciences Corporation/
Goddard Space Flight Center

Greenbelt, Maryland 20711, U.S.A.

ABSTRACT

Spectral overlap between urban and rural land use/land cover categories can lead to unacceptable map accuracy levels in the classification of Landsat multispectral scanner (MSS) data. The four MSS bands used alone are not always adequate to distinguish among various land uses and cover types having similar spectral responses. This study investigates the use of thermal data from the Heat Capacity Mapping Mission (HCMM) satellite as a means of improving MSS land cover classification accuracies for urban versus rural categories.

1. INTRODUCTION

The Heat Capacity Mapping Mission (HCMM) satellite was launched on April 26, 1978 to acquire thermal and reflectance data for potential applications in a variety of disciplines. A number of investigators have reported on the utility of the data for discriminating geologic types (Cole and Edmiston, 1980), mapping soil moisture (Reginato et al., 1976; Kocin, 1979), measuring plant canopy temperatures (Harlan et al., 1981; Wiegand et al., 1981), and studying patterns of thermal circulation in large water bodies (Schowengerdt, 1982). While the potential for using satellite-acquired thermal data to detect and study urban heat islands has been explored by Carlson et al. (1977), Matson et al. (1978), Price (1979), and Rao (1972), there has been no practical application of HCMM data to delineate urban areas using digital classification techniques. This paper documents classification procedures for using HCMM data along with Landsat MSS data in order to improve the separability of urban and non-urban areas.

2. DESCRIPTION OF THE HCMM SATELLITE AND DATA

The HCMM satellite, whose mission lasted until October 1980, carried a two-channel radiometer to sense emitted data in a thermal infrared (IR) band (10.5-12.5 μm) and reflected data in a visible band (.5-1.1 μm). The thermal channel's NEΔT is 0.3°K at 280°K, with a nominal spatial resolution of 600 meters at nadir for both bands. The three types of data obtained by HCMM were reflectance and infrared data during a daytime pass, and infrared data at night. Thermal inertia and temperature difference (day minus night) data were also calculated. These data sets are available individually or as registered day/night pairs in image format and on computer-compatible tapes (CCTs).

The HCMM satellite thermal sensor was calibrated at launch to measure a range of temperature values between 260°K and 340 K (-13°C to 67°C, or 8.6°F to 152.6°F) . With an eight-bit (0-255) configuration, the sensitivity of the thermal channel is such that HCMM was capable of measuring .3°K or less than .6°F changes in temperature. This high thermal sensitivity suggests that it should be possible to differentiate between relatively dense man-made materials and surfaces, and natural cover types, vegetation and water, on the basis of their relative emissivities.

3. RATIONALE FOR COMBINED CLASSIFICATION

This study was an extension of a land use/land cover change detection project conducted by the Eastern Regional Remote Sensing Applications Center (ERRSAC) of NASA/Goddard Space Flight Center, and the Ohio Environmental Planning Agency/Office of the Planning Coordinator (EPA/OPC). Clark County, a largely agricultural county in west-central Ohio, was the area selected for classification of two MSS data sets. The city of Springfield (population 70,000) and two other large towns are the only extensive urban developments within Clark County. When the initial classifications of MSS data, both unsupervised and supervised, failed to distinguish between commercial-industrial areas and bare agricultural lands, as well as lower density residential areas and cropland, a decision was made to integrate HCMM thermal data. The underlying assumption made was that the emitted temperatures associated with urban land uses are generally higher than those of surrounding rural land covers, and that this dichotomy would allow urban areas to be delineated on the basis of characteristic temperature differences.

The data sources used included the following:

- Landsat MSS scenes imaged on 2 June 1974 and 26 May 1977.
- HCMM digital data sets (daytime thermal IR and visible) acquired 12 March 1979.
- Ohio Department of Natural Resources (ODNR), Ohio Capability Analysis Program (OCAP) 1:24,000 scale level I land use map of Clark County.

- U. S. Geological Survey 7.5' topographic quadrangle maps of Clark County.

- Aerial photography at 1:60,000, 1:30,000, and 1:24,000 scales for portions of Clark County.

The OCAP level I (Anderson et al., 1976) land use map represented the existing land use classification encoded in the State's data base, and was used as the main source of ground truth information in carrying out the accuracy assessment.

4. PROCEDURES

Three classifications were developed in order to compare how the addition of HCMM data influenced the map accuracy of the results. A classification of the two MSS scenes covering Clark County was the first stage of the study. Unsupervised signatures were developed by using a clustering technique, and supervised signatures were derived by selecting training sites. The final signatures were input to a maximum likelihood classifier to produce classifications of the entire County. When the results did not prove satisfactory, it was found that changing or eliminating certain ambiguous signatures did not lead to significant improvement.

Two general approaches, referred to herein as the masking and merging techniques, were used to integrate the HCMM data. With the masking technique, HCMM thermal data were density sliced to create binary masks representing urban/non-urban areas of the image. The 'urban' and 'rural' masks then were multiplied by the raw MSS data set to create two complementary images. These were classified using separate sets of signatures, and the resulting images were recombined to form the final classification. In the merging technique, two HCMM bands (day infrared and visible) were combined with the four MSS bands, and a subset of the merged data set was input for unsupervised clustering. The resulting 64 clusters were labelled, and the same maximum likelihood algorithm was applied to the entire image in the final classification. Throughout this paper, the three techniques are referred to as the MSS only, MSS-HCMM masked, and the MSS-HCMM merged classifications.

4.1. MSS Only Classification

Subsections of the geocorrected Landsat data sets for the Clark County study area were sent to the Ohio EPA/OPC. The agency analyzed the digital Landsat data by using the ORSER-OCCULT software package. The OCCULT software is a user-friendly and conversational system for utilizing the ORSER software and its modifications. The ORSER-OCCULT package includes a series of analytical routines which are used to develop signatures (sets of means and standard deviations) from known, uniform areas in the subscene, and classifies the entire study area based on these signatures.

Classification results were modified by using progressively refined signatures to achieve a classification comparable to the ground truth information. The training site statistics developed by the Ohio EPA/OPC were sent to ERRSAC for classification there with the Interactive Digital Image Manipulation System (IDIMS) software on a Hewlett-Packard (HP) 3000 computer. Using these signatures, the entire subscene was classified with a maximum likelihood algorithm to duplicate the classification at ERRSAC. All spectral classes were renumbered to the six land use/land cover categories and reclassified to smooth out single-pixel discrepancies.

Visual inspection of this classified image found unacceptable rates of confusion between urban areas and bare fields that were being classified mistakenly as either commercial-industrial or residential land use. The first means of correcting this problem was to attempt a minimum distance classification. Critical limits were raised for agriculture/bare soils signatures and, when this did not result in significant improvement, limits for commercial-industrial and residential signatures were lowered. These changes had little effect on the classification. All signatures were then evaluated individually to detect and discard those which were causing the most confusion. The resulting classification still showed no major improvement, and the MSS only classified image did not undergo any further refinement.

4.2. MSS-HCMM Masked Classification

Data from the HCMM satellite, scene ID AA0320-18200 (image center at N39 55' and W82 15'; sun elevation of 45), were received from the National Space Science Data Center at Goddard Space Flight Center. Both data sets acquired, day thermal infrared (day IR; 10.5-12.5 m) and day visible (day VIS; 0.5-1.1 m), were taken simultaneously at about 1:30 in the afternoon, the time of maximum surface temperature. The image quality was good, and though there was ten per cent cloud cover, no clouds or haze obscured the study area. The digital images came registered to one another, but not to any map base. Thus the images had to be registered to available maps of the study area, and resampled using a nearest neighbor procedure to overlay the MSS data at the same scale and pixel size (50 meters). Appropriate subsets corresponding to the MSS image of the study area were created. After doing a contrast stretch of both the infrared and visible bands to the fullest range possible, it was determined that the thermal band offered better definition of urban areas when compared with maps of Clark County. Therefore, the thermal image was selected to be used in developing the binary urban and rural masks.

The binary image was created by mapping the thermal IR data set to desired levels. A stochastic method was employed to determine the threshold which represented the boundary between the higher emissivity of warmer, urban areas, and the lower emissivity of cooler, non-urban areas. All digital values above this level were mapped (that is, renumbered) to one, and all digital values at this level and below were mapped to zero. After repeated experimentation with the mapping, the values 46 and below were mapped to zero and the values 47 and above were mapped to one, creating an urban area mask. The rural area mask was complementary to the latter image; it was obtained by mapping 0 1 values to 1 0 values. These masks were used to create complementary MSS images representing urban and non-urban areas. Each of the four MSS bands was multiplied by the urban mask and then united to form a masked MSS image representing only urban areas. Similarly utilizing the MSS image and the non-urban mask, a masked MSS image representing rural areas was formed.

Separate groups of signatures were used to classify the 'urban' and 'rural' data sets. The urban group of signatures included commercial-industrial, residential, agriculture, forest and water signatures (all except bare soil), while the rural group included agriculture, bare soil, forest and water. No pixels in the rural image could be classified as commercial-industrial or residential, thus eliminating some small towns and low-density suburban areas. However, urban areas did encompass some farmland and parkland, which were classified correctly as agriculture in the broad sense.

After the 'urban' and 'rural' images were classified separately using a minimum distance algorithm with modified threshold limits for all signatures, three additional steps were necessary to complete the classification. Each of the classified images was renumbered to reflect the final number of land use/land cover categories, the complementary images were added together, and the united classification was reclassified to smooth out single-pixel discrepancies.

4.3. MSS-HCMM Merged Classification

The third classification was generated from a Landsat MSS-HCMM merged data set. The registered and resampled HCMM subscene was united with the MSS subscene to form a six-band image, containing four MSS bands and two HCMM bands (day IR and visible). Unsupervised signatures were developed from a representative test area using a clustering technique. These signatures then were applied to the entire study area using a maximum likelihood classifier. The spectral classes were labelled as the same six land use/land cover categories in the other two classifications. The classified image was renumbered and reclassified as previously described, thus completing the processing phase of the study.

4.4. Accuracy Assessment Procedures

The accuracy assessment was conducted by comparing land cover maps from the three classifications with the 1:24,000 scale OCAP land use map of level I categories. Three U.S.G.S. quadrangles in central Clark County - Springfield, New Moorefield, and Donnelsville - were selected for this procedure. Thematic grayscale overlays of each classification were output on a Versatec plotter at the same scale to facilitate the comparison.

Each of the thematic overlays (one per quad for three classifications, or a total of nine) was registered to the OCAP land use map on a light table. A grid with a cell size of 25 pixels (5x5 pixel blocks) was superimposed on the maps, and numbered along its x and y axes. Using a random numbers table, 50 cells were sampled on each overlay and compared with the ground truth map on a pixel-by-pixel basis. A count of the number of correct and incorrect pixels was kept; these results are presented in Tables I and II.

Because the Landsat and OCAP categories were not identical, some categories were combined for the accuracy assessment. The Landsat commercial-industrial and residential categories were combined so that they could be compared with urban/built-up and barren land categories in OCAP. The agriculture and bare soil categories (Landsat) also were added together and compared with OCAP's combined agriculture and rangeland. Forest and wetlands were combined (OCAP) and compared with forest in Landsat. The remaining level I category was water; thus it was possible to compare four general land cover types.

5. RESULTS

The results of the three classification procedures are presented below in two tables. Table I compares the classification acreage counts for the four level I categories with the ground truth acreages from OCAP, for the total area of the three 7.5' quads on which the accuracy assessment was conducted. Table II shows the percentage of agreement derived from the accuracy assessment, again for all categories and the three quads totalled.

From the first table it can be seen that the MSS-HCMM merged classification comes closest to estimating the actual extent of urban and agricultural lands, according to the OCAP information. Only the forest category is overestimated, and this may be a result of cluster mislabelling. It became evident during the accuracy assessment that many pixels classified as forest on the HCMM merged classification should have been agriculture or urban instead. If this problem were to be corrected, the agreement between the OCAP and HCMM merged acreages would be even greater.

Table I: ACREAGE COUNTS BY CLASSIFICATION

Level I Category	MSS only (1977)	HCMM masked (1977-79)	HCMM merged (1977-79)	OCAP (1979)
Urban	30706	11958	17985	19834
	28%	10.9%	16.4%	18.1%
Agric.	67101	89043	75278	78524
	61.1%	81.1%	68.6%	71.5%
Forest	7430	6804	14545	8482
	6.8%	6.2%	13.2%	7.7%
Water	2026	1997	1993	2942
	1.8%	1.8%	1.8%	2.7%
Unclas.	2537	7	-	-
	2.3%	-	-	-
Totals	109800	109802	109801	109782

Conversely, in terms of locational accuracy, the HCMM merged classification delineated forest better than the other techniques (56.3 per cent, as opposed to 52.9 per cent and 37.7 per cent for the MSS only and HCMM masked classifications). This reflects a low rate of omission errors and a high rate of commission errors for the forest category, which led to decreased locational accuracies for the urban and agricultural categories in the MSS-HCMM merged classification. In general, however, the MSS-HCMM merged classification had the virtues of the other two classifications without the faults of either. It provided more highly correlated estimates of urban and agricultural lands, and locational accuracies as good as or better than the MSS only classification.

The other fact that emerges from the comparison of acreage counts is that the HCMM masked classification underestimates urban land and overestimates agricultural land by approximately the same amounts that the MSS only classification does the reverse. For urban land, the HCMM masked estimate is 7.2 per cent low and the MSS only estimate is 9.9 per cent high; while for agriculture the MSS only estimate is 10.4 per cent low and the HCMM masked estimate is 9.6 per cent high. This is the result of a fairly indiscriminant urban classification with the MSS data (high rates of bare soil being classified as commercial-industrial, and cropped fields being mapped as residential areas), and a very restrictive urban classification on the HCMM masked image. There were very few errors of commission in the urban category for the HCMM masked classification as a result (see Witt and Sekhon, 1982). The fact that only areas within the urban mask could be classified as commercial-industrial or residential led to the relatively high locational accuracies for urban (65.7 per cent) and agriculture (87.9 per cent) on the HCMM masked classification.

Table II: ACCURACY ASSESSMENT RESULTS

Level I	Classification Percentage Correct*		
Category	MSS only	HCMM merged	HCMM masked
Urban	61.8	60.0	65.7
Agric.	82.0	84.0	87.9
Forest	52.9	56.3	37.7
Water	81.4	89.1	81.6
Totals	73.8	76.1	79.6

(*calculated as 100% minus the average of commission and omission errors)

Thus, the MSS-HCMM masked classification had the highest overall per-pixel accuracy for the three 7.5' quadrangles checked, while the MSS-HCMM merged classification better represented overall acreage totals for the two principal categories assessed. Furthermore, the MSS-HCMM merged classification was clearly superior in terms of locational accuracy for the forest and water categories, while the MSS only classification included more than 2 per cent unclassified data.

6. CONCLUSIONS

The addition of HCMM thermal day infrared (and visible) data to the MSS classification did lead to better results in map accuracy for level I land use/land cover categories. Although neither the masking nor the merging procedure led to dramatic increases in classification accuracy, both techniques show potential for improved delineation of urban land from surrounding non-urbanized areas.

The masking technique was more effective in delineating the city of Springfield and larger towns, while excluding small towns and linear developments which were below the spatial resolution of the HCMM sensor. Thus, small communities which were not "bright" or warm enough to saturate a single/multiple HCMM pixels could not be classified as urban using the binary masking (urban/rural stratification) technique.

On the other hand, the merging technique relied on the thermal information that was added directly to the classification of every MSS pixel. Obviously, large blocks within the MSS data set had the same thermal value due to the coarser resolution of HCMM, but variation of the four MSS bands was enough to eliminate any appearance of blockiness in the final classified image. The clustering method by which the merged classification was executed allowed smaller towns and major transportation corridors to be properly labelled as urban even if they were below the HCMM spatial resolution. There was still some confusion, however, between urban and rural land cover categories associated with the cluster labelling process, which resulted in locational accuracies lower than those anticipated by the researchers.

It may be possible to further improve classification results by using a hybrid procedure incorporating both techniques. In this procedure, a binary (urban/non-urban) image would be created first from the HCMM thermal data. After multiplying the binary image with the merged MSS-HCMM data set, the resulting 'urban' and 'rural' six-banded images would be subjected to separate clustering and cluster labelling. The classification process then would reflect a bias toward urban land uses within the 'urban' image, and non-urban land covers within the 'rural' image.

6.2. Additional Research

Future research relating to the integration of HCMM and MSS data for improved classification results should focus on several key topics. More work needs to be done to determine optimal conditions under which to employ HCMM data for urban area delineation. This probably is dependent not only on the density and extent of the particular urban area, but on such factors as the time of year of the HCMM and MSS images, atmospheric conditions on a given day, and the types of vegetation present in the surrounding area. Some combination of these factors may determine whether it is appropriate to utilize HCMM data for delineating various types of urban areas, or whether (for example) the analysis would profit more from the digitization of urban area boundaries.

Research is now being carried out to test the radiometric stability of HCMM data for the same scene from date to date. Because of the problems with the absolute calibration of the HCMM thermal sensor, the non-experimental usage of HCMM thermal data has been subject to question. If it can be proven that there is little variation in sensor performance over time, it is expected that the use of HCMM thermal data for the type of application discussed above might become more widely accepted.

REFERENCES

1. Anderson, J. R., Hardy, E. E., Roach, J. T., and R. E. Witmer. 1976. A land cover classification system for use with remote sensing data. U. S. Geological Survey Professional Paper 964, Washington, D.C., 33 p.

2. Augustine, J. A. 1978. A detailed analysis of urban ground temperature and albedo using high-resolution satellite measurements. M.S. thesis, Department of Meteorology, Pennsylvania State University (DM/PSU).

3. Boland, F. E. 1977. A model for determining surface temperatures and sensible heat fluxes over the urban-rural complex. M.S. thesis, DM/PSU.

4. Carlson, Toby N. 1980. Applications of HCMM satellite data to the study of urban heating patterns: remote estimate of the surface energy flux, moisture availability and thermal inertia over urban and rural terrain. NAS5-24264 Final Report, DM/PSU, College Park, Pa., 62 p.

5. ______, Boland, F. E., and J. A. Augustine. 1977. Potential application of satellite temperature measurements in the analysis of urban land use over urban areas. Bulletin of the American Meteorological Society, 58:1301-1303.

6. Cole, M. M., and D. J. Edmiston. 1980. HCMM and Landsat imagery for geologic mapping in Northwest Queensland, in Fourteenth International Symposium on Remote Sensing of Environment, San Jose, Costa Rica, April 23-30, 1980, Environmental Research Institute of Michigan, Ann Arbor, Mich., Proceedings: Vol. 3, pp. 1849-1857.

7. DiCristofaro, D. C. 1980. Remote estimation of the surface characteristics and energy budget over an urban-rural area and the effects of surface heat flux on plume spread and concentration. M.S. thesis, DM/PSU, 106 p.

8. Dodd, J. K. 1979. Determination of surface characteristics and energy budget over an urban-rural area using satellite data and a boundary layer model. M.S. thesis, DM/PSU, 87 p.

9. Harlan, J. C., Rosenthal, W. D., and B. J. Blanchard. 1981. Dryland pasture and crop conditions as seen by HCMM. NAS5-24383 Final Report 3712, Texas A&M University Remote Sensing Center, College Station, Texas, 53 p.

10. Kocin, P. J. 1979. Remote estimation of surface moisture over a watershed. M. S. thesis, DM/PSU, 62 p.

11. Matson, Michael, McClain, E. P., McGinniss, D. F., and J. A. Pritchard. 1978. Satellite detection of urban heat islands. Monthly Weather Review 106:1725-1734.

12. NASA/Goddard Space Flight Center. 1978 (revised 1979, 1980). Heat Capacity Mapping Mission Data Users' Handbook for Applications Explorer Mission-A (AEM). 120 p.

13. Price, J. C. 1979. Assessment of the urban heat island effect through the use of satellite data. Monthly Weather Review 107:1554-1557.

14. Rao, P. K. 1972. Remote sensing of urban heat islands from an environmental satellite. Bulletin of the American Meteorological Society 53:647-648.

15. Reginato, R. J., Idso, S. B., Vedder, J. F., Jackson, R. D., Blanchard, M. B., and R. Goettelman. 1976. Soil water content and evaporation determined by thermal parameters obtained from ground-based and remote measurements. Journal of Geophysical Research 81:1617-1620.

16. Schowengerdt, Robert. 1982. Enhanced thermal mapping with Landsat and HCMM digital data, in American Congress on Surveying and Mapping-American Society of Photogrammetry 48th Annual Meeting, Denver, Colorado, March 14-20, 1982, Proceedings: American Society of Photogrammetry, Falls Church, Va., pp. 414-422.

17. Short, N. M. 1981. The Heat Capacity Mapping Mission, in Second Eastern Regional Remote Sensing Applications Conference, Danvers, Mass., March 9-11, 1981, Proceedings: NASA Conference Publication 2198, Goddard Space Flight Center, Greenbelt, Md., pp. 1-5.

18. ______, and L. M. Stewart, Jr. 1982. The HCMM Anthology. NASA publication in press.

19. Wiegand, C. L. 1981. Plant cover, soil, temperature, freeze, water stress, and evapotranspiration conditions. AgRISTARS EW/CCA Final Report EW-UI-04103, JSC-17143, 159 p.

20. Witt, R. G., Sekhon, R. S., and Anthony Sasson. 1982. An urban-rural land use/land cover inventory of Clark County, Ohio, using Landsat digital data. Eastern Regional Remote Sensing Applications Center (ERRSAC) open file project report, NASA/Goddard Space Flight Center, Greenbelt, Md., 30 p.