A Hierarchical Fuzzy Classification Approach for High-Resolution Multispectral Data Over Urban Areas

Aaron K. Shackelford, *Student Member, IEEE,* and Curt H. Davis, *Senior Member, IEEE*

Abstract—In this paper, we investigate the usefulness of high-resolution multispectral satellite imagery for classification of urban and suburban areas and present a fuzzy logic methodology to improve classification accuracy. Panchromatic and multispectral IKONOS image datasets are analyzed for two urban locations in this study. Both multispectral and pan-sharpened multispectral images are first classified using a traditional maximum-likelihood approach. Maximum-likelihood classification accuracies between 79% to 87% were achieved with significant misclassification error between the spectrally similar *Road* and *Building* urban land cover types. A number of different texture measures were investigated, and a length–width contextual measure is developed. These spatial measures were used to increase the discrimination between spectrally similar classes, thereby yielding higher accuracy urban land cover maps. Finally, a hierarchical fuzzy classification approach that makes use of both spectral and spatial information is presented. This technique is shown to increase the discrimination between spectrally similar urban land cover classes and results in classification accuracies that are 8% to 11% larger than those from the traditional maximum-likelihood approach.

Index Terms—Fuzzy classification, high-resolution satellite imagery, urban remote sensing.

I. INTRODUCTION

URBAN and economic growth places a heavy demand on local governments to seek better planning and management approaches to deal with the numerous problems associated with increasing urbanization. Timely and accurate information products are required by federal, state, and local government agencies and officials to make effective decisions regarding a wide variety of issues affecting the urban environment. High-resolution commercial satellite imagery has been shown to be a cost-effective alternative to aerial photography for the generation of digital image basemaps [1], which are digital images with map-quality positional accuracies. Information products derived from positionally accurate high-resolution satellite imagery, such as land cover maps, can be easily integrated into existing state and local government GIS databases and utilized to aid officials in planning and decision making processes [2]. Applications for urban land cover maps include environmental planning and assessment, land use change detection/attribution, utility and transportation planning, infrastructure inventory, stormwater planning/mitigation, and water quality management.

Analysis of urban areas using medium-resolution remote sensing imagery (e.g., Landsat) has typically focused on the identification of built-up areas or discrimination between residential, industrial, and commercial zones. However, with the recent availability of commercial high-resolution satellite multispectral imagery from sensors such as IKONOS and QuickBird, it is now possible to produce more detailed urban land cover maps by identifying features such as individual roads and buildings in the urban environment. High-resolution data over urban areas have been classified using morphological profiles [3] and neural network techniques [4]. In addition, various methods for road extraction from high-resolution satellite imagery and aerial photography have been investigated [5]–[7]. Studies have been conducted on the use of texture and contextual information in the classification of high-resolution satellite imagery of urban areas [8], [9]. In addition to pixel-based approaches, high-resolution urban imagery can be analyzed using segmentation and object-based classification approaches [10], [11]. In [12], a supervised fuzzy classification method for Landsat Thematic Mapper (TM) data is presented.

Because of the complex nature and diverse composition of land cover types found within the urban environment, the production of urban land cover maps from high-resolution satellite imagery is a difficult task. The materials found in the urban environment include concrete, asphalt, metal, plastic, glass, shingles, water, grass, trees, shrubs, and soil, to list just a few. Moreover, many of these materials are spectrally similar, and this leads to problems in automated or semiautomated image classification of these areas. In addition, these materials form very complex arrangements in the imagery such as housing developments, transportation networks, industrial facilities, and commercial/recreational areas. Conventional methods for classification [13] of multispectral remote sensing imagery such as parallelepiped, minimum distance from means, and maximum likelihood, only utilize spectral information and consequently have limited success in classifying high-resolution urban multispectral images. As many classes of interest in the urban environment have similar spectral signatures, spatial information such as texture and context must be exploited to produce accurate classification maps.

Another disadvantage of conventional classification methods is that they only produce crisp classifications, i.e., each pixel

The authors are with the Department of Electrical and Computer Engineering, University of Missouri–Columbia, Columbia, MO 65211 USA (e-mail: DavisCH@missouri.edu).

can only be classified as one class. However, remote sensing images contain mixed pixels and many land cover types have similar spectral signatures. These problems are particularly severe in urban environments. Fuzzy classification techniques allow pixels to have membership in more than one class and therefore better represent the imprecise nature of the data. In this paper, a hierarchical fuzzy classification method that incorporates both spectral and spatial information is presented. This technique produces a substantial increase in classification accuracy of urban land cover maps compared to the traditional maximum-likelihood classification approach.

The remainder of this paper is organized as follows. The accuracy and limitations of maximum-likelihood classification of high-resolution satellite imagery over urban and suburban areas are presented in Section II. In addition to spectral data, several types of spatial information can be extracted from the high-resolution imagery. These are investigated and corresponding results are presented in Section III. In Section IV, we describe a hierarchical fuzzy classifier that utilizes both spectral and spatial information to produce more accurate urban land cover maps. Finally, the conclusions are presented in Section V.

II. CLASSIFICATION OF HIGH-RESOLUTION SATELLITE IMAGERY

We first investigated the effectiveness of high-resolution satellite imagery for classification of urban and suburban scenes using a traditional maximum-likelihood classifier. The imagery used for this study was acquired by the IKONOS commercial remote sensing satellite and consists of four multispectral (MS) bands with 4-m resolution and a single panchromatic (PAN) band with 1-m resolution. The four MS bands collect data at the red, green, blue, and near-infrared wavelengths, and the data in each band is stored with 11-b quantization. Two IKONOS image datasets are used in this study: an image of Columbia, MO acquired on April 30, 2000, and an image of Springfield, MO acquired on September 17, 2000. Both image datasets include a variety of urban and suburban land cover types making them ideal for this study. Two separate datasets were used to provide multiple evaluations of the algorithms presented in this paper and to ensure that the algorithms were not so highly specialized as to be applicable to only a single dataset.

The Columbia image is shown in Fig. 1. The IKONOS images went through several preprocessing steps before classification. First, the images were orthorectified to increase the planar accuracy from 25 m RMS to approximately 3 m RMS. Map-quality positional accuracy is needed so that the image data and derivative products (e.g., land cover map) can be effectively incorporated into GIS databases [1]. After orthorectification, a color normalization method [14] was used to fuse the PAN data with the four MS bands to produce a four-band pan-sharpened multispectral (PS-MS) image with 1-m resolution. The PS-MS imagery retained the 11-b quantization of the original data.

Both the 4-m MS and 1-m PS-MS image datasets were classified using the traditional supervised maximum-likelihood approach. A more detailed classification of the urban landscape is possible from the high-resolution IKONOS imagery

Fig. 1. One-meter resolution panchromatic IKONOS image of Columbia, MO.

TABLE I
MAXIMUM LIKELIHOOD CLASSIFICATION RESULTS FOR 4-m MS AND 1-m PS-MS IMAGE DATASETS

Pixel Size	Image	Overall Accuracy (%)	Kappa Coefficient
4-m	Columbia	78.6	0.731
1-m	Columbia	80.9	0.763
4-m	Springfield	85.7	0.818
1-m	Springfield	87.4	0.840

compared to medium-resolution multispectral image data (e.g., Landsat). Accordingly, the identification of fine-scale urban features (residential houses, individual trees, etc.) in the image can be achieved. The urban land cover classes used in this study were *Road, Building, Grass, Tree, Bare Soil, Water*, and *Shadow*. The *Shadow* class is required to minimize the problem of shaded pixels in the urban environment, e.g., building shadows, being classified as *Water*. An accuracy assessment of the resulting classification was performed making use of reference pixels that were independent of the pixels used to train the classifier. The reference pixel datasets were generated via photo interpretation of the 1-m PS-MS IKONOS imagery. Approximately 175 randomly distributed test site polygons were manually digitized in the imagery. The Columbia dataset had 9410 training pixels and 80 895 reference pixels, and the Springfield dataset had 13 602 training pixels and 184 056 reference pixels. The same training and reference pixel sets were used for all classification results presented in this paper.

Supervised maximum-likelihood classifications were produced for both the 4-m MS and the 1-m PS-MS images

TABLE II
CONFUSION MATRIX FOR MAXIMUM LIKELIHOOD CLASSIFICATION OF 1-m PS-MS COLUMBIA IMAGE DATASET

	Road Ref.	Building Ref.	Grass Ref.	Tree Ref.	Bare Soil Ref.	Water Ref.	Total	%
Road	8109	2682	1	0	316	23	11,131	73
Building	3020	10,850	53	5	55	299	14,282	76
Grass	107	23	22,567	1225	0	0	23,922	94
Tree	54	57	4179	10,055	0	0	14,345	70
Bare Soil	177	49	10	0	8864	0	9100	97
Water	5	537	0	0	0	4961	5503	90
Shadow	15	712	0	0	0	1885	2612	
Total	11,487	14,910	26,810	11,285	9235	7168	80,895	
%	71	73	84	89	96	69		

from both study locations. The confusion matrix, the overall accuracy, and the Kappa coefficient of agreement [15]–[17] were computed for each classification. The overall accuracy was computed by dividing the number of correctly classified reference pixels by the total number of reference pixels. The Kappa coefficient adjusts the overall accuracy value by subtracting the estimated contribution of chance agreement between classified pixels and reference pixels [18]. The overall accuracies and Kappa coefficients are presented in Table I. The overall accuracies for the Springfield image were higher than those corresponding to the Columbia image for both the 4-m MS and the 1-m PS-MS datasets. This is most likely due to the presence of a small amount of haze in the Columbia image. The classification accuracies and Kappa coefficients of the 1-m PS-MS data are several percent higher than those of the 4-m MS data for both datasets, indicating that the pan-sharpened images produced by the color normalization method can be effectively used for classification purposes.

The confusion matrix for the PS-MS classification of the Columbia image is shown in Table II. The largest source of error is due to misclassifications between the *Road* and *Building* classes, with 26% of the *Road* reference pixels classified as *Building* and 18% of the *Building* reference pixels classified as *Road*. The other major source of error is confusion between the *Grass* and *Tree* classes, with 16% of the *Grass* reference pixels classified as *Tree* and 11% of the *Tree* reference pixels classified as *Grass*. In addition, 26% of the *Water* reference pixels are classified as *Shadow*. Suburban and urban image subsets of the maximum-likelihood classification for Columbia are shown in Fig. 2. The confusion matrix for the PS-MS classification of the Springfield image shows similar misclassification characteristics. The confusion matrix for the Springfield PS-MS classification is shown in Table III. As with the Columbia PS-MS classification, the largest source of error in the Springfield classification is caused by misclassifications between the *Road* and *Building* classes, with 30% of the *Road* reference pixels classified as *Building* and 31% of the *Building* reference pixels classified as *Road*. Unlike the classification of Columbia image, there is virtually no confusion between the *Grass* and *Tree* classes in the Springfield image. There is more spectral variation between these classes in the Springfield data because the image was acquired in the early fall time period, resulting in less confusion between the classes. In addition, 24% of the *Water* reference pixels are classified as *Shadow*.

The *Road* and *Building* classes in both images and the *Grass* and *Tree* classes in the Columbia image are spectrally similar and have a significant amount of spectral overlap. This is the primary reason for the large number of misclassifications between these classes. Traditional supervised classification methods that only take into account spectral information, such as maximum likelihood, are unable to differentiate between these classes with a high degree of accuracy. Methods that utilize spatial information in addition to spectral information are needed to produce more accurate classifications of high-resolution image data over urban areas.

III. SPATIAL INFORMATION EXTRACTION

Spatial features such as texture contain information about the spatial distribution of tonal variations within a band and are typically derived from windows of data surrounding the area being analyzed [19]. By combining spatial information and spectral information, the amount of overlap between classes can be decreased, thereby yielding higher classification accuracies and more accurate urban land cover maps. For example, while the *Grass* and *Tree* classes can have similar spectral signatures, areas in the image covered with grass appear much more homogeneous than tree-covered areas. This difference in homogeneity between regions can be used to decrease the confusion between the classes. This is illustrated in Fig. 3, where an entropy texture measure is used to differentiate between the *Grass* and *Tree* land cover types.

A variety of texture measures utilizing different window sizes were evaluated to test the usefulness of different texture measures. Each texture image was then added to the four PS-MS bands as an extra channel of data and then classified using maximum-likelihood classification. The following occurrence texture measures were evaluated: entropy, data range, skewness, and variance [20]. The texture features were calculated from the normalized gray-level histogram, $p(z_i)$, of the pixel window $w(z_i)$, where $0 \leq i \leq L-1$, and L is the number of gray levels in the image. The texture measures were calculated as follows:

$$\text{entropy} = -\sum_{i=0}^{L-1} p(z_i) \log_2[p(z_i)] \tag{1}$$

$$\text{data range} = \max[w(z_i)] - \min[w(z_i)] \tag{2}$$

$$\text{variance} = \sigma^2 = \sum_{i=0}^{L-1}(z_i - m)^2 p(z_i) \tag{3}$$
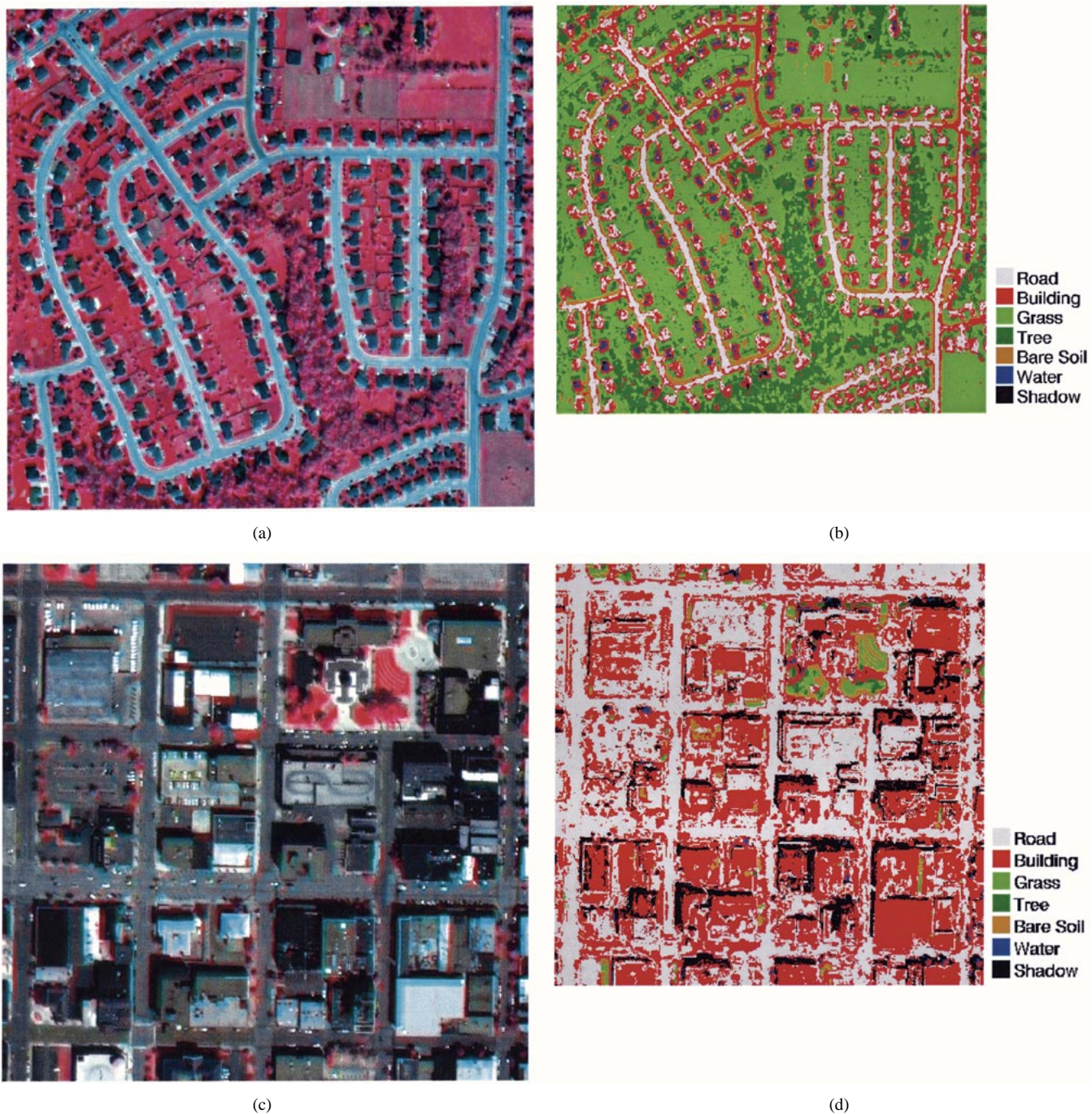
Fig. 2. Maximum-likelihood classification for (b) suburban area, (d) urban area from the Columbia, MO image subsets shown in (a) and (c), respectively. Note the significant misclassifications between the *Road* and *Building* land cover types.

$$\text{skewness} = \frac{1}{\sigma^3} \sum_{i=0}^{L-1} (z_i - m)^3 p(z_i) \tag{4}$$

where m is the mean value of the gray levels in the window, i.e.,

$$m = \sum_{i=0}^{L-1} z_i p(z_i). \tag{5}$$

Each texture measure was calculated with a 5×5, 10×10, and 20×20 pixel window. The window sizes tested were chosen to be no larger than the objects of interest in the image from which the texture measures were to extract information from. For that

reason, a 20-m-wide window was the largest texture kernel size tested. While there were areas in the image, such as fields and large tree covered areas, that were much larger than this, the texture measures needed to be applicable to urban and suburban areas where the objects of interest are on the order of 10–20 m in size. All of the texture measures discussed here were extracted from the panchromatic band of the IKONOS image datasets.

The average classification accuracy for the *Road* and *Building* classes and the *Grass* and *Tree* classes from the Columbia image is shown in Table IV. The first row in the table is the average classification accuracies from the maximum-likelihood classification of the PS-MS data with no added texture measures. The

TABLE III
CONFUSION MATRIX FOR MAXIMUM LIKELIHOOD CLASSIFICATION OF 1-m PS-MS SPRINGFIELD IMAGE DATASET

	Road Ref.	Building Ref.	Grass Ref.	Tree Ref.	Bare Soil Ref.	Water Ref.	Total	%
Road	21,219	10,667	57	20	33	0	31,996	66
Building	8968	24,165	82	4	1390	0	34,609	70
Grass	54	16	48,753	47	1	0	48,871	100
Tree	15	2	249	49,448	0	206	49,920	99
Bare Soil	0	5	0	0	13,266	0	13,271	100
Water	5	0	0	0	0	4009	4014	100
Shadow	9	0	0	0	0	1366	1375	
Total	30,270	34,855	49,141	49,519	14,690	5581	184,056	
%	70	69	99	100	90	72		

entropy texture measures using both a 10×10 and a 20×20 pixel window have a significant effect on the average classification accuracy of the *Grass* and *Tree* classes, where the classification accuracy of those classes increases approximately 10% in both cases. Although the classification accuracies of both the 10×10 and 20×20 entropy texture measures were essentially the same, the 10×10 window was chosen to help reduce edge effects associated with large texture windows [21]. Several of the other texture measures show moderate increase in the accuracy of these classes, but not as large as the increase found when using the entropy texture measure. Most of the texture measures actually <u>decrease</u> the average classification accuracies for the *Road* and *Building* classes, and the best result (entropy 20×20) only yields a 1.5% increase over the PS-MS classification with no texture features.

It was found in the previous section that the largest source of confusion in the classification of the high-resolution urban scenes is between the *Road* and *Building* classes. Thus, a spatial measure that can increase discrimination between these two classes is highly desirable. One such spatial measure is to examine the context of each pixel, measuring the spatial dimensions of groups of spectrally similar connected pixels. Roads tend to consist of groups of spectrally similar pixels oriented along a long narrow line. Buildings, on the other hand, usually consist of a group of pixels with a similar spectral response oriented in a more rectangular or square shape.

A simple algorithm was developed to extract the length and width of spectrally similar connected groups of pixels from the PS-MS imagery. The algorithm calculates a length and width value for each pixel of interest in the image. These values are found by searching along a predetermined number of equally spaced lines radiating from the central pixel. The Euclidean distance

$$D_{\text{ED}} = \left(\sum_{i=1}^{dim} (x_i^c - x_i^n)^2 \right)^{\frac{1}{2}} \quad (6)$$

is calculated between the spectrum of the central pixel and the spectrum of each new pixel, where dim is the dimensionality of the data; x_i^c is the value of the ith band of the central pixel; and x_i^n is the value of the ith band of the pixel in question. If that value is less than a similarity threshold, the search continues until the maximum allowed length is reached. Once all of the directions have been searched, the maximum value is stored as the

length and the minimum value is stored as the width. The output of the algorithm is a two-band length–width feature image.

Three parameters control the length–width extraction algorithm: the number of search directions, N_{dir}, the maximum length, L_{max}, and the similarity threshold, TH_{ED}. The similarity threshold, TH_{ED}, has the largest effect on the performance of the algorithm. The algorithm extracts accurate length and width values if TH_{ED} is set to between 2.5 to 4.0 times the average standard deviation of the Euclidean distance of the training pixel data from the class means. The length–width extraction algorithm is summarized by the psuedocode shown in Fig. 4. We found that if the data were median filtered before the length–width algorithm was applied, then the length and width measurements were more accurate representations of the data. The median filter was chosen because of its inherent properties of reducing tonal variations while retaining edges [22]. A 7×7 window for the median filter was found to work well. The kernel size for the median filter was chosen to be smaller than the desired objects being analyzed for contextual information (i.e., roads and buildings). However, the 7×7 window was large enough so that extremely fine-scale features in the image, such as automobiles and linework on the roads, were removed. Note that the effect of the median filtering is not the same as simply working with lower resolution imagery, as the edges between objects of interest are still preserved at the 1-m resolution.

The outputs of the length–width extraction algorithm applied to both an urban and a suburban scene are shown in Fig. 5. The length values are displayed in the red channel of an RGB display and width is displayed in the blue and the green channels. Vegetation pixels have been masked out so the effect of the length–width measure on road and building pixels can be more clearly seen. Pixels that have large length values and small width values, such as road pixels, appear more red in color, while pixels with similar length and width values, such as building pixels, appear more blue in color. The parameters used for the length–width extraction were: $N_{\text{dir}} = 36$ ($10°$ azimuth sampling), $L_{\text{max}} = 200$ pixels, and $TH_{\text{ED}} = 50$. This algorithm was applied to the Columbia image and the resulting two bands of data were added to the four PS-MS bands and classified using maximum-likelihood classification. The average classification accuracy for the *Road* and *Building* classes increased by 5% when the length–width features were added. However, the average classification accuracy for the *Grass* and *Tree* classes de-
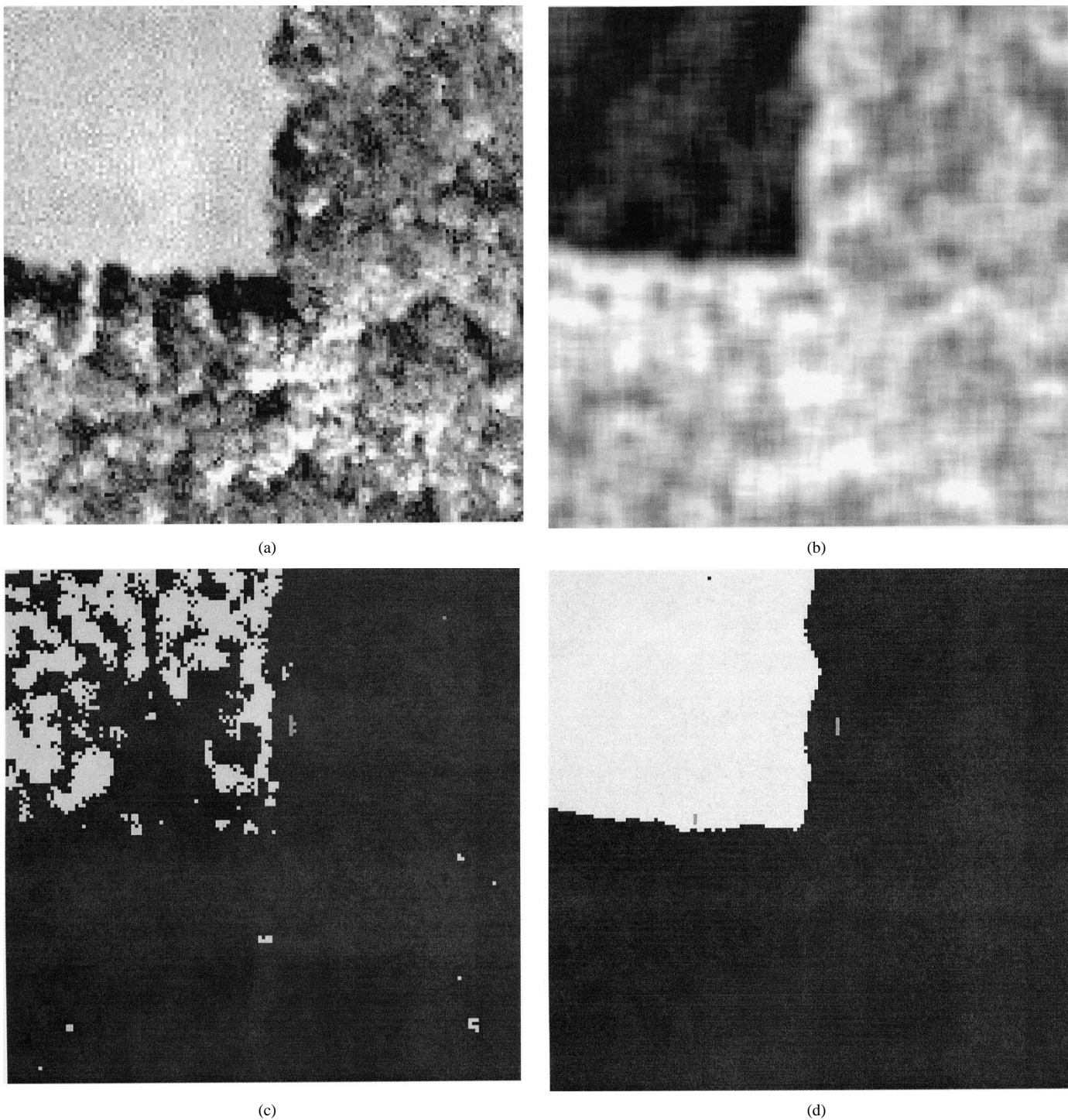
Fig. 3. Effect of entropy texture measure on classification of *Grass* and *Tree* classes. (a) Image subset. (b) 10×10 entropy texture measure. (c) Maximum-likelihood classification of (a) (light gray $=$ *Grass*, dark gray $=$ *Tree*). (d) Maximum-likelihood classification of PS-MS data $+$ entropy.

creased by 9%. Finally, after inspection of the distributions of the length–width measures, it was found that they were not normally distributed and the maximum-likelihood classification is therefore not the best choice for classification using this type of spatial feature.

IV. HIERARCHICAL FUZZY CLASSIFICATION APPROACH

Spatial measures extracted from the high-resolution multispectral imagery can help decrease the number of misclas-

sifications between the spectrally similar *Road/Building* and *Tree/Grass* classes. However, while one spatial feature might increase the classification accuracy between one set of classes, it might decrease the accuracy between another set using traditional classification methods. For example, the length–width contextual measure discussed in the previous section increased the maximum-likelihood classification accuracy between *Road* and *Building* by 5%, but the classification accuracy between *Grass* and *Tree* decreased by 9%. The entropy texture measure increased the *Grass* and *Tree* maximum-likelihood classifica-

TABLE IV
AVERAGE MAXIMUM LIKELIHOOD CLASSIFICATION ACCURACIES WITH TEXTURE INFORMATION INCLUDED FOR 1-m PS-MS COLUMBIA IMAGE DATASET

Texture Feature	Window Size	Road – Building Avg. Accuracy (%)	Grass – Tree Avg. Accuracy (%)
None	NA	71.4	86.6
Entropy	5 x 5	70.4	87.8
Data Range	5 x 5	68.2	90.1
Skewness	5 x 5	70.7	90.7
Variance	5 x 5	65.7	91.4
Entropy	10 x 10	71.8	97.5
Data Range	10 x 10	69.1	89.9
Skewness	10 x 10	71.1	92.6
Variance	10 x 10	68.0	91.5
Entropy	20 x 20	72.9	97.3
Data Range	20 x 20	66.0	95.8
Skewness	20 x 20	69.9	84.3
Variance	20 x 20	70.2	87.8
Length-Width	NA	76.2	77.9

```
for each pixel i
    for direction j = 1 to N_dir
        counter[j] = 0
        k = 1
        data_matrix = spectrum of each pixel along direction j
        repeat
            D_ED = Euclidean distance between pixel i and data_matrix[k]
            k = k + 1
            counter[j] = counter[j] + 1
        until (counter[j] > L_max) or (D_ED > TH_ED)
    endfor
    length[pixel i] = max{counter}
    width[pixel i] = min{counter}
endfor
```

Fig. 4.   Psuedocode for length–width extraction algorithm.

tion accuracy by 10%, but this had almost no effect on *Road* and *Building* classification accuracy (Table IV).

Ideally, different classes should only be classified using the spatial measures best suited for those classes. Toward that end, we developed a fuzzy classification scheme that allows the image to be hierarchically classified using different spatial measures for different sets of classes. First, the maximum-likelihood classification of the PS-MS data is used to split the data into four initial sets: *Grass-Tree, Road-Building, Water-Shadow*, and *Bare Soil*. A membership value for each class in each set is then calculated from membership functions generated from the PS-MS data plus the appropriate spatial measure. The 10×10 entropy texture measure is used for the *Grass-Tree* set, and the length–width contextual measure is used for both the *Road-Building* and the *Water-Shadow* sets. As the classification accuracy of *Bare Soil* is already high and no spatial measures were found to increase the classification accuracy of this class, only the PS-MS data is used to generate the membership value for the *Bare Soil* class. After membership values are calculated for each class in the set, the result is a fuzzy classification with each input pixel having a membership value in each class in the set. A crisp classification is generated
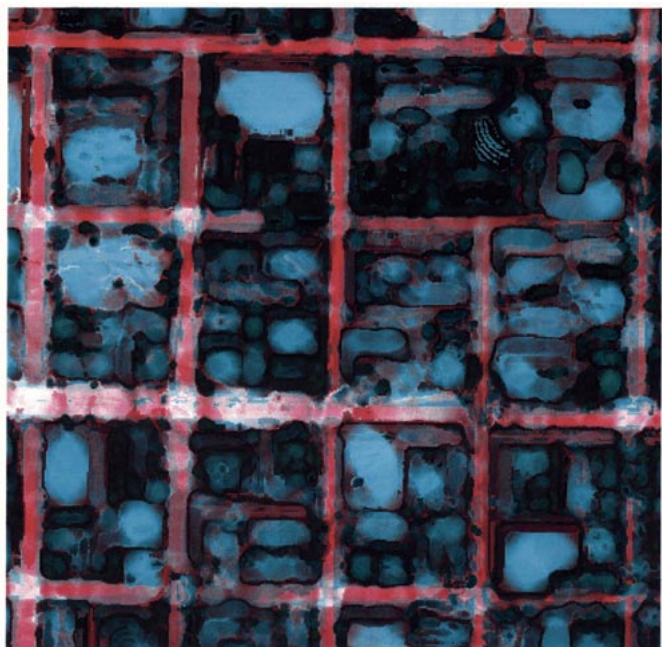
in a defuzzification step using the max operator. A block diagram of this hierarchical fuzzy classification approach is shown in Fig. 6. The membership values for each class are calculated in parallel, so the pair classification order has no influence on the final outcome. This differs from a decision-tree approach where the pair-classification branching is done sequentially and the order of the pair branching is critical in the final classification outcome. Once divided into the initial sets, pixels can only be classified as one of the set members to which they belong. This does not have a negative impact on classifier performance as the sets are chosen to include the classes that have the largest amount of spectral confusion.

A. Fuzzy Classifier Implementation

As in [12], the membership functions used for the PS-MS and entropy data are Gaussian shaped functions. The membership functions are defined with two parameters: the mean vector μ_c and covariance matrix Σ_c, which are calculated from the training data. The mean vector μ_c is used to represent the ideal pixel in class C. If an input pixel x has the value μ_c, then it will have a membership value of 1.0, and as x moves away from μ_c

$\boldsymbol{x}$ contains only the PS-MS data. Once the membership value in each class has been calculated, a primitive fuzzy membership vector is formed for $\boldsymbol{x}$

$$\vec{F}'_{\mathrm{MS}}(\vec{x}) = \left[f'^{\mathrm{MS}}_1(\vec{x}), f'^{\mathrm{MS}}_2(\vec{x}), \ldots, f'^{\mathrm{MS}}_N(\vec{x}) \right]^T \quad (8)$$

where N is the number of classes in the set. After the membership values for the PS-MS and entropy data have been calculated for each class, they are rescaled to normalize the membership values, forming the fuzzy membership vector

$$\vec{F}_{\mathrm{MS}}(\vec{x}) = \left[f^{\mathrm{MS}}_1(\vec{x}), f^{\mathrm{MS}}_2(\vec{x}), \ldots, f^{\mathrm{MS}}_N(\vec{x}) \right]^T \quad (9)$$

where

$$f^{\mathrm{MS}}_i(\vec{x}) = \frac{f'^{\mathrm{MS}}_i(\vec{x})}{\sum_{j=1}^{N} f'^{\mathrm{MS}}_j(\vec{x})}. \quad (10)$$

This normalization takes place within all classes. The vector $\boldsymbol{F}_{\mathrm{MS}}(\boldsymbol{x})$ represents the degree to which $\boldsymbol{x}$ belongs to each class in terms of the PS-MS and entropy data.

A second membership value is calculated for the pixels in the *Road-Building* and *Water-Shadow* sets using the length–width contextual measure. The length–width values are not normally distributed, so Gaussian-shaped functions are <u>not</u> appropriate for the membership functions. Instead, the membership functions are learned using a multilayer perceptron neural network. The use of a neural network allows the membership functions to be learned from training data without any prior assumptions about the distribution of the data. The multilayer perceptron was chosen for its ability to approximate arbitrarily shaped functions and because of its ease in training. The multilayer perceptron is trained using the standard back-propagation algorithm [23].

The membership functions for all the data could have been generated using the multilayer perceptron, however this approach was not chosen because the spectral and entropy data were normally distributed and best represented with Gaussian shaped functions. After the neural network has learned the membership functions from the training data of the length–width contextual data, membership values in the *Road-Building* classes and the *Water-Shadow* classes are found for the pixels in those partitions resulting in a fuzzy membership vector

$$\vec{F}_{\mathrm{LW}} = \left[f^{\mathrm{LW}}_1(\vec{x}), f^{\mathrm{LW}}_2(\vec{x}), \ldots, f^{\mathrm{LW}}_N(\vec{x}) \right]^T \quad (11)$$

where $f^{\mathrm{LW}}_i(\boldsymbol{x})$ is the membership value of pixel $\boldsymbol{x}$ in the length–width membership function for class i. The vector $\boldsymbol{F}_{\mathrm{LW}}(\boldsymbol{x})$ represents the degree to which $\boldsymbol{x}$ belongs to each class in terms of the length–width contextual measure. Because the length–width contextual measure contains no information useful for the characterization of the *Grass, Tree*, and *Bare Soil* classes, $f^{\mathrm{LW}}_i(\boldsymbol{x})$ is set to zero for those classes.

At this point each pixel has two fuzzy membership vectors, $\boldsymbol{F}_{\mathrm{MS}}(\boldsymbol{x})$ and $\boldsymbol{F}_{\mathrm{LW}}(\boldsymbol{x})$. These two vectors are combined using a fuzzy union max operator [24] to produce a single fuzzy membership vector

$$\vec{F}(x) = [f_1(\vec{x}), f_2(\vec{x}), \ldots, f_N(\vec{x})] \quad (12)$$

(a)

(b)

Fig. 5. Length–width contextual measures of (a) suburban subset shown in Fig. 2(a), and (b) urban subset shown in Fig. 2(c).

the membership value decreases. The covariance matrix Σ_c governs the width of the function. The membership value in class C for the PS-MS and entropy data is calculated as

$$f'^{\mathrm{MS}}_C(\vec{x}) = \exp\left(-0.5 \cdot (\vec{x} - \vec{\mu}_C)^T \cdot \Sigma_C^{-1} \cdot (\vec{x} - \vec{\mu}_C) \right) \quad (7)$$

and this is a scalar value representing the degree to which input vector $\boldsymbol{x}$ belongs to class C. In the case of the *Grass-Tree* set, $\boldsymbol{x}$ is a five-dimensional vector containing the PS-MS data and the 10×10 entropy texture measure. For the other three class sets (*Road-Building, Water-Shadow*, and *Bare Soil*) the input vector
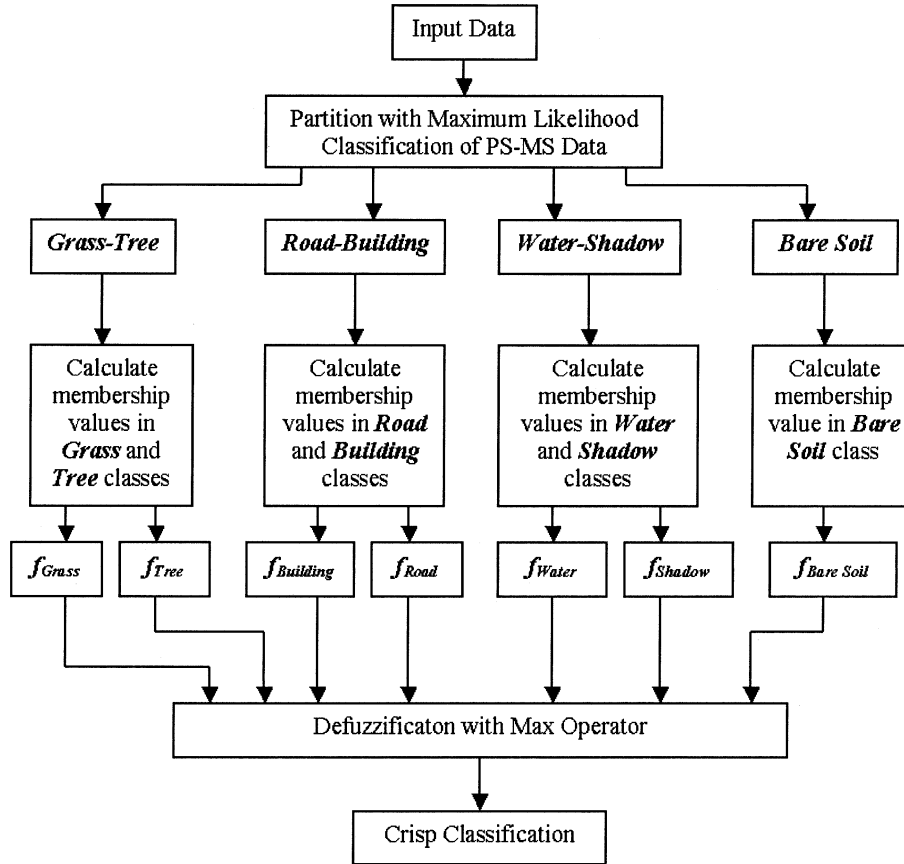
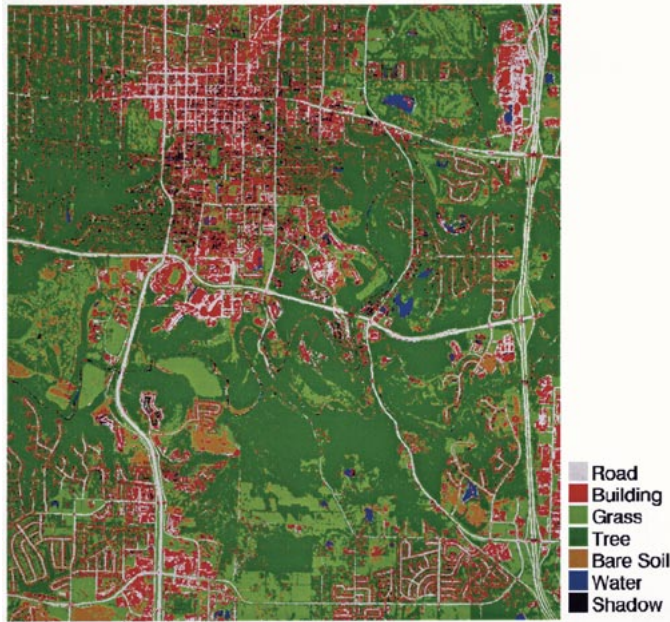Fig. 6. Block diagram of hierarchical fuzzy classification scheme.

Fig. 7. Crisp output of fuzzy classifier for Columbia, MO imagery shown in Fig. 1. Note the excellent delineation of road and building features.

where

$$f_i(\overrightarrow{x}) = \max\left\{\left(1 - \alpha_i^{\mathrm{MS}}\right) \cdot f_i^{\mathrm{MS}}(\overrightarrow{x}), \left(1 - \alpha_i^{\mathrm{LW}}\right) \cdot f_i^{\mathrm{LW}}(\overrightarrow{x})\right\} \tag{13}$$

TABLE V
OVERALL ACCURACIES OF CRISP OUTPUT OF FUZZY CLASSIFIER

Image	Overall Accuracy (%)	Kappa Coefficient
Columbia	92.7	0.909
Springfield	95.8	0.946

and α_i^{MS} and α_i^{LW} are values between 0.0 and 1.0 representing the uncertainty in the PS-MS data and the length–width contextual measure for class i. The input pixel x now has one membership value in each of the N classes. Since a crisp classification is desired, the fuzzy classification must be defuzzified to produce a single class label for each pixel in the image. Defuzzification is performed using the max operator such that x is classified as the class C with the highest membership value

$$\text{Class} = \arg\max_{j=1}^{N}\{f_1(\overrightarrow{x}), f_2(\overrightarrow{x}), \ldots, f_N(\overrightarrow{x})\}. \tag{14}$$

B. Hierarchical Fuzzy Classifier Results

The hierarchical fuzzy classifier was applied to both the Columbia and Springfield image datasets using the same training data that was used to generate the maximum-likelihood classification results presented in Section II. The classification map of the Columbia imagery generated using the fuzzy classifier

TABLE VI
CONFUSION MATRIX FOR CRISP OUTPUT OF FUZZY CLASSIFICATION OF 1-m PS-MS COLUMBIA IMAGE DATASET

	Road Ref.	Building Ref.	Grass Ref.	Tree Ref.	Bare Soil Ref.	Water Ref.	Total	%
Road	10,147	1019	2	1	231	16	11,416	89
Building	982	12,513	52	4	140	306	13,997	89
Grass	99	23	25,458	53	0	0	25,633	99
Tree	62	57	1288	11,227	0	0	12,634	89
Bare Soil	177	49	10	0	8864	0	9100	97
Water	2	318	0	0	0	6798	7118	96
Shadow	18	931	0	0	0	48	997	
Total	11,487	14,910	26,810	11,285	9235	7168	80,895	
%	88	84	95	99	96	95		

is shown in Fig. 7. The accuracy assessments of the crisp classifications using the hierarchical fuzzy classifier are shown in Table V. The overall accuracy of the Columbia image increased by approximately 11% over the maximum-likelihood accuracy when the fuzzy classification scheme was implemented. Moreover, the Kappa coefficient increased by 0.146. The overall accuracy of the Springfield image increased by approximately 8% over the maximum-likelihood accuracy when the fuzzy classification scheme was implemented and the Kappa coefficient increased by 0.106.

The confusion matrix for the crisp output of the fuzzy classification of the Columbia image is shown in Table VI. The average *Road-Building* classification accuracy increased from 71% to 86%, and the average *Grass-Tree* accuracy increased from 87% to 97%. In addition, the *Water* classification accuracy increased from 69% to 95%. Fig. 8 shows the crisp classification of suburban and urban area subsets from the Columbia image. As the classification maps in Fig. 8 show, the fuzzy classifier performs better in suburban areas than in urban areas, where the problems of spectral overlap and within class variance are most severe. However, when the classification maps in Figs. 8 and 2 are compared, it is clear that the fuzzy classifier outperforms the maximum-likelihood classifier in both suburban and urban areas. The confusion matrix for the crisp output of the fuzzy classification of the Springfield image is shown in Table VII. The average *Road-Building* classification accuracy increased from 70% to 92%. The average *Grass-Tree* accuracy remained at 99%. In addition, the *Water* classification rate increased from 72% to 93%.

For comparison purposes, the hierarchical fuzzy classifier was applied to the 4-m MS Columbia dataset. The same training and reference sites used for the Columbia PS-MS dataset were used, however with the decrease in resolution of the imagery, the number of training and reference pixels decreased accordingly. All algorithm parameters were kept the same except the maximum length for the length–width feature extraction algorithm was decreased from $L_{max} = 200$ pixels to $L_{max} = 50$ pixels, reflecting the decrease in resolution of the imagery. Also, the imagery was not smoothed with a median filter prior to application of the length–width feature extraction. The entropy texture measure was calculated using a 7×7 pixel window (28×28 m). This window size was the best compromise between minimizing edge effects and still extracting usable information from the objects of interest in the image.
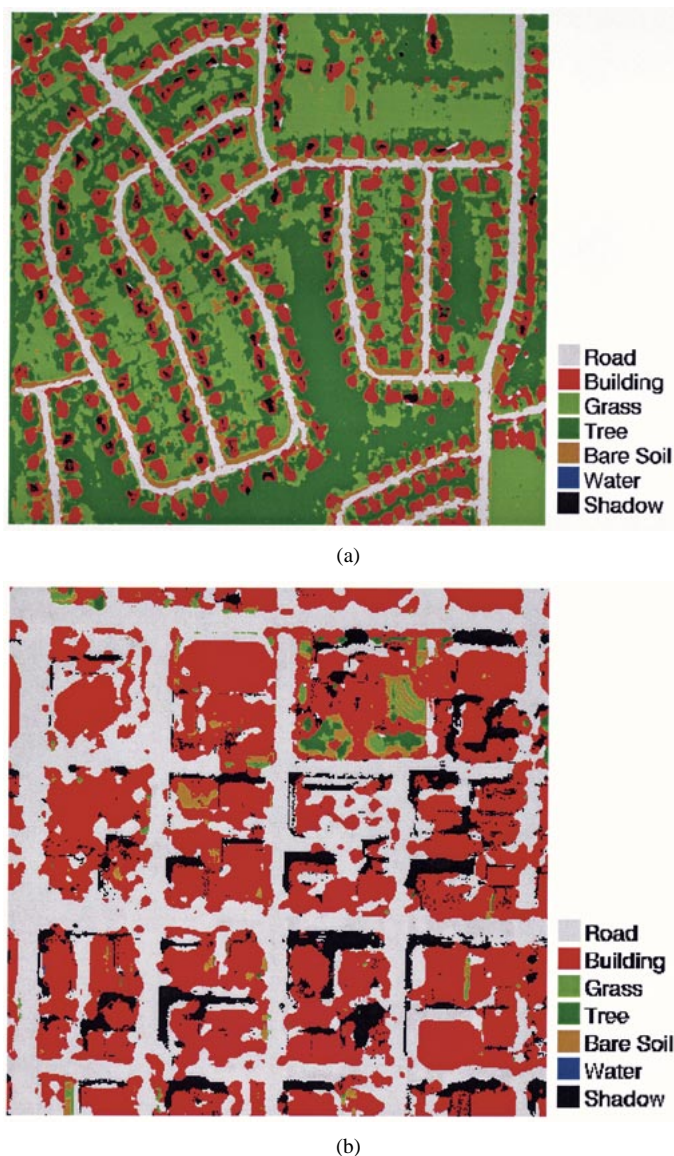

(a)

(b)

Fig. 8. Crisp output of fuzzy classifier for (a) suburban scene, (b) urban scene from the Columbia, MO image subsets shown in Fig. 2(a) and (c), respectively. Note the significant improvement over the maximum-likelihood classification results also shown in Fig. 2.

The classification accuracies for the 4-m MS Columbia dataset are presented in Table VIII. The confusion matrices from the fuzzy and maximum-likelihood classifications of the 4-m MS Columbia dataset are shown in Tables IX and X

TABLE VII
CONFUSION MATRIX FOR CRISP OUTPUT OF FUZZY CLASSIFICATION OF 1-m PS-MS SPRINGFIELD IMAGE DATASET

	Road Ref.	Building Ref.	Grass Ref.	Tree Ref.	Bare Soil Ref.	Water Ref.	Total	%
Road	28,967	4215	12	1	423	0	33,618	86
Building	1220	30,617	127	23	1000	0	32,987	93
Grass	54	16	48,755	2	1	0	48,828	100
Tree	15	2	247	49,493	0	206	49,963	99
Bare Soil	0	5	0	0	13,266	0	13,271	100
Water	6	0	0	0	0	5183	5189	100
Shadow	8	0	0	0	0	192	200	
Total	30,270	34,855	49,141	49,519	14,690	5581	184,056	
%	96	88	99	100	90	93		

TABLE VIII
CLASSIFICATION ACCURACIES FOR 4-m MS COLUMBIA IMAGE DATASET

	Overall Accuracy (%)	Kappa Coefficient
Fuzzy Classifier	83.5	0.791
Max. Likelihood	78.6	0.731

TABLE IX
CONFUSION MATRIX FOR CRISP OUTPUT OF FUZZY CLASSIFICATION OF 4-m MS COLUMBIA IMAGE DATASET

	Road Ref.	Building Ref.	Grass Ref.	Tree Ref.	Bare Soil Ref.	Water Ref.	Total	%
Road	559	79	0	16	39	3	696	80
Building	145	821	34	30	54	18	1102	74
Grass	6	6	1579	278	0	0	1869	84
Tree	0	2	28	407	0	0	437	93
Bare Soil	9	4	0	0	487	0	500	97
Water	0	47	0	0	0	409	456	90
Shadow	0	12	0	0	0	33	45	
Total	719	971	1641	731	580	463	5105	
%	78	85	96	56	84	88		

TABLE X
CONFUSION MATRIX FOR MAXIMUM LIKELIHOOD CLASSIFICATION OF 4-m MS COLUMBIA IMAGE DATASET

	Road Ref.	Building Ref.	Grass Ref.	Tree Ref.	Bare Soil Ref.	Water Ref.	Total	%
Road	521	135	11	0	57	3	727	72
Building	183	765	23	46	36	18	1073	71
Grass	6	6	1560	273	0	0	1845	85
Tree	0	2	47	412	0	0	461	89
Bare Soil	9	4	0	0	487	0	500	97
Water	0	30	0	0	0	269	299	90
Shadow	0	29	0	0	0	173	202	
Total	719	971	1641	731	580	463	5105	
%	72	79	95	56	84	58		

for comparison. As was the case with the 1-m PS-MS data, the hierarchical fuzzy classification accuracy for the 4-m MS data is higher ($\sim$6%) than the maximum-likelihood classification accuracy. However, the increase in accuracy from the maximum-likelihood classification to the fuzzy classification was larger ($\sim$12%) for the 1-m PS-MS data. The confusion between the *Road* and *Building* classes is decreased, however there is little change in the classification accuracies of the *Grass* and *Tree* classes. The most likely explanation for this is that the texture information useful for discrimination between these two classes is represented primarily in the 1-m resolution panchromatic band. Thus, it is clear that the 1-m PS-MS

imagery is better suited for urban land cover mapping than the 4-m MS imagery by itself.

C. Postprocessing

A majority filter was implemented to operate on the *Water, Shadow, Road*, and *Building* classes to increase the accuracy of the fuzzy classification result and clean up the appearance of the classification image. A majority filter operates by extracting a window of pixels around the pixel of interest and reclassifies the central pixel as the class with the largest number of pixels in the window. The majority filter was first applied to the *Water* class, but instead of allowing the *Water* pixels to be

reclassified as any class, the *Water* pixels were only allowed to be reclassified as *Water, Shadow, Road*, or *Building*. This was done to keep *Water* pixels from being reclassified into one of the vegetation classes.

After the *Water* pixels, the *Shadow* pixels were majority filtered next. The pixels were reclassified as *Water, Road*, or *Building* thus removing the *Shadow* class from the image. It is important to remove the *Shadow* class, as it is not a real urban land cover class. Finally, the *Road* and *Building* pixels were majority filtered and reclassified as *Road, Building, Water*, or *Bare Soil*. As was the case with the other majority-filtered classes, *Road* and *Building* pixels were not allowed to be reclassified as one of the vegetation classes. The result of the majority filter postprocessing is a modest increase in classification accuracy of 1% to 2% and a more spatially coherent classification image.

V. CONCLUSION

The results presented here demonstrate the usefulness of high-resolution satellite imagery for urban land cover mapping and some of the shortcomings of conventional classification techniques such as maximum likelihood. It was found that maximum-likelihood classification of high-resolution multispectral imagery over urban areas produced significant amounts of misclassification errors between spectrally similar classes such as *Road* and *Building* classes. Different spatial measures such as texture and contextual methods were investigated and found to increase the discrimination between certain spectrally similar classes. In particular, the 10×10 entropy texture window measure and the length–width contextual measures were both found to increase discrimination between the *Grass-Tree* and *Road-Building* classes, respectively.

Finally, a hierarchical fuzzy classification method was developed that utilized both spectral and spatial information to classify the data. The classification accuracies of the fuzzy classifier were approximately 10% greater than the maximum-likelihood classification results for 1-m PS-MS image datasets. Accordingly, there were significant decreases in the number of misclassifications between spectrally similar classes. Further work is needed to improve the performance of the fuzzy classifier in dense urban areas and to produce even more detailed urban land cover maps by identifying features such as parking lots and side walks. We believe an image segmentation approach combined with morphological feature operators may be used to further improve upon the results presented here.

ACKNOWLEDGMENT

REFERENCES

[1] C. H. Davis and X. Wang, Planimetric accuracy of Ikonos 1-m panchromatic orthoimage products and their utility for local government GIS basemap applications, in *Int. J. Remote Sens.*, to be published.

[2] J. R. Jenson and D. C. Cowen, "Remote sensing of urban/suburban infrastructure and socio-economic attributes," *Photogramm. Eng. Remote Sens.*, vol. 65, no. 5, pp. 611–622, May 1999.

[3] J. A. Benediktsson, K. Arnason, and M. Persaresi, "The use of morphological profiles in classification of data from urban areas," *Proc. IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion Over Urban Areas*, pp. 30–34, Nov. 2002.

[4] A. J. Tatem, H. G. Lewis, P. M. Atkinson, and M. S. Nixon, "Super-resolution mapping of urban scenes from IKONOS imagery using a hopfield neural network," in *Proc. IGARSS*, vol. 7, 2001, pp. 3203–3205.

[5] I. Couloigner and T. Ranchin, "Mapping of urban areas: A multiresolution modeling approach for semi-automatic extraction of streets," *Photogramm. Eng. Remote Sens.*, vol. 66, no. 7, pp. 867–874, July 2000.

[6] C. Steger, "An unbiased detector of curvilinear structures," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, pp. 113–125, Feb. 1998.

[7] T. Chen, J. Wang, and K. Zhang, "A wavelet transform based method for road extraction from high-resolution remotely sensed data," in *Proc. IGARSS*, vol. 6, Toronto, ON, Canada, June 24–28, 2002, pp. 1621–1623.

[8] M. Pesaresi, "Textural classification of very high-resolution satellite imagery: Empirical estimation of the interaction between window size and detection accuracy in urban environment," *Proc. ICIP*, vol. 1, pp. 114–118, Oct. 1999.

[9] P. van Teeffelen, S. de Jong, and L. van der Berg, "Urban monitoring: New possibilities of combining high spatial resolution IKONOS images with contextual image analysis techniques," *Proc. IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion Over Urban Areas*, pp. 265–269, Nov. 2002.

[10] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, pp. 309–320, Feb. 2001.

[11] F. P. Kressler, T. B. Bauer, and K. T. Steinnocher, "Object-oriented per-parcel land use classification of very high resolution images," *Proc. IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion Over Urban Areas*, pp. 164–167, Nov. 2002.

[12] F. Melgani, B. A. R. AL Hashemy, and S. M. R. Taha, "An explicit fuzzy supervised classification method for multispectral remote sensing images," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, pp. 287–295, Jan. 2000.

[13] J. R. Jenson, *Introductory Digital Image Processing: A Remote Sensing Perspective*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1996.

[14] J. Vrabel, "Multispectral imagery advanced band sharpening study," *Photogramm. Eng. Remote Sens.*, vol. 66, no. 1, pp. 73–79, Jan. 2000.

[15] R. G. Congalton, R. G. Oderwald, and R. A. Mead, "Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques," *Photogramm. Eng. Remote Sens.*, vol. 49, no. 12, pp. 1671–1678, Dec. 1983.

[16] W. D. Hudson and C. W. Ramm, "Correct formulation of the kappa coefficient of agreement," *Photogramm. Eng. Remote Sens.*, vol. 53, no. 4, pp. 421–422, Apr. 1987.

[17] R. G. Congalton, "A review of assessing the accuracy of classifications of remotely sensed data," *Remote Sens. Environ.*, vol. 37, pp. 35–46, 1991.

[18] J. B. Campbell, *Introduction to Remote Sensing*, 2nd ed. New York: Guilford, 1996.

[19] R. M. Haralik, K. Shanmugam, and D. Its'hak, "Textural features for image classification," *IEEE Trans. Syst. Man Cybnet.*, vol. SMC-3, pp. 610–621, 1973.

[20] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 2002.

[21] C. J. S. Ferro and T. A. Warner, "Scale and texture in digital image classification," *Photogramm. Eng. Remote Sens.*, vol. 68, no. 1, pp. 51–63, Jan. 2002.

[22] *Handbook of Image and Video Processing*, A. Bovik, Ed., Academic, San Diego, CA, 2000, pp. 101–116. Morphological filtering for image enhancement and detection.

[23] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Upper Saddle River, NJ: Prentice-Hall, 1999.

[24] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Upper Saddle River, NJ: Prentice-Hall, 1995.

Aaron K. Shackelford (S'97) was born in Kansas City, MO, on January 1, 1977. He received the B.S. and M.S. degrees in electrical engineering from the University of Missouri-Columbia, Columbia, in 1999 and 2001, respectively. He is currently pursuing the Ph.D. degree in electrical engineering from the University of Missouri-Columbia.

Since January of 2000, he has been a Research Assistant in the Department of Electrical Engineering, University of Missouri-Columbia. He was a Research Scholar in the Department of Electronic Engineering, City University of Hong Kong, Hong Kong, for three months in 2000. He is currently a Research Assistant in the Remote Sensing Laboratory, University of Missouri-Columbia. His research interests include application of pattern recognition approaches to remote sensing imagery and patch antenna design.

Mr. Shackelford is a member of Tau Beta Pi. He was awarded the NASA Graduate Student Researchers Program fellowship in 2001.
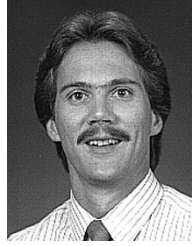
Curt H. Davis (S'90–M'92–SM'98) was born in Kansas City, MO, on October 16, 1964. He received the B.S. and Ph.D. degrees in electrical engineering from the University of Kansas, Lawrence, in 1988 and 1992, respectively.

He has been actively involved in experimental and theoretical aspects of microwave remote sensing of the ice sheets since 1987. He has participated in two field expeditions to the Antarctic continent and one to the Greenland ice sheet. From 1989 to 1992, he was a NASA Fellow at the Radar Systems and Remote Sensing Laboratory, University of Kansas where he conducted research on ice-sheet satellite altimetry. He is currently the Croft Distinguished Professor of Electrical and Computer Engineering at the University of Missouri-Columbia. His research interests are in the areas of mobile radio signal propagation, RF/microwave systems, satellite remote sensing, and remote sensing applications for urban environments.

Dr. Davis is a member of the Tau Beta Pi, Eta Kappa Nu, and URSI-Commission F. He is a former Chairman of the Instrumentation/Future Technologies committee of the IEEE Geoscience and Remote Sensing Society. In 1996, he was selected by the International Union of Radio Science for their Young Scientist Award. He was awarded the Antarctica Service Medal from the National Science Foundation.