in I. Balderjahn, R. Mathar and M. Schader (eds.): Data Highways and Information Flooding, a Challenge for Classification and Data Analysis, Springer 1997.
21st Annual Meeting of the Gesellschaft für Klassifikation, GfKl'97, Potsdam, March 12–14, 1997.

Unsupervised Fuzzy Classification of Multispectral Imagery Using Spatial-Spectral Features

Rafael Wiemker

II. Institut für Experimentalphysik, Universität Hamburg,
KOGS / Informatik, Vogt-Kölln-Str. 30, D-22527 Hamburg, Germany,
http://kogs-www.informatik.uni-hamburg.de/~wiemker

Abstract: Pixel-wise spectral classification is a widely used technique to produce thematic maps from remotely sensed multispectral imagery. It is commonly based on purely spectral features. In our approach we additionally consider additional spatial features in the form of local context information. After all, spatial context is the defining property of an image. Markov random field modeling provides the assumption that the probability of a certain pixel to belong to a certain class depends on the pixel's local neighborhood. We enhance the ICM algorithm of Besag (1986) to account for the fuzzy class membership in the fuzzy clustering algorithm of Bezdek (1973). The algorithm presented here was tested on simulated and real remotely sensed multispectral imagery. We demonstrate the improvement of the clustering as achieved by the additional spatial fuzzy neighborhood features.

1 Introduction

Spectral classification is a widely used technique to produce *thematic maps* from remotely sensed multispectral imagery. The *classification* or *labeling* of each pixel relies on its spectrum or *spectral signature*, which consists of n spectral values of the spectral bands $i = [1 \dots n]$ of wavelength λ_i. The similarity between an observed spectral vector $\mathbf{x} = [\dots, x(\lambda_i), \dots]^t \in \mathbb{R}^n$ and a given reference spectrum $\mathbf{m}$ is commonly evaluated by the *spectral distance* $d = \|\mathbf{x} - \mathbf{m}\|$ in the feature space. The n-dimensional spectral feature space is spanned by the radiance or reflectance signals $x(\lambda_i)$ as received in the n spectral bands of the sensor or camera.

Given the set of all observed spectral vectors $\mathbf{x}$ corresponding to the pixels of a particular multispectral image, unsupervised clustering algorithms are employed to find k cluster centers in the spectral feature space $\mathbb{R}^n$ around which the observed spectra are scattered. Each pixel of the image can then be classified (labeled) to the class to which its spectral distance is minimal.

Most classification techniques applied in multispectral remote sensing (Richards 1993) rely on purely *spectral* features and consider only one pixel at a time. More recently, a method for utilizing additional contextual information from neighboring pixels has been derived from Markov random field modeling (Besag 1986). This 'ICM-algorithm' has been shown to improve classification results on multispectral imagery (Solberg et al. 1996). So far, this spectral-spatial labeling approach has been used in conjunction with supervised classification only, *i.e.*, the reference classes were established from training data by the analyst. In this paper we describe the effects of incorporating spatial context information into *unsupervised* clustering techniques such as the hard and fuzzy k-means algorithms.

2 Hard and Fuzzy Clustering with Spectral and Spatial Features

All k-means algorithms or 'migrating means'-algorithms work iteratively. Also, all k-means algorithms, as used in multispectral image classification, determine for each pixel the Euclidean distance $d(\mathbf{x}|\mathbf{m}_c) = \sqrt{\sum_i (x_i - m_{c,i})^2}$ between the spectral vector $\mathbf{x}$ and the respective mean spectrum $\mathbf{m}_c$ of each class ω_c $(c = 1, \dots, k)$ in the spectral feature space which is spanned by the n spectral bands i of the imaging sensor $(i = 1, \dots, n)$, where $x_i = x(\lambda_i)$.

The hard k-means algorithm (Ball and Hall 1967) assigns each pixel to the class ω_c to which the spectral distance $d(\mathbf{x}|\mathbf{m}_c)$ is minimal. Then the k cluster centers $\mathbf{m}_c$ are recomputed as the means of all pixels which currently belong to the respective class. The process is repeated until convergence.

The fuzzy k-means algorithm (originally 'c-means', Bezdek 1973) relies on a fuzzy membership $p_{\text{spec}}(\mathbf{x}|\omega_c)$ which is inversely proportional to the spectral distance $d(\mathbf{x}|\mathbf{m}_c)$:

$$p_{\text{spec}}(\mathbf{x}|\omega_c) \quad = \quad \frac{d^{-1}(\mathbf{x}|\mathbf{m}_c)}{\sum_{c'} \left[d^{-1}(\mathbf{x}|\mathbf{m}_{c'})\right]} \quad , \quad \sum_{c=1}^{k} p_{\text{spec}}(\mathbf{x}|\omega_c) = 1 \quad . \tag{1}$$

The algorithm iterates two alternating steps: (a) updating the membership weights $p_{\text{spec}}(\mathbf{x}|\omega_c)$, and (b) re-estimating the cluster centers $\mathbf{m}_c = \sum_{\mathbf{x}} p_{\text{spec}}(\mathbf{x}|\omega_c)\,\mathbf{x} / \sum_{\mathbf{x}} p_{\text{spec}}(\mathbf{x}|\omega_c)$. In contrast to the hard k-means, *all* pixels are used for the computation of each cluster center, weighted with their respective fuzzy membership $p_{\text{spec}}(\mathbf{x}|\omega_c)$. Convergence to a local minimum of a global objective function has been shown (Bezdek 1981).

Considering contextual image information, we use Markov random field modelling, and assume that the conditional spatial probability $p_{\text{spat}}(\mathbf{x}|\omega_c)$ of pixel $\mathbf{x}$ depends only on the pixels $\mathbf{x}'$ in its spatial neighborhood $\mathcal{N}(\mathbf{x})$ (Li 1995). As the neighborhood $\mathcal{N}(\mathbf{x})$ we here use a $l \times l$ window around pixel $\mathbf{x}$ except the pixel itself.

For the interaction between neighboring pixels, Besag (1986) has suggested a neighborhood potential $U(\mathbf{x}) = \sum_{\mathbf{x}' \in \mathcal{N}(\mathbf{x})} [1 - \delta(\mathbf{x}, \mathbf{x}')]$, with $\delta(\mathbf{x}, \mathbf{x}') = 1$ for equal classes $\omega(\mathbf{x}) = \omega(\mathbf{x}')$, and 0 otherwise.

In this paper we use a refined 'fuzzy' neighborhood potential $U(\mathbf{x}|\omega_c)$, based on the current memberships $P(\omega_c|\mathbf{x}')$ (defined in Eq. 4) of the neighboring pixels $\mathbf{x}'$. The potential $U(\mathbf{x}|\omega_c)$ and then the spatial membership $p_{\text{spat}}(\mathbf{x}|\omega_c)$ are defined as

$$U(\mathbf{x}|\omega_c) = \sum_{\mathbf{x}' \in \mathcal{N}(\mathbf{x})} [1 - P(\omega_c|\mathbf{x}')] \quad (2)$$

$$p_{\text{spat}}(\mathbf{x}|\omega_c) = \frac{1}{Z} e^{-\beta U(\mathbf{x}|\omega_c)} \quad , \quad \sum_{c=1}^{k} p_{\text{spat}}(\mathbf{x}|\omega_c) = 1 \quad (3)$$

where $\beta > 0$ is a factor to weight the influence of the spatial context. The spatial membership $p_{\text{spat}}(\mathbf{x}|\omega_c)$ for class ω_c is large if the neighboring pixels $\mathbf{x}'$ have large memberships $P(\omega_c|\mathbf{x}')$ for the same class ω_c, and small if they tend to belong to other classes $\omega_{c'}$. Computation of the normalization constant Z is unnecessary here, as it cancels out in Eq. (4). The joint spectral-spatial membership $P(\omega_c|\mathbf{x})$ of pixel $\mathbf{x}$ to belong to class ω_c then is defined to be :

$$P(\omega_c|\mathbf{x}) = \frac{p_{\text{spec}}(\mathbf{x}|\omega_c) \cdot p_{\text{spat}}(\mathbf{x}|\omega_c)}{\sum_{c'} [p_{\text{spec}}(\mathbf{x}|\omega_{c'}) \cdot p_{\text{spat}}(\mathbf{x}|\omega_{c'})]} \quad , \quad \sum_{c=1}^{k} P(\omega_c|\mathbf{x}) = 1 \quad . \quad (4)$$

Using the additional spatial features, we again can perform hard and fuzzy k-means clustering, depending on whether the cluster centers are re-estimated from only those pixels currently assigned to each cluster, or from all pixels using their current memberships as weights.

3 Results on Simulated and Remotely Sensed Multispectral Imagery

In order to evaluate the effect of the additional contextual memberships (Eq. 3), we have simulated a test image with $n = 2$ spectral bands and $k = 2$ spectral classes ω_1 and ω_2 (Fig. 1). The spectral vectors $\mathbf{x} = [x_1, x_2]^t$ of each class are scattered randomly around the two cluster centers $\mathbf{m}_{\omega_1}$ and $\mathbf{m}_{\omega_2}$, forming uncorrelated Gaussian distributions. In the original spectral feature space the two clusters (Fig. 1, right, true cluster centers marked by crosses) are almost indistinguishable by eye appraisal due to the extreme scatter. The root mean square scatter of both clusters is equal to the Euclidean distance between the cluster centers.

A number of runs was performed with four different algorithms. The number of clusters $k = 2$ and random initial centers (seed coordinates) for the cluster centers were supplied. The resulting classification accuracies (i.e., the relative number of correctly labeled pixels) and the cluster center estimation accuracy (mean relative deviation of coordinates between true and estimated class centers) are given in Table 1. Typical classification results of each method are shown in Fig. 2. We observe that the fuzzy k-means performs slightly better in the estimation of the class center coordinates, but not in the classification. Also, the hard k-means is improved by the additional spatial features in classification, but not in cluster center accuracy. For fuzzy k-means with additional contextual memberships, however, the results indicate clearly that not only the classification results are improved, but that also the coordinates of the cluster centers are estimated with significantly improved accuracy.

Good convergence of the iterative algorithm can be achieved by starting the iteration with $\beta = 0$, *i.e.*, without spatial influence, and then increasing β gradually towards $\beta = 1$. For each intermediate β-value, convergence is waited for before the spatial influence is increased (Fig. 4).

Another interesting observation is that the classification and cluster center accuracy does not deteriorate when the number of classes k is over-estimated. We performed another series of runs with $k = 3$ and $k = 4$ and random cluster seeds provided. With a fuzzy k-means on purely spectral features this decreases the classification accuracy as well as the cluster center estimation (Fig. 3). In contrast, with contextual fuzzy k-means the superfluous classes are basically not populated at all, and the centers of the actually existing Gaussian distributions are found correctly. Classification and cluster center accuracy with over-estimated $k = 4$ clusters is indeed the same as for the correct $k = 2$ case (Table 1, bottom line).

The fuzzy clustering with purely spectral features on the one hand, and with spectral-spatial features on the other hand, was applied to remotely sensed multispectral scanner imagery (flight altitude 300 m) with $n = 10$ spectral bands. The airborne line scanner of DAEDALUS, Inc., is operated on board of a Do 228 aircraft by the German Aerospace Research Establishment (DLR).

Some examplary classification results can be inspected in Fig. 5. The classification results which utilize both spectral and spatial features appear smoother and without grainy, pixel-size noise. Note that not only the pixel-wise classification but also the cluster centers differ when utilizing additional spatial features. The such established clusters are clearly better suitable for thematic image segmentation (see e.g. the forest areas in Fig. 5, top).

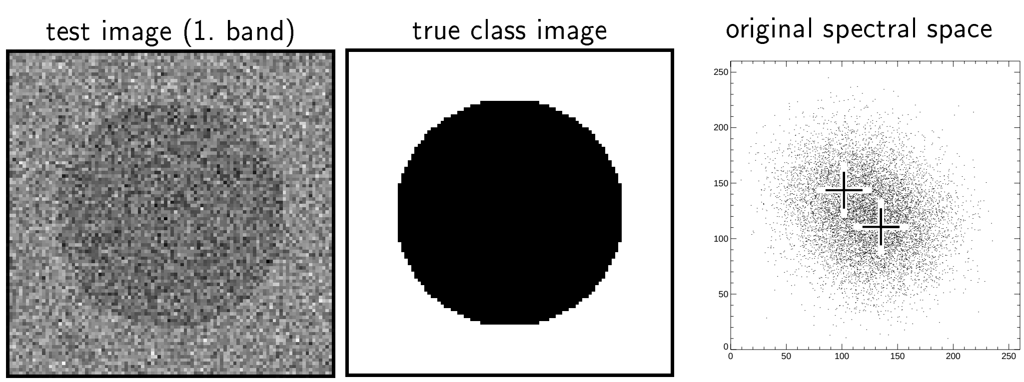

Figure 1: Simulated test image (left) with $n = 2$ spectral bands and $k = 2$ spectral classes (center) with rms scatter equal to the distance between the cluster centers in the spectral feature space (right).

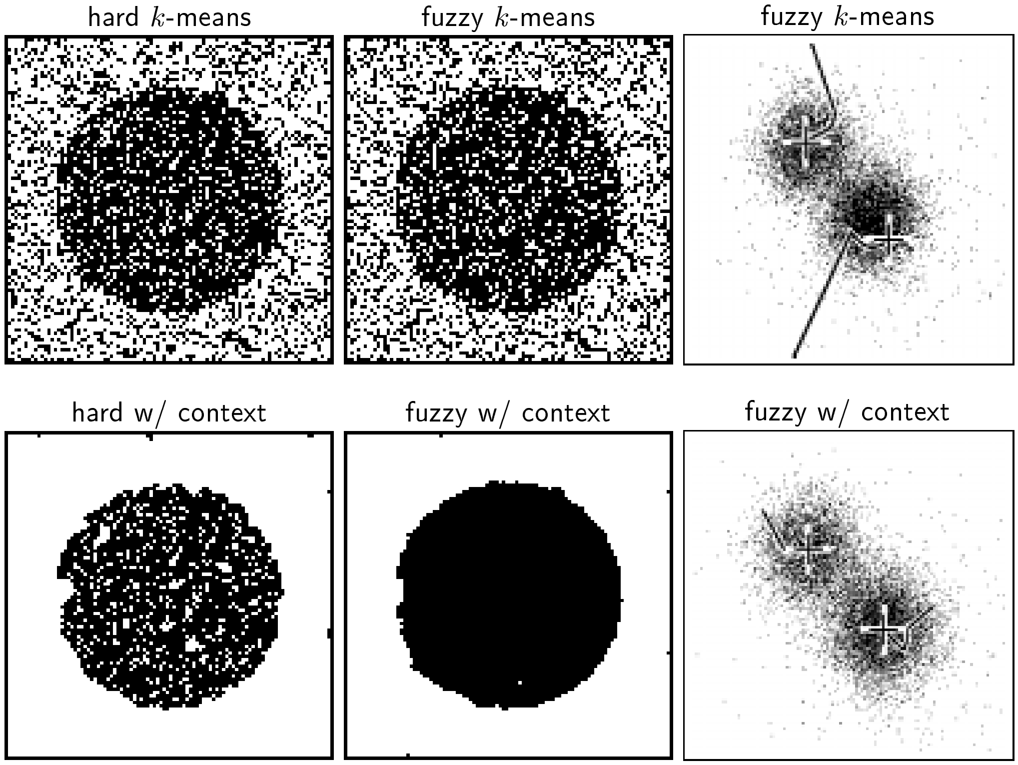

Figure 2: Typical classification results of the various algorithms (left and center). On the right-hand side, estimated cluster centers are depicted in the spectral feature space. The path of the migrating means from random seeds is indicated by solid lines. (For improved visualization, the original feature space was magnified and the scatter reduced.)

algorithm	**classification accuracy** (correctly classified pixels)	**cluster center accuracy** (deviation from true centers)
hard k-means	75% (±1%)	5% (±1.5%)
with context	91% (±1%)	5% (±1.5%)
fuzzy k-means	75% (±1%)	3.5% (±1%)
with context	**99%** (±1%)	**0.3%** (±0.1%)

Table 1: Results on simulated data. Accuracy of cluster center estimation and classification for various algorithms. Error margins are estimated from several runs with random cluster seeds.

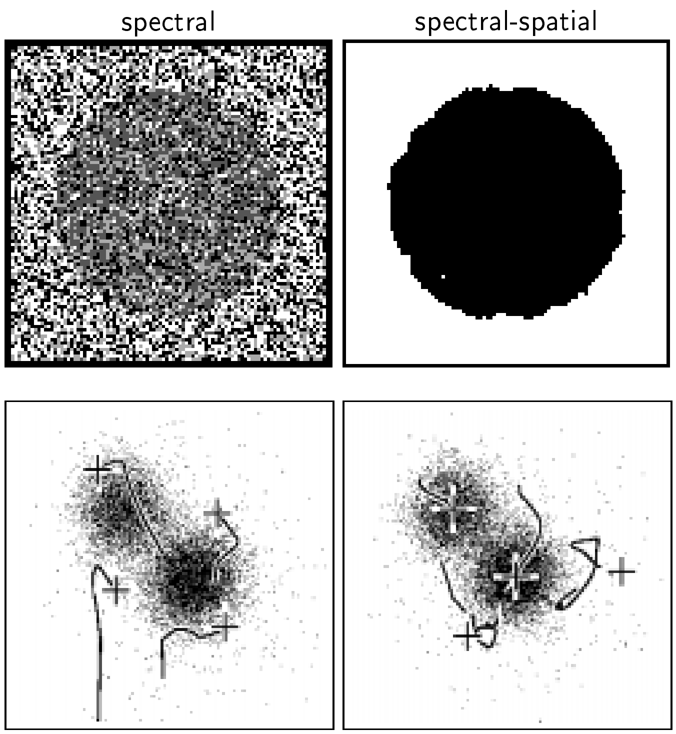

Figure 3: Typical classification results (top), and cluster centers found (bottom), with over-estimated $k = 4$ (instead of the correct $k = 2$) for fuzzy k-means clustering,
on purely spectral (left) and spatial-spectral features (right). On the right-hand side, estimated cluster centers are depicted in the spectral feature space. The path of the migrating means from random seeds is indicated by solid lines. (For improved visualization, the original feature space was magnified and the scatter reduced.)

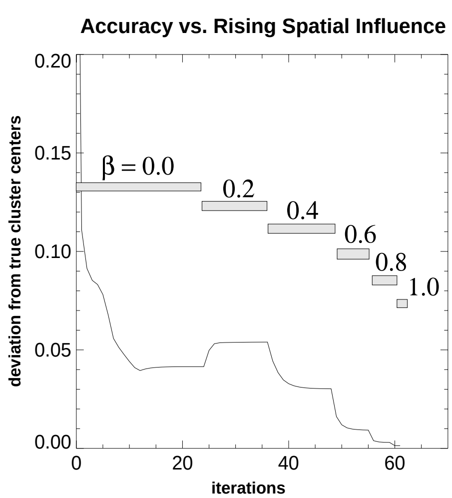

Figure 4: Typical iteration process. Convergence is achieved for each level of spatial influence β. The spatial influence is raised gradually and yields increasing accuracy of the cluster center estimation.

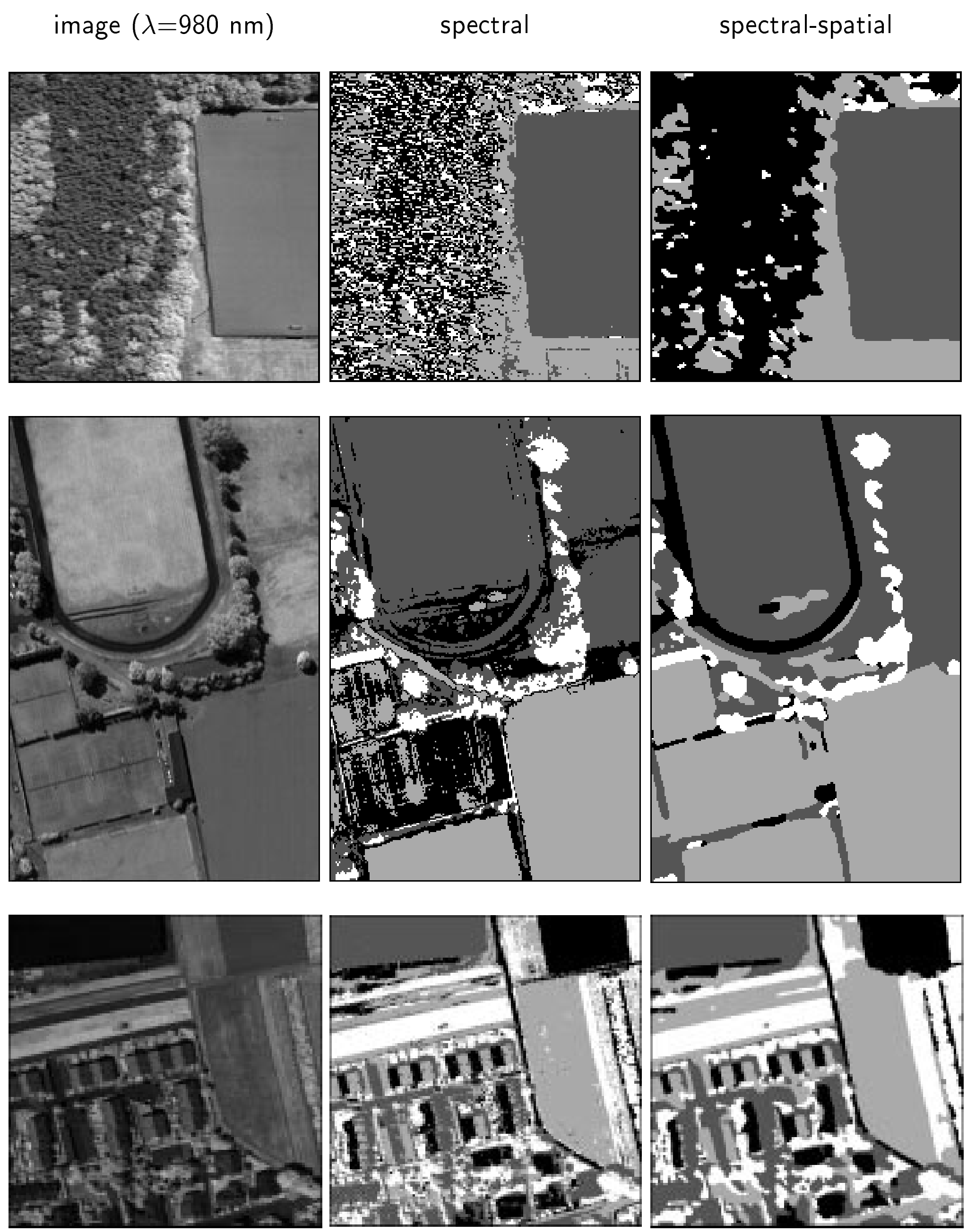

Figure 5: Multispectral aerial imagery of Nürnberg with $n = 10$ spectral bands ($\lambda_1 = 435\,\mathrm{nm}$, $\lambda_{10} = 2215\,\mathrm{nm}$), unsupervisedly classified by fuzzy k-means clustering into $k = 4$ classes (recorded in 1995, 300 m altitude, atmospherically corrected, various scales).

4 Conclusions

We have tested the effects of using spatial context features in addition to spectral features for unsupervised clustering and classification of multispectral imagery. Our observations with simulated and remotely sensed multispectral imagery can be summarized as follows:

- ▷ The additional use of spatial features can significantly improve the classification (labeling) results of unsupervised clustering. The full benefits of additional spatial features are experienced when used in conjunction with fuzzy clustering (in contrast to hard clustering).
- ▷ Moreover, also the accuracy of the cluster center coordinates as estimated by unsupervised clustering is significantly improved when using additional spatial features in conjunction with *fuzzy* k-means.
- ▷ The additional spatial features avoid effectively the deteriorating effect of over-estimating the number of clusters k. The cluster centers are found accurately by fuzzy k-means clustering even if the spectral feature space in fact contains fewer than k clusters of Gaussian distribution.

References

BALL, G. and HALL, D. (1967): A Clustering Technique for Summarizing Multivariate Data. *Behavioral Sciences, 12, 153–155.*

BESAG, J. (1986): On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society B, 48, 3, 259–302.*

BEZDEK, J.C. (1973): Fuzzy Mathematics in Pattern Classification. PhD-thesis, Applied Math Center, Cornell University, Ithaca.

BEZDEK, J.C. (1981): Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York.

LI, S.Z. (1995): Markov Random Field Modeling in Computer Vision, Springer, Tokyo.

RICHARDS, J.A. (1993): Remote Sensing Digital Image Analysis. Springer, Heidelberg, New York.

SOLBERG, A.H.S., TAXT, T., and JAIN, A.K. (1996): A Markov Random Field Model for Classification of Multisource Satellite Imagery. *IEEE Transactions on Geoscience and Remote Sensing, 34, 1, 100–134.*

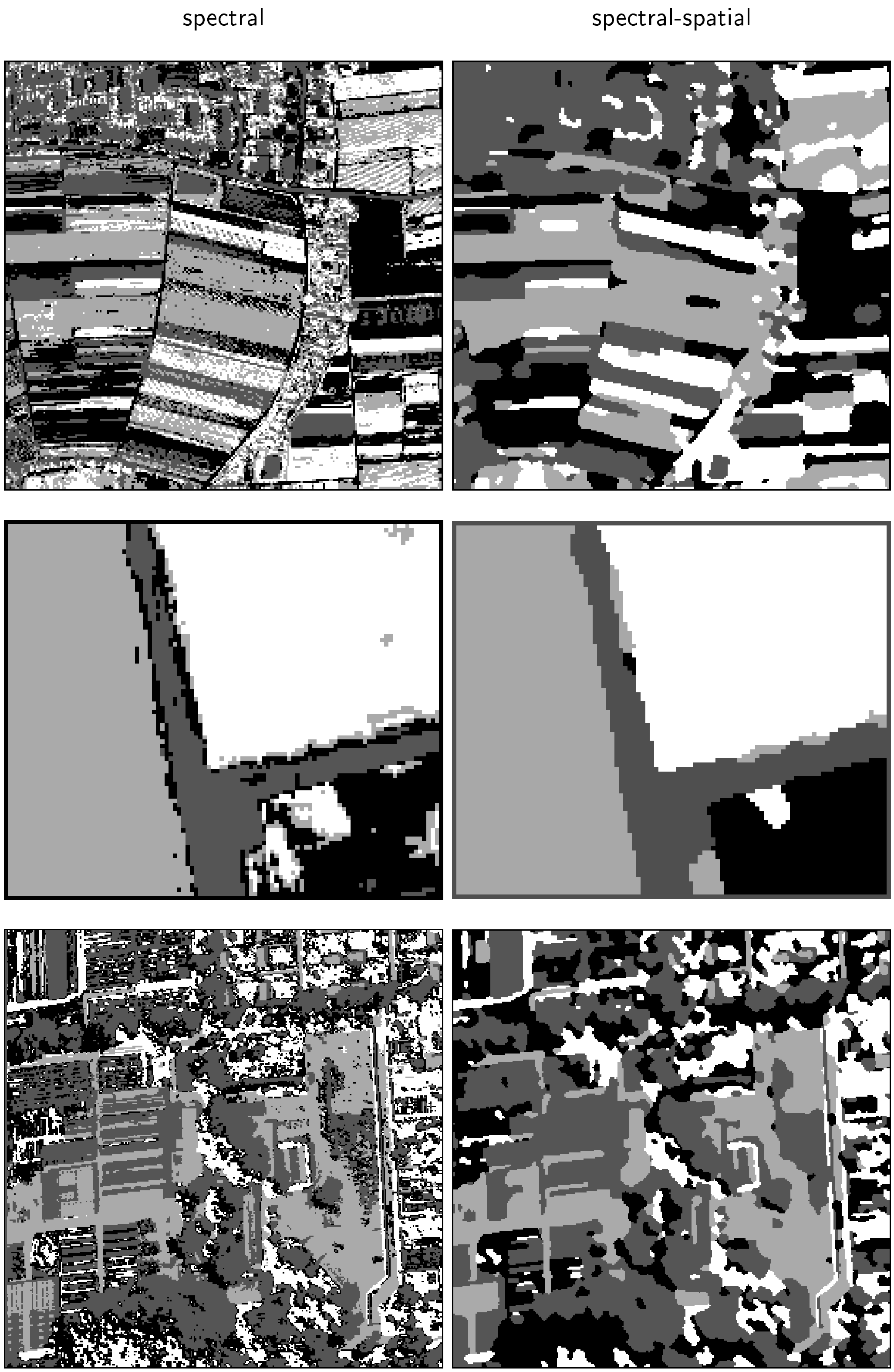

Figure 5b: Multispectral aerial imagery of Nürnberg with $n = 10$ spectral bands ($\lambda_1 = 435\,\mathrm{nm}$, $\lambda_{10} = 2215\,\mathrm{nm}$), classified by fuzzy k-means clustering into $k = 4$ classes (recorded in 1995, 300 m altitude, atmospherically corrected).