Unsupervised Fuzzy Clustering of Multi-variate Image Data

Klaus Baggesen Hilger and Allan Aasbjerg Nielsen
IMM, Department of Mathematical Modelling
Technical University of Denmark, 2800 Lyngby, Denmark
e-mail: kbh@imm.dtu.dk, aa@imm.dtu.dk

1. Introduction

An often occurring task in image analysis is the segmentation of multispectral (or multi-temporal) image data into a number of clusters/classes. Given an image with p spectral bands, the job is to assign to each observation or pixel a degree of membership. This can be done based on spectral characteristics alone, on spatial characteristics alone, or on combined spectral-spatial characteristics.

By applying the fuzzy c-means (FCM) algorithm, [1], we are able to segment an image into meaningful regions. For a given number of classes, the algorithm estimates the cluster centres in the p-dimensional feature space, including the degree of membership of each pixel. The membership corresponds to the a posteriori estimates of the class of the observation. Thus the image can be segmented by assigning each pixel to the class with the maximum a posteriori (MAP) estimate.

The resulting cluster centres can be emperically classified according to the most significant related sources. This allows for an unmixing of signals related to different sources. Here the influence of cloud signals, represented by the corresponding cluster centres, is reduced by means of orthogonal subspace projection (OSP), [3].

2. The FCM algorithm

The spectral fuzzy c-means algorithm

1. assigns values to p-dimensional feature vectors for C cluster centres, $\boldsymbol{r}_c,\ c = 1, \ldots, C$;
2. calculates membership weight for cluster $c = 1, \ldots, C$;
$$u_c = \frac{1/d_c^{2/(m-1)}}{\sum_{i=1}^{C} 1/d_i^{2/(m-1)}}$$
where d_c is the (Euclidean) spectral distance from the running observation $\boldsymbol{r}$ to each cluster centre $d_c^2 = (\boldsymbol{r} - \boldsymbol{r}_c)^T(\boldsymbol{r} - \boldsymbol{r}_c)$, and $m > 1$ is a user defined weight to control the degree of fuzziness which increases with m (default value $m = 2$);
3. calculates new cluster centres from
$$\boldsymbol{r}_c = \frac{\sum_{i=1}^{N} u_c^m \cdot \boldsymbol{r}}{\sum_{i=1}^{N} u_c^m}$$

where N is the number of observations (both u_c and $\boldsymbol{r}$ depend on i). Steps 2 and 3 are iterated until the largest change in cluster membership becomes small or zero.

To boost performance, the FCM algorithm can be embedded into a multi-resolution inheriting hierarchy. In [5] a spatial element is added. [2] adds a multi-resolution aspect.

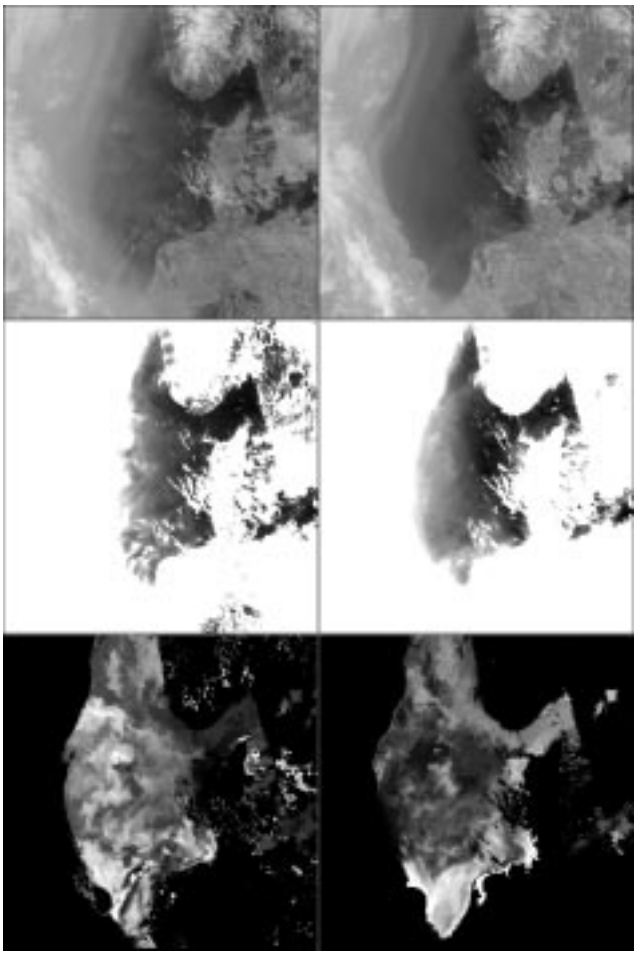

Figure 1. SeaWiFS bands 3 and 6. Before OSP cloud signal reduction: (row 1) stretched under the whole image, (row 2) stretched under the water mask obtained by FCM. After OSP cloud signal reduction: (row3) stretched under the water mask.

3. SeaWiFS example

SeaWiFS is an 8 channel optical scanner on the SeaStar spacecraft which orbits sun synchronously at a 705 km altitude. On a daily basis, SeaWiFS provides 10 bit data in the 402-422, 433-453, 480-500, 500-520, 545-565, 660-

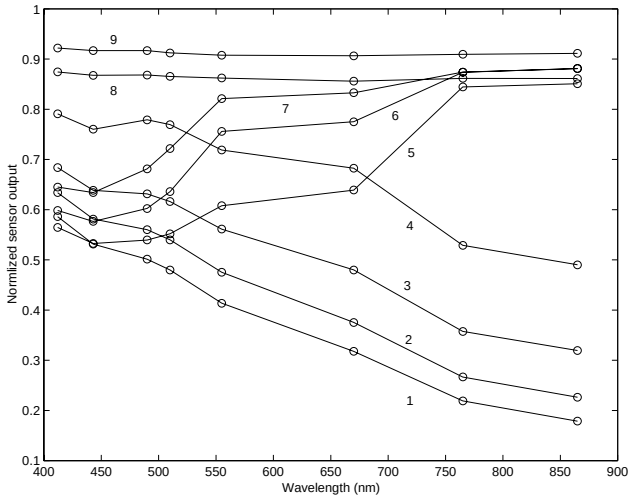

Figure 2. Cluster centres 1 through 9.

680, 745-785 and 845-885 nm regions. The pixel size is 1.1 km $\times$ 1.1 km. See also [4].

Figure 1 shows channels 3 and 6 of a SeaWiFS scene acquired on 14 May 1998. The figure also shows the enhanced ocean related signal after OSP cloud signal reduction. In Figure 2 the cluster centres estimated by FCM are presented. The OSP is performed on the cluster centres 2-4. The memberships of each pixel to the clusters are illustrated in Figure 3, along with the MAP estimated classes.

Acknowledgements

The work reported here is carried out under the GEOSONAR (http://manicoral.kms.min.dk/PK/ geosonar.-html) project funded by the Danish National Research Councils (http://www.forskraad.dk) under the Earth Observation Programme.

Use of the SeaWiFS data is in accord with the SeaWiFS Research Data Use Terms and Conditions Agreement.

References

[1] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, 1981.

[2] C. A. Bouman and M. Shapiro. A multiscale random field model for Bayesian image segmentation. *IEEE Transactions on Image Processing*, 3(2):162–177, 1994.

[3] A. A. Nielsen. Linear mixture models, full and partial unmixing in multi- and hyperspectral image data. In B. E. Ersbøll and P. Johansen, editors, *Proceedings of the Scandinavian Image Analysis Conference (SCIA'99), vol. 2*, pages 898–902, Kangerlussuaq, Greenland, 7-11 June 1999.

[4] SeaWiFS. http://seawifs.gsfc.nasa.gov/SEAWIFS.html. Goddard Space Flight Center, National Aeronautics and Space Administration, Greenbelt, Maryland, USA.

[5] R. Wiemker. Unsupervised fuzzy classification of multispectral imagery using spatial-spectral features. In I. Balderjahn, R. Mathar, and M. Schader, editors, *Data Highways and Information Flooding, A Challenge for Classification and Data Analysis*. Springer, 1997.

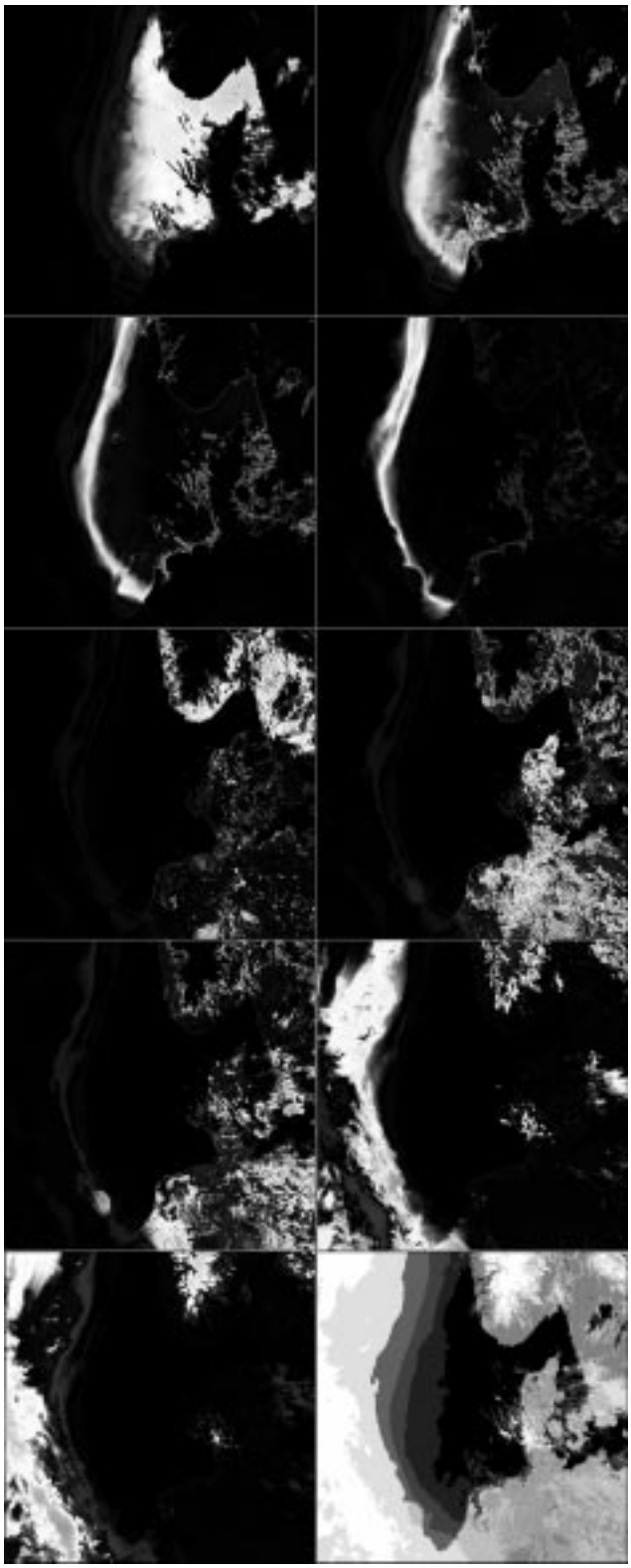

Figure 3. SeaWiFS, fuzzy, spectral segmentation, cluster memberships 1-9 row-wise. Most significant sources emperically related to each cluster: 1 water, 2-4 clouds, 5-7 vegetation, and 8-9 clouds and ice. Botton right frame is the MAP estimated classes. The black region, corresponding to cluster 1, is the water mask used in Figure 1 for stretching. The brightest region corresponds to cluster 9.