

Unsupervised Classification of Changes in Multispectral Satellite Imagery

Morton J. Canty

Forschungszentrum Jülich, D-52425 Jülich, Germany

Allan A. Nielsen

Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

ABSTRACT

The statistical techniques of multivariate alteration detection, maximum autocorrelation factor transformation, expectation maximization, fuzzy maximum likelihood estimation and probabilistic label relaxation are combined in a unified scheme to classify changes in multispectral satellite data. An example involving bitemporal LANDSAT TM imagery is given.

Keywords: Change detection, Clustering

1. INTRODUCTION

In this paper we describe a simple analysis sequence, based on standard algorithms and essentially free of adjustable parameters, for the unsupervised classification of changes in bitemporal, multispectral satellite imagery. The procedure is founded on the *multivariate alteration detection* or MAD transformation for change enhancement proposed originally by Nielsen et al.¹ which, because of its favorable statistical properties, provides a suitable starting point. The MAD transformation generates a set of mutually orthogonal difference images (MAD variates) having the same spectral dimension as the original multispectral images to which it is applied. It is explained in Section 2, together with a further processing step, the *scaled maximum autocorrelation factor* (SMAF) transformation,² which serves to reduce dimensionality and to enhance signal to noise. In Section 3 a mixture-modelling procedure based on a suggestion by Bruzzone and Prieto³ is applied to the SMAF/MAD variates individually in order to obtain better estimates of the variances of the no-change pixels. Section 4 deals with the application of the *fuzzy maximum likelihood estimation* (FMLE) clustering algorithm⁴ to the change pixels and the subsequent use of probabilistic label relaxation⁵ to include spatial information in the final change classification. In Section 5 examples of the techniques applied to the classification of changes in a LANDSAT TM bitemporal image are presented. Some conclusions are given in Section 6.

2. STATISTICAL CHANGE DETECTION

Briefly, the MAD transformation is as follows: We first represent two N -dimensional multispectral images of a scene acquired at times t_1 and t_2 by random vectors \mathbf{F} resp. \mathbf{G} , assumed to be multivariate normally distributed with zero mean. After forming the scalar difference

$$D = \mathbf{a}^\top \mathbf{F} - \mathbf{b}^\top \mathbf{G}, \quad (1)$$

the vectors \mathbf{a} and \mathbf{b} are chosen, analogously to the principal components transformation, so as to maximize the variance of D with the additional constraints imposed that $\text{var}(\mathbf{a}^\top \mathbf{F}) = \text{var}(\mathbf{b}^\top \mathbf{G}) = 1$. This means that the resulting difference image D will show maximum spread in its pixel intensities. If we assume that the spread is primarily due to actual changes that have taken place in the scene over the interval $t_2 - t_1$, then this procedure will enhance those changes as much as possible. The determination of the \mathbf{a} and \mathbf{b} that satisfy (1) and (2) is equivalent to two coupled, generalized eigenvalue problems. Their solution determines N eigenvalues ρ_i^2 and N pairs of eigenvectors $\mathbf{a}_i, \mathbf{b}_i$, $i = 1 \dots N$ and, accordingly, N difference components of the form (1), referred to as the *MAD variates*. The quantities $\mathbf{a}_i^\top \mathbf{F}$ and $\mathbf{b}_i^\top \mathbf{G}$ are called *canonical variates*. The square root of the eigenvalues ρ_i are the correlations between them. The MAD variates themselves are orthogonal (uncorrelated):

$$\text{cov}(D_i, D_j) = 2\delta_{ij}(1 - \rho_i) \quad (2)$$

and invariant under affine transformations. This invariance can be exploited to determine time-invariant features for relative radiometric normalization.⁶ If the eigenvectors $\mathbf{a}_i, \mathbf{b}_i, i = 1 \dots N$, are sorted according to decreasing eigenvalue ρ_i , the MAD variates will be sorted according to increasing variance. The sum of squares of standardized variates

$$\chi_N^2 = \left(\frac{D_1}{\sigma_{D_1}}\right)^2 + \dots + \left(\frac{D_N}{\sigma_{D_N}}\right)^2, \quad (3)$$

is approximately chi-square distributed with N degrees of freedom.

The MAD transformation can be augmented by subsequent application of the *maximum autocorrelation factor* (MAF) transformation, in order to improve the spatial coherence of the difference components, see Nielsen et al.¹ When image noise is estimated as the difference between intensities of neighboring pixels, the MAF transformation is equivalent to a *minimum noise fraction* (MNF) transformation,⁷ which generates image components with maximum signal to noise ratio. A further refinement normalizes each MAF variate such that the covariance matrix of its noise component is the identity matrix. This results in *scaled* MAF (SMAF) variates whose variances are given by $\text{SNR}_i + 1$, where SNR_i is the signal to noise ratio of the i th variate.² The SMAF/MAD variates thus generated are also orthogonal and invariant under affine transformations. In the present investigation, we have used the SMAF transformation to give more weight to the signal-rich change components by eliminating variates with signal to noise ratios smaller than 1.

3. DECISION THRESHOLDS

Decision thresholds for change pixels can be set in terms of standard deviations about the mean for each SMAF/MAD component separately. This may be done *ad hoc*, for example by saying that all pixels in an image component D whose intensities are in the interval $[-2\sigma_D, 2\sigma_D]$ are no-change pixels. A better technique^{3,8} involves application of a probability mixture model to the variates of the difference images, running a simple version of the EM algorithm to determine automatically the density functions for the change and no-change pixels and thence the optimal (error-minimizing) decision thresholds for discrimination of change from no-change pixels. Figure 1 shows a fit of a three-Gaussian mixture model to a typical SMAF/MAD variate. The decision thresholds are chosen at the left and right intersections of the three distribution functions. Since it is the no-change pixels which we expect to obey the chi-square distribution (3) we can replace the denominators σ_{D_i} with σ_{NC_i} , where $\sigma_{NC_i}^2$ is the variance of the no-change component for the i th SMAF/MAD variate as determined in the mixture model.

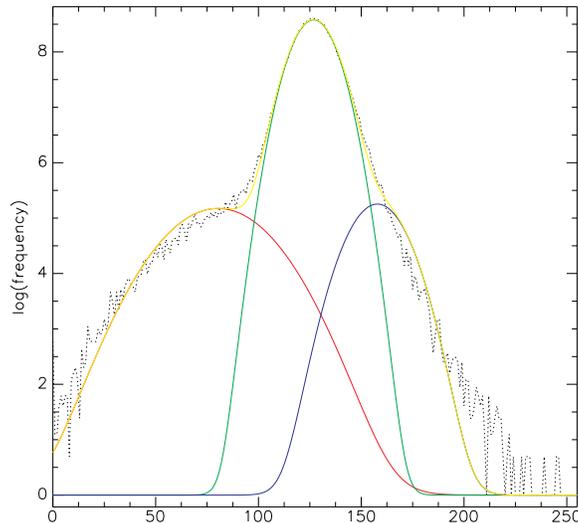


Figure 1. Mixture model fit to the histogram of a SMAF/MAD variate.

4. UNSUPERVISED CLASSIFICATION

The approximate chi-square distribution of the sum of squares of the standardized variates (3) allows the labelling of pixels with high change probability. These can then be clustered in N dimensions using any convenient clustering algorithm. We have chosen the method of Gath and Geva,⁴ sometimes referred to as *fuzzy maximum likelihood estimation* (FMLE). Unlike many other clustering algorithms, FMLE can accommodate elongated clusters and clusters of widely varying memberships, both of which are encountered for change pixels in the SMAF/MAD feature space. Moreover the “fuzzy” cluster memberships calculated in FMLE are in fact posterior probabilities which can be exploited for post classification processing.

We begin by defining the change pixels as those having a chi-square (3) exceeding a threshold $t = \chi_{N,P}^2$, where P is the probability of observing a value lower than t (typically we choose $P = 0.995$). Denote the corresponding SMAF/MAD vectors by \mathbf{d}_i , $i = 1 \dots n$, where n is the number of change pixels. Let u_{ki} be the membership probability of the i th change pixel to the k th cluster. If there are K clusters in all, we have

$$\sum_{k=1}^K u_{ki} = 1, \quad i = 1 \dots n, \quad (4)$$

and the probability of cluster k is given by

$$P(k) = \frac{1}{n} \sum_{i=1}^n u_{ki}, \quad k = 1 \dots K. \quad (5)$$

The cluster means or prototypes \mathbf{m}_k and “fuzzy” covariance matrices \mathbf{F}_k can then be written in the form

$$\begin{aligned} \mathbf{m}_k &= \frac{1}{nP(k)} \sum_{i=1}^n u_{ki} \mathbf{d}_i \\ \mathbf{F}_k &= \frac{1}{nP(k)} \sum_{i=1}^n u_{ki} (\mathbf{d}_i - \mathbf{m}_k)(\mathbf{d}_i - \mathbf{m}_k)^\top, \quad k = 1 \dots K. \end{aligned} \quad (6)$$

These moments in turn determine the membership probabilities according to $u_{ki} = P(k | \mathbf{d}_i)$, the posterior probability for cluster k conditional upon observation \mathbf{d}_i . That is, invoking Bayes’ Theorem and assuming each cluster to have a multivariate normal density distribution, we have

$$\begin{aligned} u_{ki} &= P(k | \mathbf{d}_i) = cP(k)P(\mathbf{d}_i | k) \\ &= cP(k) \frac{1}{\sqrt{|\mathbf{F}_k|}} \exp\left(-\frac{1}{2}(\mathbf{d}_i - \mathbf{m}_k)^\top \mathbf{F}_k^{-1}(\mathbf{d}_i - \mathbf{m}_k)\right), \end{aligned} \quad (7)$$

where c is a normalization constant independent of k which can be determined from (4). Starting from some initial choice of the memberships u_{ik} , the FMLE algorithm consists of a simple iteration of equations (4-7) until convergence.

Because of the exponential distance dependence of the fuzzy cluster membership in (7), the algorithm is very sensitive to initialization conditions, and can even become unstable. To avoid this problem we follow the suggestion of Gath and Geva⁴ and first obtain initial values for the u_{ki} by preceding the calculation with the fuzzy K-means algorithm,⁹ for which the class memberships can be chosen to follow an inverse square law.

In order to discriminate among different choices for the number K of change clusters we again follow Gath and Geva⁴ and first define a “fuzzy hypervolume” given by

$$F_{HV} = \sum_{k=1}^K \sqrt{|\mathbf{F}_k|}.$$

The *partition density* P_D is then defined according to

$$P_D = S/F_{HV}, \quad (8)$$

where S is the number of change pixels within one standard deviation of a cluster prototype. An optimal choice for K is one which maximizes the partition density.

After running the algorithm, the no-change class is reintroduced as $k = 0$ with mean and covariance matrix $\mathbf{m}_0 = 0$ and \mathbf{F}_0 , respectively, where \mathbf{F}_0 is estimated from the no-change pixels. Thus with (7)

$$u_{0i} = cP(0) \frac{1}{\sqrt{|\mathbf{F}_0|}} \exp\left(-\frac{1}{2} \mathbf{d}_i^\top \mathbf{F}_0^{-1} \mathbf{d}_i\right)$$

The priors $P(k)$, $k = 0 \dots K$, are renormalized appropriately. Finally, all pixels are labelled according to the maximum membership probability:

$$\text{label}(i) = \arg \max_{k \in \{0 \dots K\}} u_{ki}. \quad (9)$$

With Jia et al.⁵ we note that a possible misclassification of a pixel in (9) could in principle be corrected by examining the membership probabilities of pixels in its neighborhood. They describe a method referred to as *probabilistic label relaxation*, which we have adapted here to improve our classification result and take spatial information into account. The class membership vectors $\mathbf{u}_i = (u_{1i} \dots u_{Ki})^\top$ are updated according to

$$\mathbf{u}'_i = \mathbf{u}_i \otimes \frac{\mathbf{Q} \mathbf{u}_n}{\mathbf{u}_i^\top \mathbf{Q} \mathbf{u}_n}. \quad (10)$$

where $\mathbf{Q} = (Q)_{kl}$ is a $K \times K$ matrix of compatibility measures expressing the probability that a pixel in class k has a neighbor in class ℓ , \mathbf{u}_n is the average class membership vector for a 4-neighborhood of pixel i and \otimes denotes Hadamard (component-by-component) multiplication. \mathbf{Q} is easily estimated directly from the originally classified image.

The probabilistic label relaxation procedure can be iterated arbitrarily often. However too many iterations may lead to a widening of the effective neighborhood of a pixel to such an extent that irrelevant spatial information may falsify the final classification. Experience shows that the best results are obtained after 3–4 iterations.⁵

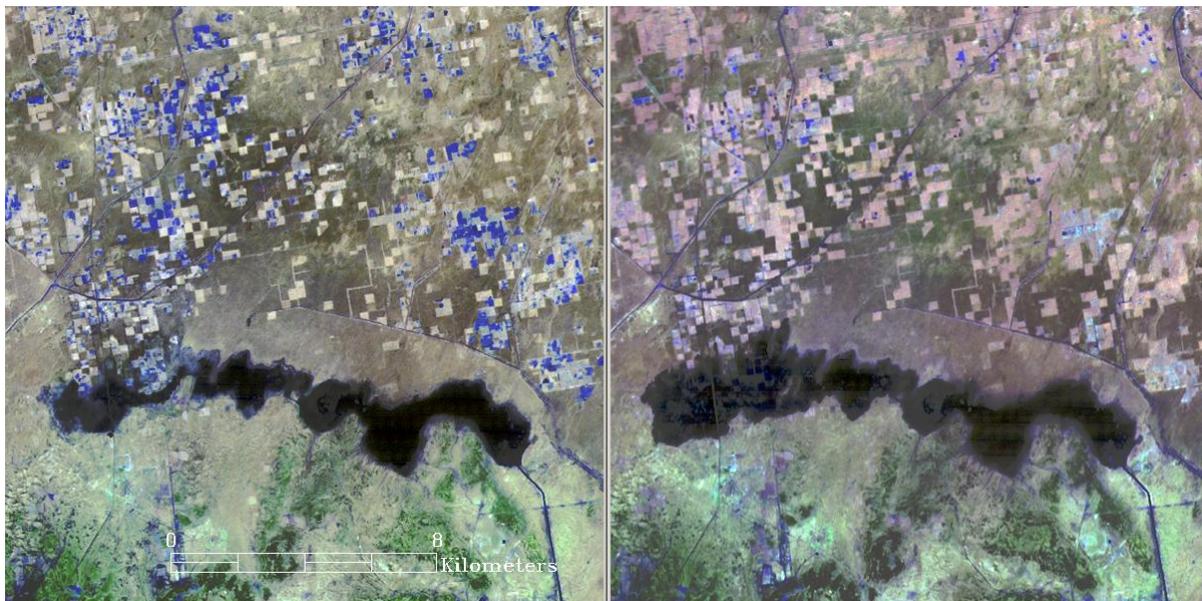


Figure 2. LANDSAT TM5 images (RGB composite of bands 7, 5, 4, Gaussian stretch) acquired on March 29,1998 (left) and May 16, 1998 (right).

5. RESULTS

The procedures described in the foregoing sections are illustrated on the basis of a bitemporal multispectral image consisting of two LANDSAT TM5 scenes acquired on March 29 and then again on May 16, 1998 over a semi-arid agricultural area in Hindustan, India (Figure 2). The scenes were registered to one another by applying an automatic contour matching algorithm due to Li et al.¹⁰ and using first-order polynomial, nearest-neighbor re-sampling. The RMS errors were less than or equal to 0.5 pixel. Because of the good atmospheric conditions and the invariance of the MAD variates under linear transformations, no radiometric corrections were applied to the two scenes. The most evident change that took place between the acquisitions is some shallow flooding at the eastern end of the lake or reservoir.

After performing the MAD transformation as described in Section 2, the SMAF transformation was carried out on the MAD variates. The resulting SMAF/MAD variates were ordered according to increasing signal to noise ratio. The values obtained are shown in Table 1. The first three components are seen to have a signal to noise ratio smaller than 1. These were eliminated from subsequent analysis.

Table 1. Signal to noise ratios after SMAF transformation.

SMAF	SNR
1	0.40
2	0.56
3	0.61
4	2.30
5	5.51
6	14.83

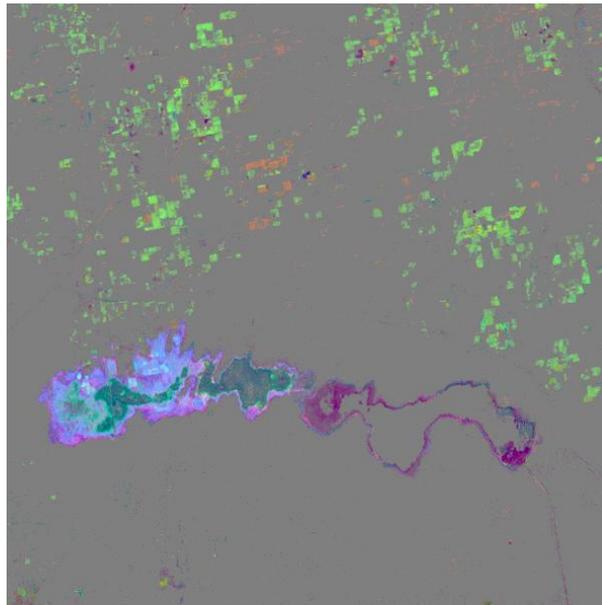


Figure 3. Change image with thresholds on SMAF components 4, 5 and 6 displayed in blue, green and red, respectively.

In Figure 3 we see the result of thresholding the difference components with the automatically determined values as described in Section 3. An SMAF/MAD component lying within the lower and upper no-change thresholds is colored middle gray (byte intensity 127). This serves to make negative changes more visible. Moreover the pixel intensities are stretched across ± 8 standard deviations in the respective component so that

changes less than or exceeding this amount are saturated in their respective colors at 0 or 255. In addition to the changes associated with the lake level and water quality, substantial changes in the agricultural fields to the north of the lake are evident. These changes are predominantly in green, that is in SMAF/MAD variate 5.

Next we consider clustering the no-change pixels with the procedure of Section 4. The first question is the appropriate number of clusters to use. Figure 4 displays the partition density, Eq. (8), as a function of chosen number of clusters ranging from 2 to 12, determined after convergence of the FMLE algorithm. A fairly convincing maximum is evident for 5 clusters and this number was accordingly chosen.

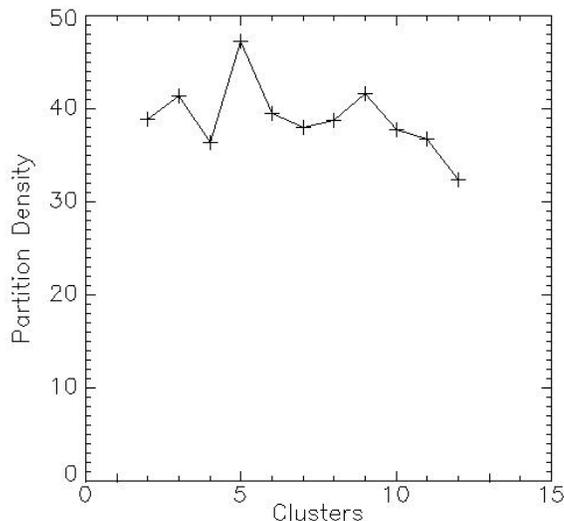


Figure 4. Plot of partition density vs. cluster number.

Figure 5 shows the clustered change pixels after convergence using 5 cluster centers, projected onto the 5–6 SMAF/MAD plane. Note the absence of the dominant no-change cluster.

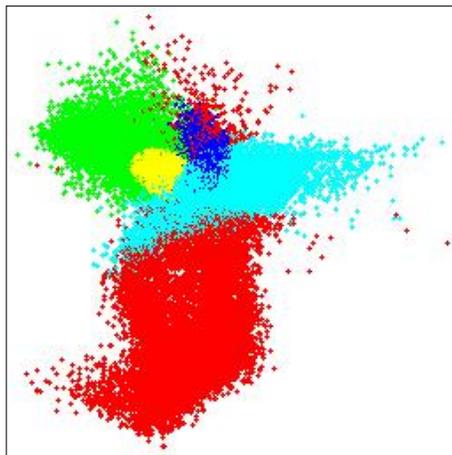


Figure 5. Projection of 5 clusters onto the 5–6 (X–Y) SMAF/MAD plane.

Figure 6 illustrates the effect of post-classification processing with probabilistic label relaxation (Section 4). The spatial coherence of the change classes improves remarkably.

The final product, that is, the unsupervised change classification, is shown in Figure 7. By complementing

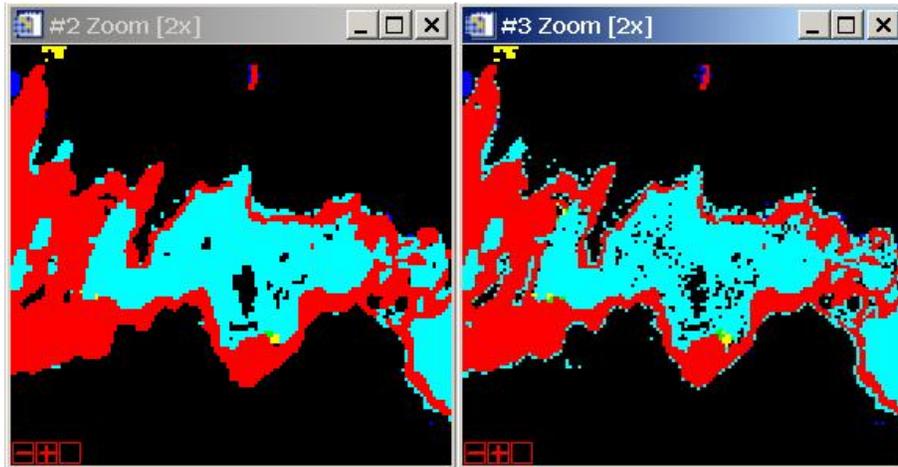


Figure 6. Portion of clustered image before (right) and after four iterations of probabilistic label relaxation.

this representation with ground reference data, the five classes may be assigned to real change categories, thereby giving a quantitative analysis of ground cover changes that have taken place.

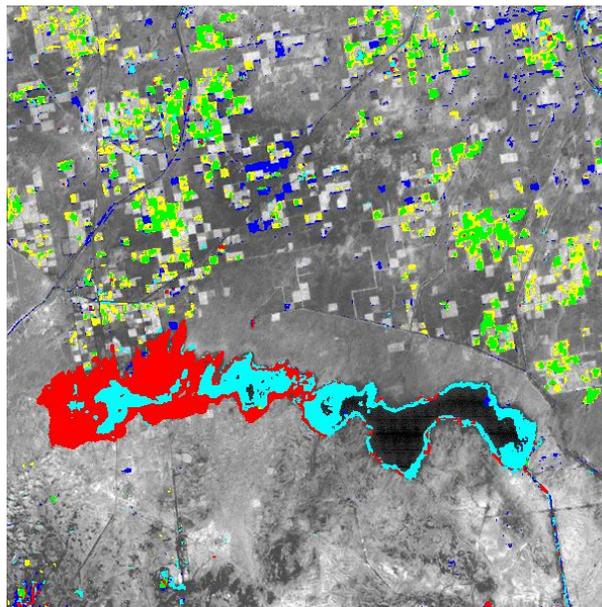


Figure 7. Final classification of changes with 5 clusters.

6. CONCLUSION

We have introduced and demonstrated a generally applicable and straightforward method for classification of changes in multispectral satellite imagery, based on well-established algorithms in image analysis and pattern recognition and building upon the robustness of the MAD and SMAF transformations for extracting change information from bitemporal multispectral data and for enhancing signal to noise.

The threshold determination via mixture modelling is closely related to change vector analysis,¹¹ a technique most commonly used with untransformed spectral bands or with the tasseled cap transformation. In those cases a direct physical interpretation of the change signals is sometimes possible. Invariably, however, the change – no

change decision thresholds are treated as adjustable parameters. The automatic threshold determination procedures could in principle be applied here as well, although we are not aware of anyone having done so. We have preferred to use the SMAF/MAD transformed variables because of their robustness, their statistical orthogonality, their approximate normal distribution and their ability to enhance signal-to-noise. Physical interpretation of change can still be accomplished by examining correlations with the spectral bands in the original scenes.¹

The unsupervised FMLE classification of the change pixels, identified on the basis of the statistics of the SMAF/MAD variates, allowed for elongated (hyper-ellipsoidal) clusters and clusters of differing sizes, and included a criterion for choosing the best number of classes. Apart from the choice of the chi-square percentile for selection of change pixels, no other adjustable parameters are involved. Advantage can be taken of the probabilistic nature of class memberships established by the FMLE algorithm to apply post classification processing to improve the spatial coherence of the change classes obtained. Quantitative comparisons of unsupervised classification methods are notoriously difficult, see for example Duda and Canty.¹² In any case we are unaware of other work which could be critically compared with the clustering approach described here.

REFERENCES

1. A. A. Nielsen, K. Conradsen, and J. J. Simpson, "Multivariate alteration detection (MAD) and MAF postprocessing in multispectral, bitemporal image data: New approaches to change detection studies," *Remote Sens. Environ* **64**, pp. 1–19, 1998.
2. K. B. Hilger, A. A. Nielsen, and R. Larsen, "A scheme for initial exploratory data analysis of multivariate image data," *Proceedings of 12th Scandinavian Conference on Image Analysis Bergen, Norway*, pp. 717–724, 2001.
3. L. Bruzzone and D. F. Prieto, "Automatic analysis of the difference image for unsupervised change detection.," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**(4), pp. 1171–1182, 2000.
4. I. Gath and A. B. Geva, "Unsupervised optimal fuzzy clustering.," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **3**(3), pp. 773–781, 1989.
5. X. Jia, J. A. Richards, and D. E. Ricken, *Remote Sensing Digital Image Analysis*, Springer-Verlag, 1999.
6. M. J. Canty, A. A. Nielsen, and M. Schmidt, "Automatic radiometric normalization of multispectral imagery," *Remote Sens. Environ.* **91**, pp. 441–451, 2004.
7. A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Transactions on Geoscience and Remote Sensing* **26**(1), pp. 65–74, 1988.
8. L. Bruzzone and D. F. Prieto, "An adaptive semi-parametric and context-based approach to unsupervised change detection in multitemporal remote sensing images. Technical Report No. DIT-02-0030, Department of Information and Communication Technology, University of Trento," 2002.
9. J. C. Dunn, "A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters.," *Journal of Cybernetics* **PAM1-1**, pp. 32–57, 1973.
10. H. Li, B. S. Manjunath, and S. K. Mitra, "A contour-based approach to multisensor image registration," *IEEE Transactions on Image Processing* **4**(3), pp. 320–334, 1995.
11. A. Singh, "Digital change detection techniques using remotely-sensed data," *Int. J. Remote Sensing* **10**(6), pp. 989–1002, 1989.
12. T. Duda and M. J. Canty, "Unsupervised classification of satellite imagery: choosing a good algorithm," *Int. J. Remote Sensing* **23**(11), pp. 2193–2212, 2002.