USING MIXED PIXELS FOR THE TRAINING OF A MAXIMUM LIKELIHOOD CLASSIFICATION

J. Lesparre[a, *], B. G. H. Gorte[a]

[a] Dept. of Aerospace Engineering, Delft University of Technology, Kluyverweg 1, 2629 HS Delft, the Netherlands -
jochem.lesparre@kadaster.nl, b.g.h.gorte@lr.tudelft.nl

KEYWORDS: Training, Fuzzy Logic, Method, Estimation, Classification, Sampling, Vegetation

ABSTRACT:

Most supervised image classification methods need pure pixels for training, which complicates training when pure pixels are scarce. In these cases, it can be difficult to obtain a sufficiently large number of representative training samples to accurately estimate the spectra of the classes. The direct use of 'almost pure' pixels is not advisable. These cause the estimated mean to be biased and the estimated variance to depend on the degree of occurrence of other classes, instead of on the natural variation in the spectrum of the class. The solution for the lack of pure training samples is to be found in the use of mixed pixels to estimate the spectra of pure classes.

This article presents a method to estimate unbiased pure spectra out of mixed pixels using adjustment theory and probability model estimation. An advantage of such a fuzzy training method is that more pixels in the image can be used for training, which enables the use of heterogeneous areas for training or the random selection of training pixels. There are two conditions for this method. First, one needs to have estimates of the fractions of the classes in the mixed training samples. Secondly, the spectral values of the mixed pixels should be a linear combination of the spectra of the composing classes.

1. INTRODUCTION

Training is the first step of a supervised image classification. The objective of the training is specifying the classes that need to be distinguished and providing the features of these classes. Usually, the class spectra cannot be taken from a library. For example, atmospheric conditions and the growing stadium of vegetation at the moment of image recording have a significant influence on the spectra. Therefore, representative pixels of each class need to be selected from the image itself as training samples, to estimate the spectra of the classes.

Supervised image classification can be crisp or fuzzy. A crisp method classifies every pixel, mixed or not, always in one of the classes. When spatial variability is high, compared to the spatial resolution of the imagery, as it is often the case when monitoring natural vegetation, it is better not to be so strict and to allow pixels to be classified in more than one class at the same time. In this article a classification is called fuzzy if it classifies objects not just in one class, but to more or less degree in multiple classes, expressed in a value per class. This value can be a probability, a fraction or any other quantity.

1.1 Motivation

This article is a result of research in the use of probabilistic segmentation and fuzzy classification for imagery of natural vegetation (Lesparre, 2003; Gorte *et al.*, 2003). Therefore, some of the motivation is based on aspects of natural vegetation or fuzzy classification. Nevertheless, the presented method is applicable for other purposes too.

Most supervised image classification methods need pure pixels for training, which complicates training in cases that pure pixels are scarce, for example images of natural vegetation. In these cases, it can be difficult to obtain a sufficiently large number of representative training samples to accurately estimate the spectra of the classes. One of the possibilities to get around this problem is using 'almost pure' pixels for training. There are even methods to find the 'purest' pixels in an image, by seeking the extremes in the feature space, like the Pixel Purity Index (ENVI, 1999). However, the use of 'almost pure' pixels is not advisable. These cause the estimated mean to be biased and the estimated variance to be influenced by the degree of occurrence of other classes, instead of to represent the natural variation in the spectrum of the class. For crisp classifications, one could argue that this is not so much of a problem, because for most pixels the most likely class will still be the right one. However, for a fuzzy classification one is seeking probabilities (or another fuzzy value) for each class. These will have a systematic error as a result of the not completely pure training pixels.

Example: Suppose we wish to distinguish the classes wood and heath and the spectra of these classes both have the same variation and no correlation. For the training of the class heath we have 20 pure pixels at our disposal, but for the class wood we use 10 pure pixels and 10 pixels with 10% heath in it (figure 1). In this case the estimated variance (the ellipse in figure 1) of the class wood will be larger than the real variance (the circle). As a result of this, wood will be overrepresented in the classification.

The solution for the lack of pure training samples is not to be found in the use of 'almost pure' pixels as pure ones, but in using mixed pixels to estimate the spectra of pure classes. This can be considered as fuzzy training. In the ideal case, training, classification and validation are all fuzzy (Foody, 1999).

* Corresponding author, currently employed at Geodesy Dept., Kadaster, Hofstraat 110, 7311 KZ Apeldoorn, the Netherlands.

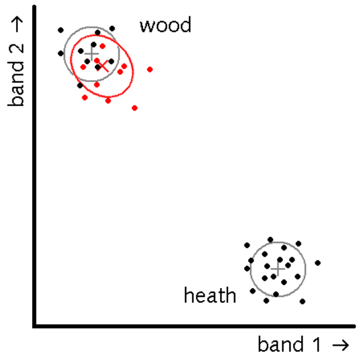

Figure 1. Feature space with pure training samples of the class heath and only partly pure training samples for the class wood.

The computed features in the training stadium of a maximum likelihood classification are for each class the mean value and the variance per spectral band and the correlation with other bands. A Gaussian distribution is assumed for this. How can these features be estimated using mixed pixels? Wang (1990) suggested computing a weighted mean and a weighted empirical (co)variance, using the fractions of the class as weight. However, this method gives biased estimates of the spectra of the pure classes (Eastman and Laney, 2002). Namely, the non-pure pixels pull the weighted mean, despite the lower weight, in the direction of the non-pure spectra. Another method to use mixed pixels as training samples is to train for each mixture proportion separately. However, this would require training samples for all the mixture proportions that need to be distinguished. For natural vegetation applications, this would require enormous amount of fieldwork to acquire the needed training samples.

1.2 Aim

This article presents a method to execute a maximum likelihood training, which estimates an unbiased mean and (co)variances of pure spectra using mixed pixels. The principle is explained in section 2.1. This is worked out in the sections 2.2-2.4, using adjustment theory and probability model estimation.

2. METHOD

2.1 Principle

To be able to use mixed pixels, a model is needed to relate the spectrum of a mixed pixel to the spectra of its comprising components. When linear mixing of the spectra is assumed, the value of a mixed pixel equals:

$$\chi_{B_j S_i} = \sum_{k=1}^{K} (f_{C_k S_i} \cdot \chi_{B_j C_k}) \qquad (1)$$

Where $\chi_{B_j S_i}$ = value for spectral band j of sample i
$f_{C_k S_i}$ = fraction of class k in sample i
$\chi_{B_j C_k}$ = mean value for spectral band j of class k
K = the number of classes

Note: The value $\chi_{B_j S_i}$ is element j of the feature vector of sample i. A feature vector is a vector containing the spectral signature of a pixel. This is indicated with χ (chi), the first character of the Greek word *chroma*, meaning colour.

The sum of all fractions should equal 1, so:

$$\sum_{k=1}^{K} f_{C_k S_i} = 1 \qquad (2)$$

Points in the feature space with different mixture proportions of two classes are situated on a straight line in between the two pure spectra. The position on this line is linear proportional to the fractions of the classes. This enables us to estimate the spectra of the pure classes using mixed pixels, provided that the mixture proportions are known.

Example: Suppose we have two training pixels, one with 25% wood and 75% heath and one with 75% wood and 25% heath. If we draw a line in the feature space, from one to the other, and extend this line with half its length on both ends, we get estimates for samples of 100% wood and 100% heath (figure 2).

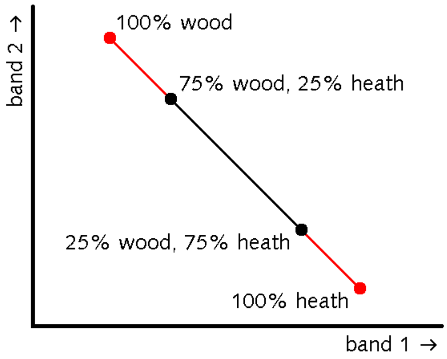

Figure 2. Feature space with estimates of two pure spectra using two mixed pixels.

In case of many training samples instead of two, these will never lie exactly on one line, resulting in an ambiguous solution. Using adjustment theory a least squares estimation of the pure spectra is obtained in section 2.2 (figure 3). For a maximum likelihood classification not only the mean values, but also the (co)variances of the pure spectra, are needed. To estimate the variances of the pure spectra, variance component model estimation is used in section 2.3.

2.2 Least squares estimation of mean values

To estimate the spectra of pure classes using mixed pixels least squares adjustment theory (Teunissen, 1999) can be used. To apply a least squares estimation, (1) and (2) need to be formulated as observation equations:

$$\mathrm{E}\{\chi_{B_j S_i}\} = \sum_{k=1}^{K-1} (f_{C_k S_i} \cdot \chi_{B_j C_k}) + (1 - \sum_{k=1}^{K-1} f_{C_k S_i}) \cdot \chi_{B_j C_K}$$

$$\begin{aligned} \mathrm{E}\{f_{C_1S_i}\} &= f_{C_1S_i} \\ &\vdots \\ \mathrm{E}\{f_{C_{K-1}S_i}\} &= f_{C_{K-1}S_i} \end{aligned} \tag{3}$$

The observation equations (3) are of the form $\mathrm{E}\{y\} = A(x)$, where the observations (y) are the spectra of the training pixels and the fractions of the classes in the training pixels. The unknowns (x) are the spectra of the pure classes and the fractions of the classes in the training pixels. As these equations are non-linear, they need to be linearised to be able to perform the least squares estimation. The linearised equations are:

$$\begin{aligned} \mathrm{E}\{\Delta\chi_{B_jS_i}\} &= \sum_{k=1}^{K-1}(f_{C_kS_i}{}^{o}\cdot\Delta\chi_{B_jC_k}) + (1-\sum_{k=1}^{K-1} f_{C_kS_i}{}^{o})\cdot \\ &\quad \Delta\chi_{B_jC_K} + \sum_{k=1}^{K-1}((\chi_{B_jC_k}{}^{o} - \chi_{B_jC_K}{}^{o})\cdot\Delta f_{C_kS_i}) \\ \mathrm{E}\{\Delta f_{C_1S_i}\} &= \Delta f_{C_1S_i} \\ &\vdots \\ \mathrm{E}\{\Delta f_{C_{K-1}S_i}\} &= \Delta f_{C_{K-1}S_i} \end{aligned} \tag{4}$$

Here the observations are noted as $\Delta y = y - y^{o}$ and the unknowns as $\Delta x = x - x^{o}$. The approximate values of the unknowns (x^{o}) can be obtained, for instance, from the purest training samples. The approximate values of the observations (y^{o}) are obtained from the equation $y^{o} = A(x^{o})$.

Next, the linearised equations (4) can be put in matrix notation $\mathrm{E}\{\Delta y\} = A\cdot\Delta x$. Then, the least squares estimates of the unknowns results from:

$$\hat{x} = x^{o} + (A^{\mathrm{T}}Q_y^{-1}A)^{-1}A^{\mathrm{T}}Q_y^{-1}\Delta y \tag{5}$$

Where $\hat{x}$ = least squares estimates of the unknowns (x)
Q_y = covariance matrix of the observations (y)

When the approximate values are not very accurate, it will be necessary to estimate the unknowns ($\hat{x}$) iteratively. By repeatedly using the result as a new approximate value, the estimates can be obtained as precise as desired.

Example: For two classes, two spectral bands and five training pixels, the linearised model of observation equations is:

$$\mathrm{E}\{\begin{bmatrix} \Delta\chi_{B_1S_1} \\ \Delta\chi_{B_2S_1} \\ \Delta f_{C_1S_1} \\ \vdots \\ \Delta\chi_{B_1S_5} \\ \Delta\chi_{B_2S_5} \\ \Delta f_{C_1S_5} \end{bmatrix}\} = [A_1 \mid A_2]\cdot\begin{bmatrix} \Delta\chi_{B_1C_1} \\ \Delta\chi_{B_2C_1} \\ \Delta\chi_{B_1C_2} \\ \Delta\chi_{B_2C_2} \\ \hline \Delta f_{C_1S_1} \\ \vdots \\ \Delta f_{C_1S_5} \end{bmatrix} \tag{6}$$

Where
$$A_1 = \begin{bmatrix} f_{C_1S_1}{}^{o} & 0 & 1-f_{C_1S_1}{}^{o} & 0 \\ 0 & f_{C_1S_1}{}^{o} & 0 & 1-f_{C_1S_1}{}^{o} \\ 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ f_{C_1S_5}{}^{o} & 0 & 1-f_{C_1S_5}{}^{o} & 0 \\ 0 & f_{C_1S_5}{}^{o} & 0 & 1-f_{C_1S_5}{}^{o} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} \chi_{B_1C_1}{}^{o} - \chi_{B_1C_2}{}^{o} & & 0 \\ \chi_{B_1C_1}{}^{o} - \chi_{B_1C_2}{}^{o} & & 0 \\ 1 & & 0 \\ & \ddots & \\ 0 & & \chi_{B_1C_1}{}^{o} - \chi_{B_1C_2}{}^{o} \\ 0 & & \chi_{B_1C_1}{}^{o} - \chi_{B_1C_2}{}^{o} \\ 0 & & 1 \end{bmatrix}$$

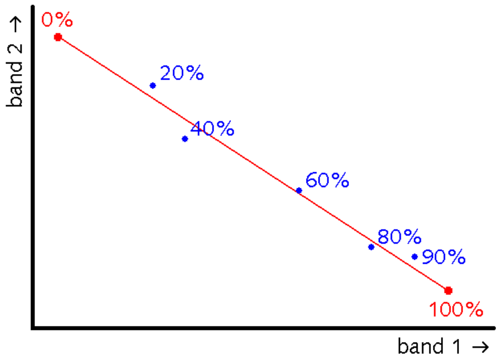

Figure 3. Feature space with least squares estimates of two pure spectra using five mixed pixels. The percentages are the fraction of one class. The fraction of the other class is 100% minus the given percentage.

The above formulas assume that the sum of the observed fractions of the classes in each training sample is 1. Therefore, the fraction of the last class (K) of each sample is not an independent observation, and consequently it is not present in the observation equations. If the fractions are observed in a way that does not guarantee a sum of all fractions equal to 1, the fraction of the last class (K) is an independent observation too, but still not an independent unknown. In this case, one observation equation for each training pixel should be added:

$$\mathrm{E}\{f_{C_KS_i}\} = 1 - \sum_{k=1}^{K-1} f_{C_kS_i} \tag{7}$$

The use of an unknown class is in theory possible. However, the fractions of this unknown class need to be estimated too, but the assumption of Gaussian distribution would probably not be justified for this class.

The quality of the estimated mean values and class fractions ($\hat{x}$) is given by the covariance matrix:

$$Q_{\hat{x}} = (A^{\mathrm{T}} Q_y^{-1} A)^{-1} \tag{8}$$

This covariance matrix should not be confused with the (co)variances estimated in the next section.

2.3 Variance component model estimation of (co)variances

To estimate the (co)variances of the spectra of pure classes using mixed pixels, the formula for the empirical (co)variance can no longer be used. In this case the more general formulas of the variance component model estimation have to be used, just like the formula for the mean needed to be replaced by the least squares estimation. Under certain circumstances the general formulas of the least squares estimation and the variance component model estimation lead to the simplified formulas of the mean and empirical (co)variance.

For variance component model estimation (Teunissen and Amiri-Simkooei, 2006) the covariance matrix to be estimated is composed as the weighted sum of P components, to estimate the factors of the components:

$$Q_y = \sum_{p=1}^{P} \sigma_p^2 Q_p \tag{9}$$

Where σ_p^2 = unknown (co)variance factor p
Q_p = cofactor matrix p

The estimates of the (co)variance factors (σ_p^2) can be computed by:

$$\hat{\sigma} = \begin{bmatrix} N_{11} & \cdots & N_{1P} \\ \vdots & \ddots & \vdots \\ N_{P1} & \cdots & N_{PP} \end{bmatrix}^{-1} \cdot \begin{bmatrix} l_1 \\ \vdots \\ l_P \end{bmatrix} \tag{10}$$

Where $\sigma = [\sigma_1^2 \;\cdots\; \sigma_P^2]^{\mathrm{T}}$
$N_{pq} = \mathrm{trace}(Q_y^{-1} P_A^{\perp} Q_p Q_y^{-1} P_A^{\perp} Q_q)$
$l_p = y^{\mathrm{T}} Q_y^{-1} P_A^{\perp} Q_p Q_y^{-1} P_A^{\perp} y$
$P_A^{\perp} = I - A \cdot (A^{\mathrm{T}} Q_y^{-1} A)^{-1} A^{\mathrm{T}} Q_y^{-1}$

Because the covariance matrix (Q_y) that is determined in (10) is also present in the formula itself, approximated values of the (co)variance factors are needed. The same covariance matrix as used in the least squares estimation can be used for this. An iterative computation should be performed to improve the approximated values until the desired precision is achieved.

Model design: The components of the model (the cofactor matrices Q_p) can be chosen in many ways. However, not every possibility is useful. To estimate the covariance matrix of the pure spectra, the covariance matrix of the training samples should be expressed as a function of the covariance matrices of the pure spectra. Based on the model of linear mixing of spectra (formula 3), this is:

$$Q_{\chi S_i} \approx \sum_{k=1}^{K-1} (f_{C_k S_i}{}^2 \cdot Q_{\chi C_k}) + (1 - \sum_{k=1}^{K-1} f_{C_k S_i})^2 \cdot Q_{\chi C_K} \tag{11}$$

The correlation between the classes is neglected, just as it is in conventional training for a maximum likelihood classification. To estimate all parameters of a normal training for a maximum likelihood classification, for every spectral band of every class a variance should be estimated, as well as covariances for all combinations of spectral bands for every class. To complete the variance component model, at least one variance should be estimated for the observation of the fractions of the classes. This results in P components:

$$P = KJ + \tfrac{1}{2} K (J^2 - J) + F \tag{12}$$

Where K = the number of classes
J = the number of spectral bands
F = the number of components for the observation of the fractions of the classes

Example: The variance component model in case of 2 classes and 2 spectral bands is:

$$Q_y = \sigma^2_{\chi B_1 C_1} \begin{bmatrix} \ddots & & & & \\ & f_{C_1 S_i}{}^2 & 0 & 0 & \\ & 0 & 0 & 0 & \\ & 0 & 0 & 0 & \\ & & & & \ddots \end{bmatrix} +$$

$$\sigma^2_{\chi B_1 C_2} \begin{bmatrix} \ddots & & & & \\ & (1 - f_{C_1 S_i})^2 & 0 & 0 & \\ & 0 & 0 & 0 & \\ & 0 & 0 & 0 & \\ & & & & \ddots \end{bmatrix} +$$

$$\sigma^2_{\chi B_2 C_1} \begin{bmatrix} \ddots & & & & \\ & 0 & 0 & 0 & \\ & 0 & f_{C_1 S_i}{}^2 & 0 & \\ & 0 & 0 & 0 & \\ & & & & \ddots \end{bmatrix} +$$

$$\sigma^2_{\chi B_2 C_2} \begin{bmatrix} \ddots & & & & \\ & 0 & 0 & 0 & \\ & 0 & (1 - f_{C_1 S_i})^2 & 0 & \\ & 0 & 0 & 0 & \\ & & & & \ddots \end{bmatrix} +$$

$$\sigma_{\chi B_1 C_1 \chi B_2 C_1} \begin{bmatrix} \ddots & & & & \\ & 0 & f_{C_1 S_i}{}^2 & 0 & \\ & f_{C_1 S_i}{}^2 & 0 & 0 & \\ & 0 & 0 & 0 & \\ & & & & \ddots \end{bmatrix} +$$

$$\sigma_{\chi_{B_1C_2}\chi_{B_2C_2}}\begin{bmatrix}\ddots & & & \\ & 0 & (1-f_{C_1S_i})^2 & 0 \\ & (1-f_{C_1S_i})^2 & 0 & 0 \\ & 0 & 0 & 0 \\ & & & & \ddots\end{bmatrix}+$$

$$\sigma^2_{f_{C_1}}\begin{bmatrix}\ddots & & & \\ & 0 & 0 & 0 \\ & 0 & 0 & 0 \\ & 0 & 0 & 1 \\ & & & & \ddots\end{bmatrix} \qquad (13)$$

If necessary, the variance component model could be improved by estimating more than one variance for the observations of the fractions. Correlation between these observations could be estimated too.

Just as in conventional training for a maximum likelihood classification enough training samples should be available to estimate all (co)variances. A negative value for the estimate of a (co)variance factor can indicate a lack of training samples. The precision of the estimated (co)variance factors depends on the redundancy of the model, The quality of the estimated (co)variance factors is given by the covariance matrix:

$$Q_{\hat{\sigma}}=2\cdot\begin{bmatrix}N_{11} & \cdots & N_{1P} \\ \vdots & \ddots & \vdots \\ N_{P1} & \cdots & N_{PP}\end{bmatrix}^{-1} \qquad (14)$$

2.4 Iterations

The variance component model estimation of the (co)variances is executed after the least squares estimation of the mean values, since the result of the least squares estimation is needed for the variance component model estimation. Because the result of the least squares estimation also depends on the presumed variance model, an iterative computation should be performed. This is a third iteration on top of the iterations mentioned before.

Third iteration
Iteration in estimation of mean values
Iteration in estimation of (co)variances

3. EXPERIMENT

The presented method, to execute a maximum likelihood training that estimates the pure spectra out of mixed pixels, was tested with an experiment. To get mixed pixels with known mixture proportions, an image and accompanying raster file with ground truth were resampled coarser. The used image is a part of a recording of the Landsat Thematic Mapper. This image with pixels of 30 by 30 meters covers a small part of the polder Flevoland around the village Biddinghuizen in the Netherlands (figure 4a). The spectral bands 3, 4 and 5 were used. For this experiment, only the three most common crops were considered: wheat, potatoes and sugar beet (figure 4b).

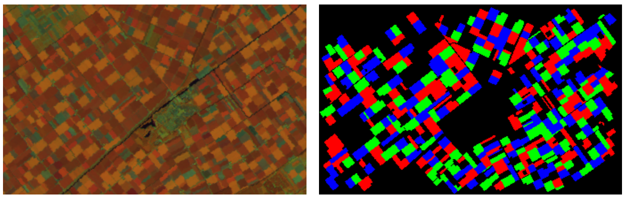

Figure 4. a. Landsat image, bands 4 (red), 5 (green) and 3 (blue); b. Raster file with ground truth, classes wheat (red), potatoes (green) and sugar beet (blue).

The Landsat image as well as the raster file with ground truth was sub-sampled with a factor 8. Per area of 8 by 8 pixels, the mean value was computed for each spectral band of the image, resulting in a new image with three bands with pixels of 240 by 240 meters (figure 5a). In the raster file with ground truth, the fractions of each class were computed for each area of 64 pixels (figure 5b).

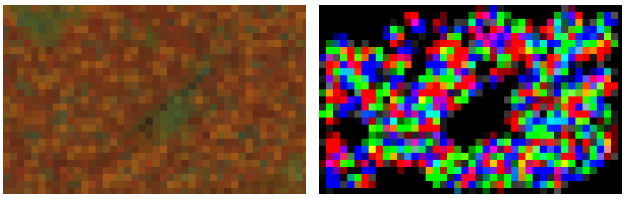

Figure 5. a. Resampled Landsat image as if recorded with a sensor with coarser pixels; b. Resampled raster file with ground truth, resulting in area proportions of the three crops as red, green and blue.

Because more classes than wheat, potatoes and sugar beet are present in the image, not all pixels are used for fuzzy training with the presented method. For conventional, crisp training however, even fewer pixels would be available. There are 243 pixels in which the three classes cover over 90% (figure 6a) and only 79 pixels where one class covers 90%. Of these 243 pixels, a subset of 121 pixels was used to test the presented training method. The results were compared with conventional training on a subset of the original image (figure 6b).

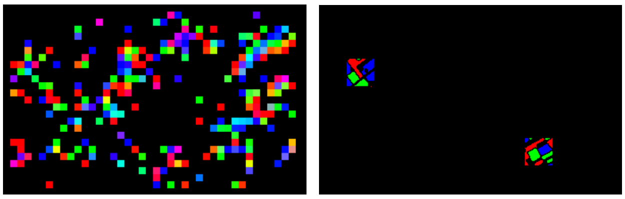

Figure 6. a. Resampled pixels in which the three classes together cover 90%, half of these are used for fuzzy training with the presented method; b. Unresampled pixels used for conventional training.

Comparison of the fuzzy method with the conventional crisp training showed similar mean values for the spectra of the classes. Covariance matrices of spectra are not so easy to compare, classification with the crisp and fuzzy spectra, however, both showed an overall accuracy of about 80%. Hence, it can be concluded that the principle of using mixed pixels for the training of a maximum likelihood classification is proven to work.

Further analysis of the results showed incidental negative estimates of fractions of classes (extreme case -0.9), probably as a result the presence of other classes than wheat, potatoes and sugar beet. Adaptation of the model to constrain positive fractions could be considered for further research. Next, the quality descriptions were examined, which indicated accurate estimation. The quality for the least squares estimation of the mean values, however, seemed a little to optimistic. Finally, the influence of the a priori values on the variance model was investigated. Because of the third iteration (section 2.4), the a priori values give negligible differences. However, this is only true as long as the a priori values are approximately right. Using a priori values with an error of a factor two causes the solution to diverge. As the resulting estimates are totally wrong when it happens, this is easily recognisable.

4. DISCUSSION

The experiment proved the functioning of the presented method. Whether it can actually give an improvement still has to be demonstrated in practice.

An advantage of the method is that more pixels in the image can be used for training. This enables the use of heterogeneous areas for training or the random selection of training pixels, which will give more representative training samples, as selection of typical training samples will.

Precise geo-referencing is very important as the in the field observed fractions of classes should match the training pixels in the image as good as possible.

The mayor disadvantage of the presented method is that it is more complicated. Although it is easily programmable, more knowledge of the operator is needed. For example, the number of needed training samples for representative estimation is no longer constant, as it depends on the mixture proportions of the training pixels. Using 'almost pure' pixels would require less pixels then pixels in all mixture proportions. In the case that all training samples have almost exactly the same mixture proportion, the estimates will be very inaccurate. Therefore, the quality descriptions (Q) need to be observed watchfully.

Error detection using the testing theory (Teunissen, 2000) can also be executed, to make sure potentially present errors and mistakes are detected and not influencing the estimates. However, this is true for conventional classification training too.

5. CONCLUSIONS

This article presented a method to execute training for maximum likelihood classification, which estimates unbiased pure spectra out of mixed pixels using adjustment theory and probability model estimation. An experiment proved the functioning of the presented method. More pixels in the image can be used for training with this method, which enables the use of heterogeneous areas for training or the random selection of training pixels.

There are two conditions for the presented training method. First, one needs to have estimates of the fractions of the classes in the mixed training samples. Secondly, the spectral values of the mixed pixels should be a linear combination of the spectra of the composing classes.

REFERENCES

Eastman, J. R. and R. M. Laney, 2002. Baysian soft classification for sub-pixel analysis: a critical evaluation. *Photogrammetric engineering & remote sensing*, 68(11), pp. 1149-1154.

ENVI, 1999. *ENVI tutorials; version 3.2.* Better Solutions Consulting, Lafayette (Colorado), USA.

Foody, G. M., 1999. The continuum of classification fuzziness in thematic mapping. *Photogrammetric engineering & remote sensing*, 65(4), pp. 443-451.

Gorte, B. G. H., J. Lesparre and R. W. L. Jordans, 2003. Probabilistic segmentation and fuzzy classification of natural vegetation in hyper-spectral imagery. *First international conference "Studying land use effects in coastal zones with remote sensing and GIS".* Kemer (Antalya), Turkey.

Lesparre, J., 2003. *Probabilistische segmentatie en fuzzy classificatie van natuurlijke vegatatie in hyperspectrale beelden.* Master thesis, Delft University of Technology, Delft, the Netherlands.

Teunissen, P. J. G., 1999. *Adjustment theory; an introduction.* Delft University Press, Delft, the Netherlands.

Teunissen, P. J. G., 2000. *Testing theory; an introduction.* Delft University Press, Delft, the Netherlands.

Teunissen, P. J. G., and A.R. Amiri-Simkooei, 2006. Least-squares variance component estimation, *VI Hotine-Marussi Symposium.* Wuhan, China.

Wang, F., 1990. Fuzzy supervised classification of remote sensing images. *IEEE transactions on geoscience and remote sensing*, 28(2), pp. 194-201.

ACKNOWLEDGEMENTS

The research presented in this paper was facilitated by the Survey Dept. (now AGI) of the Ministry of Transport, Public Works and Water Management, Derdewerelddreef 1, 2622 HA Delft, the Netherlands.