Phase-correlation-based hyperspectral image classification using multiple class representatives obtained with k-means clustering

D. Çeşmeci; M. K. Güllü

* Kocaeli University Laboratory of Image and Signal processing (KULIS), Department of Electronic and Telecommunication Engineering, University of Kocaeli, 41040 Kocaeli, Turkey


To link to this Article DOI: 10.1080/0143160902777183
URL: http://dx.doi.org/10.1080/0143160902777183

International Journal of Remote Sensing
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713722504

Please scroll down for article

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Letter

Phase-correlation-based hyperspectral image classification using multiple class representatives obtained with $k$-means clustering

D. ÇEŞMECI and M. K. GÜLLÜ*

Kocaeli University Laboratory of Image and Signal processing (KULIS), Department of Electronic and Telecommunication Engineering, University of Kocaeli, 41040 Kocaeli, Turkey

(Received 17 September 2008; in final form 24 January 2009)

In this letter, a modification to a phase–correlation-(PC-)based supervised classification method for hyperspectral data is proposed. An adaptive approach using different numbers of multiple class representatives (CRs) extracted using PC-based $k$-means clustering for each class is compared with the use of selecting a small, pre-determined number of dissimilar CRs. PC is used as a distance measure in $k$-means clustering to determine the spectral similarity between each pixel and cluster centre. The number of representatives for each class is chosen adaptively, depending on the number of training samples in each class. Classification is performed for each pixel according to the maximum value of PCs obtained between test samples and the CRs. Experimental results show that the adaptive method gave the highest classification accuracy (CA). Experiments on the effect of reducing the size of the feature vectors found that CA increased as the feature vector decreased.

1. Introduction

Hyperspectral images consist of hundreds of spectral bands ranging from visible light to mid-infrared areas of the electromagnetic spectrum. Due to the size of the hyperspectral data and the spatial variability of spectral signatures, classification and segmentation of hyperspectral images can be difficult.

Classification approaches can be divided into supervised and unsupervised methods (Duda et al. 2000). Unsupervised classification algorithms provide identification of a structure from an unlabelled dataset. Image clustering methods for example, divide the image into disjoint sets of pixel segments (Gonçalves et al. 2008). Many supervised classification methods have been proposed in the literature for hyperspectral data (Camps-Valls and Bruzzone 2005, Demir and Ertükr 2008). Demir and Ertükr (2007) classify hyperspectral images using phase correlation (PC) and multiple class representatives (CRs). In their approach, spectral signatures were sub-sampled to improve the discrimination performance of PC in the presence of additional white Gaussian noise. Three different classification strategies were examined to evaluate the performance of the PC-based classification approach. In the first approach, the mean of all hyperspectral data in each class is used as a class feature vector. The PC between each unknown pixel and each class feature vector is computed, and the pixel is labelled with the class that gives the highest PC value. In

*Corresponding author. Email: kemalg@kocaeli.edu.tr
the second approach, five representatives are chosen for each class, using training 
samples that belong to the same class, but give the lowest PC value within that class. 
The PC between each unknown pixel data and the multiple CR feature vectors is 
computed, and the new pixel is classified as the class for which the highest PC value 
is obtained. In the third approach, multiple CRs are selected, as with the preceding 
approach, but the average of the PC results of multiple representatives within each 
class is computed for a decision (Demir and Ertürk 2007).

In this letter, a modification to the classification strategy of Demir and Ertürk 
(2007) is proposed, in which representatives of each class are selected using 
unsupervised classification applied to the training data. Initially, as in Demir and 
Ertürk (2007), spectral sub-sampling is applied to the data to reduce the 
computational load and gain robustness in the presence of noise. As described 
above, Demir and Ertürk (2007) selected CRs based on dissimilarity. In the proposed 
method, the CRs are chosen using the \( k \)-means algorithm in which the number of 
clusters is obtained adaptively according to the size of the training samples for a 
given class. PC is used as distance measure for the \( k \)-means clustering to determine 
the spectral similarity of the signatures. Classification is performed for each 
unknown pixel according to the maximum value of the PC result between the 
signature of that pixel and the CRs. The impact of the sub-sampling ratio and the 
gain obtained by using multiple CRs on classification accuracy (CA) is also 
evaluated.

2. Modified PC-based supervised classification

PC is a method that can be used to determine the similarity of two signals or images 
exploiting fundamental properties of the discrete Fourier transform (DFT) (Urhan 
\textit{et al.} 2006). Mathematically, PC is defined as:

\[
r = F^{-1} \left\{ \frac{G_a G_b^*}{|G_a G_b|^2} \right\}.
\] (1)

Here, \( G_a \) and \( G_b \) are the DFTs of two spectral signatures \( g_a \) and \( g_b \), respectively; 
\( * \) denotes the complex conjugate; \(|\cdot|\) denotes the modulus; and \( F^{-1} \) represents 
the inverse DFT (IDFT). Therefore, the PC of two spectral signatures is the IDFT of 
the normalized cross power spectrum. If the hyperspectral data of two pixels are the 
same, i.e. \( g_a = g_b \), the PC result is obtained as:

\[
r = F^{-1} \left\{ \frac{G_a G_a^*}{|G_a G_a|^2} \right\} = F^{-1} \{1\} = \delta.
\] (2)

Here, \( \delta \) represents the discrete unit impulse function. Hence, the PC result of two 
identical signals will have a peak value of unity located at zero. The PC result will 
have a high peak if the two spectral signatures have a high resemblance, but the peak 
value will typically be lower than unity due to noise and spectral variability. The 
important property of PC compared to the classical cross correlation is that the peak 
in the PC can be detected much more accurately because the PC provides a distinct 
sharp peak in the case of correspondence, which is not the case in cross correlation. 
Note that spectral sub-sampling is performed (the number of bands is reduced) on 
the hyperspectral data before PC to reduce the computational load and gain 
robustness against noise. A 10th order low-pass finite impulse response filter is used
for this purpose. The PC of sub-sampled data was called the modified PC (MPC) by Urhan et al. (2006). Furthermore, reducing the feature size mitigates the Hughes phenomenon when the training set size is not large enough (Landgrebe 2003).

2.1 Modified PC-based k-means clustering

In this letter, k-means clustering of training samples is proposed to extract the CRs in hyperspectral data for the purpose of classification. The k-means is one of the simplest unsupervised clustering methods that has been used for well-known clustering problems, such as image segmentation and colour quantization (Hartigan 1975). In k-means clustering, the problem is to determine a set of \( k \) partitions in \( d \)-dimensional space by trying to minimize the total intra-cluster variance. In the proposed approach, MPC is used as a distance measure instead of the \( L_2 \) norm (Euclidean distance) because MPC gains robustness in the presence of noise and variability in the signal. Consequently, this clustering is referred to as \( k \)-means based on modified phase correlation (KM-PC). Training data of each class is clustered using KM-PC and the number of clusters (number of class representatives; NCR) is chosen using the square root of the total number of training data in that class. Note that the square root function is chosen because it provides a simple non-linear mapping. Therefore, the NCR will be small if the training data size is small for that class. Extraction of the CRs using KM-PC is shown in figure 1. In this figure, \( C_M \) represents the class label and \( M \) denotes the total number of classes. In the case of small intra-class variance, the CRs will be similar to each other. Furthermore, the number of similar CRs will potentially also be high with increased training data size. Hence, CRs with relatively high spectral similarity, as indicated by their PC, are merged to reduce the final NCR using a similarity threshold (\( T \)).

2.2 Classification

PC is used to classify the hyperspectral data, similar to the method proposed by Demir and Ertürk (2007), but the NCR is variable and extracted using KM-PC, unlike the work presented in Demir and Ertürk (2007), where the NCR was fixed and the same for each class. For classification purposes, the PC is computed between the spectral signature of the pixel to be classified and all CRs. The PC between a new pixel and the CRs are computed simultaneously and the new pixel is classified as the class for which the highest PC value is obtained.

3. Experimental results

Experimental results are presented using two hyperspectral datasets. The first dataset is the Washington DC Mall hyperspectral image, which was collected by the Hyperspectral Digital Imagery Collection Experiments (HYDICE) system over Washington DC Mall in August 1995 (Landgrebe 2003). The data consist of 1280 × 307 pixels, with 210 spectral bands in the 0.4 to 2.4 \( \mu \)m region of the visible and infrared spectrum, and has been reduced to 191 bands by excluding bands in the 0.9 and 1.4 \( \mu \)m water absorption regions. The pixel size is about 3.2 m. A total of 3420 training and 4659 testing samples are used for the DC Mall dataset, which are divided proportionally between classes. The training and test samples were constructed by systematically assigning ground reference pixels in an alternating fashion to the training and test sets. Class information is given in table 1 for the DC Mall dataset. The second hyperspectral image was acquired over the northwest
Indiana Indian Pine test site on 12 June 1992 by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) (available at: ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/92AV3C.lan). The Indian Pine dataset consists of $145 \times 145$ pixels, with

![Diagram]

Figure 1. Extraction of the class representatives using KM-PC.

Table 1. The number of training and test samples for each class of the Washington DC Mall dataset and associated number of class representatives (NCRs).

<table>
<thead>
<tr>
<th>Class</th>
<th>Training pixels</th>
<th>Test pixels</th>
<th>NCR</th>
<th>NCR (merging threshold ($T$))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.999</td>
</tr>
<tr>
<td>C1 – Roof</td>
<td>1500</td>
<td>2334</td>
<td>39</td>
<td>35</td>
</tr>
<tr>
<td>C2 – Street</td>
<td>200</td>
<td>216</td>
<td>14</td>
<td>7</td>
</tr>
<tr>
<td>C3 – Path</td>
<td>80</td>
<td>95</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>C4 – Grass</td>
<td>800</td>
<td>1128</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>C5 – Trees</td>
<td>200</td>
<td>205</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>C6 – Water</td>
<td>600</td>
<td>624</td>
<td>24</td>
<td>6</td>
</tr>
<tr>
<td>C7 – Shadow</td>
<td>40</td>
<td>57</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>3420</strong></td>
<td><strong>4659</strong></td>
<td><strong>134</strong></td>
<td><strong>99</strong></td>
</tr>
</tbody>
</table>
220 bands, but 20 bands affected by atmospheric absorption were discarded. The original ground reference data provided with the imagery has 16 classes, but the five classes given in table 2 were used in this study in order to allow a direct comparison with the results of Demir and Ertürk (2007). The training and test samples were constructed by systematically assigning pixels in an alternating fashion to the training and test sets.

Various sub-sampling factors were compared to evaluate the contribution of the sub-sampling to classification, and the obtained CAs are given for the Washington DC Mall data in table 3. It is seen from table 3 that sub-sampling increases the CA to 99.4% for a sub-sampling rate of 6. For comparative evaluation, Euclidean distance was used instead of PC as a similarity metric in the clustering and classification stages, and results are also presented in table 3. Comparing the discrimination performance, it is seen that the proposed method, which uses PC gives a higher CA, except for the analysis with the raw (not sub-sampled) data. The classification performance using PC for sub-sampling at rates of 4, 6 and 8 are similar, while the classification performance decreases for these sub-sampling rates when using Euclidean distance. In summary, both Euclidean distance and PC-based classification schemes benefit from reduction in feature vector size, but the PC method appears to benefit more, and benefits from high rates of reduction.

A CA of 74.2% was obtained if the mean of all hyperspectral data in each class was used as the class feature vector (i.e. CR) and PC-based classification is performed with a sub-sampling factor of 6. This much lower accuracy demonstrates the advantage of using multiple CRs. This shows that classification based on multiple representatives gives a superior CA compared to single representative based classification.

Table 3. Effect of the sub-sampling on the classification accuracy (CA).

<table>
<thead>
<tr>
<th>Sub-sampling factor</th>
<th>Euclidean distance</th>
<th>Phase correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 : 1</td>
<td>92.1</td>
<td>91.1</td>
</tr>
<tr>
<td>1 : 2</td>
<td>94.5</td>
<td>97.5</td>
</tr>
<tr>
<td>1 : 3</td>
<td>95.8</td>
<td>99.0</td>
</tr>
<tr>
<td>1 : 4</td>
<td>92.7</td>
<td>98.8</td>
</tr>
<tr>
<td>1 : 6</td>
<td>92.6</td>
<td>99.4</td>
</tr>
<tr>
<td>1 : 8</td>
<td>91.7</td>
<td>99.0</td>
</tr>
<tr>
<td>1 : 16</td>
<td>92.7</td>
<td>97.6</td>
</tr>
<tr>
<td>1 : 32</td>
<td>91.2</td>
<td>88.6</td>
</tr>
</tbody>
</table>
For the case of adaptive-merged NCRs, CAs and the number of reduced CRs were computed for various threshold values ($T$) for the Washington DC Mall dataset. The NCRs after merging were given in table 1 for $T=0.999$, $T=0.995$ and $T=0.990$, and CAs of 99.3%, 99.2% and 99.1%, respectively, were obtained.

The proposed method was also compared to the methods presented by Demir and Ertürk (2007) to evaluate the classification performance. The following notation is used to refer to the three classification methods proposed by Demir and Ertürk (2007): PC-1 denotes the case in which a single feature vector is used; PC-2 represents the multiple CRs case with a decision according to the maximum PC value; and PC-3 denotes the multiple CRs case with a decision according to the average PC values of each class. In the cases of PC-2 and PC-3, Demir and Ertürk (2007) proposed to assign five representatives for each class, which are selected using training samples that belong to the same class, but are least similar to each other, and PC-2 is compared to the proposed method to evaluate the advantage of extracting CRs by the KM-PC approach. To make an objective comparison with PC-2, the number of clusters is also fixed at five in the proposed method. In this case, the CA of the proposed method was 83.5%, compared to 62.4% for PC-2. Hence, the proposed KM-PC-based CR extraction approach increases the CA of the classifier by about 20%. The main reason for this gain is likely because PC-2 identifies CRs at the boundaries of the training dataset. For example, the obtained CRs for class 2 (streets) of both methods are shown in figure 2. As can be seen from figure 2, for PC-2, the CRs are placed in the extremes of the class distribution, which may result in spectral confusion between the different classes, and therefore may reduce the classification performance of PC-2. It is notable that the adaptive extraction of the NCRs, proposed in this letter, increased the CA by about 16%.

Finally, the CA of the proposed method was compared to PC-1, PC-2 and PC-3 for the Indian Pine data. Note that the training and test sets were used as in Demir and Ertürk (2007). The proposed KM-PC-based classification algorithm is evaluated for adaptive, merged-adaptive and fixed NCR cases. Classification results for three, four and five classes of the Indian Pine data are given in table 4 to compare the performance of the methods. The results show that the proposed method with variable and fixed NCRs resulted in overall better CAs. Furthermore, the use of adaptive NCRs increased the CA. For the case of adaptive-merged NCRs, three

![Figure 2: Class representatives obtained for both the proposed KM-PC and the method proposed by Demir and Ertürk (2007) for class 2 of the Washington DC Mall data.](image-url)
Threshold values were chosen arbitrarily for the PC. Note that the number of reduced CRs is given in Table 2 for the Indian Pine data. Note, however, that the accuracy was reduced through the merging of CRs, suggesting that higher numbers of CRs are preferable.

4. Conclusions

In this letter, a modification to PC for supervised classification developed by Demir and Ertürk (2007) based on an adaptive number of multiple representatives is proposed. Multiple representatives were extracted using PC-based $k$-means clustering. The number of clusters was determined according to the number of class training samples. The adaptive method was found to produce higher accuracies compared to the static number of dissimilar CRs selected with the methods proposed by Demir and Ertürk (2007). However, merging clusters lowered the accuracy. Both Euclidean distance and PC-based clustering and classification benefited from sub-sampling of the feature vector, although the PC method benefited the most, and benefited from higher degrees of sub-sampling.

Acknowledgements

This work was supported by the Scientific Research Projects Unit, University of Kocaeli, Turkey under grant 2007/45 and the Turkish Scientific and Technological Research Council (TUBITAK) project EEEAG/107E011.

References


