TYPICAL SEQUENCE CLASSIFICATION METHOD IN HYPERSPECTRAL IMAGES WITH REDUCED BANDS

Samir Youssif Wehbi Arabi¹, ², David Fernandes², Marco Antonio Pizarro³, Marcelo da Silva Pinho²

¹ Instituto Federal de Educação, Ciência e Tecnologia de Goiás – IFG,
² Instituto Tecnológico de Aeronáutica – ITA/DCTA,
³ Instituto Nacional de Pesquisas Espaciais – INPE

E-mails: samir.arabi@ifg.edu.br, marco.pizarro@inpe.br, {david, mpinho}@ita.br

ABSTRACT

This work presents a new method for hyperspectral spectra classification based on the Typical Sequence (TS) derived from the Asymptotic Equipartition Theorem and Information Theory. Each Endmember (EM) of a scene is represented by a Hidden Markov Model (HMM) and a spectrum is classified in a given class if it can be considered a TS generated by the HMM associated with the EM related to the class. The Discrete Wavelet Transform (DWT) is used in the orthogonal decomposition of the original spectrum and the HMM parameters are estimated using this orthogonal decomposition. The proposed method is tested with AVIRIS spectra of a scene with 13 EM and the classification results show that 32 spectral bands can be used instead of the original 209 bands, without significant loss in the classification process.

Index Terms — Typical Sequence, Hyperspectral, Image Classification, HMM

1. INTRODUCTION

The hyperspectral technology improves the spectral resolution, but the high number of bands requires a higher data storage capability, a high data transmission rate and has a high computational cost in the processing systems. Several works have been developed in order to reduce the data dimensionality without loss of information in the desired results [1, 2, 3]. This work presents a new methodology for hyperspectral data classification that does not require all hyperspectral bands to have a good classification performance.

The methodology is based on three steps: a) Decomposition of original spectrum in K orthogonal functions using the Discrete Wavelet Transform (DWT), b) Hidden Markov Model (HMM) parameters estimation of the Endmembers (EM) orthogonal functions and c) Spectrum classification based on the Typical Sequence (TS) concept that verifies if a given spectrum can be consider a TS generated by a HMM associated to an EM. The concept of TS is derived from the Asymptotic Equipartition Theorem in the Information Theory context [4, 5].

Chapter 2 gives details about the methodology based on the TS and the HMM parameters estimation. In Chapter 3 the methodology is applied in a set of AVIRIS spectra with 13 EM and 209 bands. It is shown that with only 32 bands the proposed classification method can achieved equivalent results when compared with other classification methods that use all bands. In Chapter 4 some final remarks are presented.

2. METHODOLOGY OF CLASSIFICATION WITH HMM AND TYPICAL SEQUENCE

Each pixel (i, j) of a hyperspectral image with bands \( t = 1, 2, \ldots, T \) will be represented by the spectral vector \( E_i^{t} = [e_1^{ij}, e_2^{ij}, \ldots, e_T^{ij}]^T \). For each spectral vector the DWT generates a set of K orthogonal functions, with T bands, represented by \( O_i^{r,k,T} = \{O_{1,k,T}^{r}, O_{2,k,T}^{r}, \ldots, O_{k,T}^{r}\} \), that can be modeled as a HMM with parameters \( \lambda^{i,j} = (\pi^{i,j}, A^{i,j}, B^{i,j}) \), where \( \pi^{i,j} \) is the initial state probability, \( A^{i,j} \) is the state transition matrix and \( B^{i,j} \) is the observation probability matrix [6, 7, 8]. The HMM process related to \( \lambda^{i,j} \) will be represented by \( HMM(\lambda^{i,j}) \).

Figure 1 shows the generation of \( O_i^{r,k,T} \) through the DWT, a zero padding operation and the DWT\(^{-1}\). With this proceeding each spectra with T bands generates K orthogonal signals also with T bands.

It will be considered that the R Endmembers (EM), or pure spectra, \( E_{1,r} = [e_1^{r}, e_2^{r}, \ldots, e_T^{r}]^T \), \( r = 1, 2, \ldots, R \), of a scene is given, and its DWT orthogonal function representation \( O_{1,r,T} = \{O_{1,k,T}^{r}, O_{2,k,T}^{r}, \ldots, O_{k,T}^{r}\} \), \( r = 1, 2, \ldots, R \) has the HMM parameters \( \lambda^{r,k} \).
Wavelet coefficients:

\[
W_{L_1}^T:2^{-k+1} \rightarrow \text{Zero Padding} \rightarrow L_1
\]

\[
W_{L_2}^T:2^{-k+1} \rightarrow \text{Zero Padding} \rightarrow L_2
\]

\[
\cdots
\]

\[
W_{L_k}^T:2^{-k+1} \rightarrow \text{Zero Padding} \rightarrow L_k
\]

Figure 1: Spectra Decomposition

Figure 2 illustrates the HMM estimation parameters procedure considering the initial parameter \( \lambda_P^v \) obtained by the K-Means Algorithm. The observation \( O_{k,i,t}^{P,r} \) is composed by \( T \) vectors \( O_{k,i,t}^{P,r} \), and this vectors are grouped by the K-Means Algorithm in \( N \) cluster in the \( \mathbb{R}^k \) space, where \( N \) is the number of states chosen previously. With this procedure a state sequence with \( N \) symbols and length \( T \) is generated and the initial parameters \( \lambda = (\pi, A, B) \) can be also estimated. After this, the Baum-Welch Algorithm is used for the \( \lambda_P^v \) estimation and \( P(O_{k,i,t}^{P,r} | \lambda_P^v) \) calculation [9, 10].

\[
H(O_{k,i,t}^{P,r} | \lambda_P^v) = -\frac{1}{T} \log \left[ P(O_{k,i,t}^{P,r} | \lambda_P^v) \right]
\]

(1)

where \( P(O_{k,i,t}^{P,r} | \lambda_P^v) \) is the probability of the observation \( O_{k,i,t}^{P,r} \) belongs to the HMM \( (\lambda_P^v) \) with

\[
0 \leq H(O_{k,i,t}^{P,r} | \lambda_P^v) \leq H(O_{k,i,t}^{P,r})
\]

(2)

with the equality only if \( v = r \).

Let \( \Theta^{P,r} \) be the set of all possible orthogonal functions \( O_{k,i,t}^{P,r} \) that can be generated by the HMM \( (\lambda_P^v) \). A partition \( \Theta_{\delta}^{P,r} \) of \( \Theta^{P,r} \) can be defined as [4]:

\[
\forall \epsilon, \delta > 0 \exists T_0 \in \mathbb{N} : T > T_0 \Rightarrow \exists \Theta_{\delta}^{P,r} \subseteq \Theta^{P,r} \text{ with } \Theta_{\delta}^{P,r} = \left\{ O_{k,i,t}^{P,r} : |H(O_{k,i,t}^{P,r} | \lambda_P^v) - H(\lambda_P^v)| \leq \delta \right\}
\]

(3)

and for \( O_{k,i,t}^{P,r} \in \Theta_{\delta}^{P,r} \) we have also that [4]

\[
P\left[ |H(O_{k,i,t}^{P,r} | \lambda_P^v) - H(\lambda_P^v)| \leq \delta \right] \geq 1 - \epsilon
\]

(4)

The subset \( \Theta_{\delta}^{P,r} \) defined in (3) is a Typical Sequence Set (TSS) and \( O_{k,i,t}^{P,r} \in \Theta_{\delta}^{P,r} \) is called TS of the HMM \( (\lambda_P^v) \). The HMM \( (\lambda_P^v) \) entropy is estimated by

\[
H(\lambda_P^v) \equiv H(O_{k,i,t}^{P,r} | \lambda_P^v)
\]

(5)

It is desired that the observation \( O_{k,i,t}^{P,r} \) associated to a class \( \omega \in \Omega \) be considered an element of the TSS related to the HMM \( (\lambda_P^v) \) and the same observation \( O_{k,i,t}^{P,r} \) should not be considered in TSS related to the HMM \( (\lambda_{P,v}) \) \( v \neq r \). To achieve this condition we must have that:

\[
O_{k,i,t}^{P,r} \in \Theta_{\delta}^{P,r} \text{ and } O_{k,i,t}^{P,r} \not\in \Theta_{\delta}^{P,v} \text{, } v \neq r
\]

\[
\Rightarrow \delta_r < \min_{k \neq \epsilon, r} \left[ |H(O_{k,i,t}^{P,r} | \lambda_P^v) - H(\lambda_P^v)| \right]
\]

(6)

If

\[
|H(O_{k,i,t}^{P,r} | \lambda_P^v) - H(\lambda_P^v)| > \delta_r \text{, for a given } r = r_0
\]

the observation \( O_{k,i,t}^{P,r} \) related to the HMM \( (\lambda_{P,v}) \) was not a good option as EM, and thus should be replaced.

Established the concept of TS and the parameter \( \delta_r \), let us consider for a pixel \((i,j)\) the observation \( O_{i,j}^{P,r} \) that must be classified in a class of \( \Omega_{i,r} = \{\omega_1, \omega_2, \ldots, \omega_R\} \). We can state the following decision rule:

\[
O_{i,j}^{P,r} \in \omega_r \Leftrightarrow \left| H(O_{i,j}^{P,r} | \lambda_P^v) - H(\lambda_P^v) \right| < \min_{v \neq r} \left[ |H(O_{i,j}^{P,v} | \lambda_P^v) - H(\lambda_P^v)| \right]
\]

(7)

\[
\wedge \left| H(O_{k,i,t}^{P,v} | \lambda_P^v) - H(\lambda_P^v) \right| < \delta_r
\]

Then the entropy of \( O_{i,j}^{P,r} \) can be calculated as [9, 10]:

1695
3. EXPERIMENTAL EVALUATION

Some real spectra of sub-scene 4, line of flight northeast of Campo Grande / MS (Brazil) region, acquired by AVIRIS in 1995 were used to evaluate the classification process performance based on TS. In the scene it was chosen a set of \( R = 13 \) EM with \( T = 209 \) bands, composed by five types of green vegetation (GV), two types of nonphotosynthetic vegetation (NPV), three types of red soil Latossolo Vermelho (LV), two types of Neossolo Quartzarênico Órtico (RQo) soil and one type of water [12]. The spectra of the EM are shown in Figure 3.

The process described in Chapter 2, and resumed in the Figure 1, was applied in the original 209 bands spectra and also in 32 bands, randomly chosen, among the 209. The DWT with \( K = 4 \) orthogonal decomposition (were used the Symlets with order six – Sym6) was applied, and the HMM parameters were estimated by the Baum-Welch algorithm, considering three states for the water spectrum and five states for all other EM spectra. The classification was done in the neighborhood of each EM pixel using a small square window of 3x3, 5x5 and 7x7 pixels considering homogeneity in this neighborhood. For comparison it was done the classification using SAM (Spectral Angle Mapper), ED (Euclidean Distance) and SID (Spectral Information Divergence) [9]. The Kappa index was calculated for all discriminatory measurements.

Table 1 shows the classification in 13 classes, 209 and 32 bands, using only the 13 EM without its neighborhood.

<table>
<thead>
<tr>
<th>Bands:</th>
<th>209 bands</th>
<th>32 bands</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>ED</td>
<td>1.00</td>
<td>0.83</td>
</tr>
<tr>
<td>SAM</td>
<td>1.00</td>
<td>0.33</td>
</tr>
<tr>
<td>SID</td>
<td>1.00</td>
<td>0.83</td>
</tr>
</tbody>
</table>

It can be observed that if the numbers of spectral bands decrease, only the TS Kappa indices remain in excellent agreement. The reason for the TS better performance can be attributed to the independence among sparse bands, and therefore it results in a better estimation of the EM HMM parameters. The SAM classifier was severely penalized by the number of bands reduction.

Table 2 shows the classification using the EM pixel neighborhood of 3x3, 5x5 and 7x7 pixels and 209 and 32 bands.

<table>
<thead>
<tr>
<th>Bands:</th>
<th>209 bands</th>
<th>32 bands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Window:</td>
<td>3x3</td>
<td>5x5</td>
</tr>
<tr>
<td>TS</td>
<td>0.83</td>
<td>0.75</td>
</tr>
<tr>
<td>ED</td>
<td>0.98</td>
<td>0.95</td>
</tr>
<tr>
<td>SAM</td>
<td>0.93</td>
<td>0.86</td>
</tr>
<tr>
<td>SID</td>
<td>0.99</td>
<td>0.96</td>
</tr>
</tbody>
</table>

Table 3: Kappa in the classification of EM’s neighborhood.

Figure 3: EM spectra with 209 bands.
In Table 2 for 209 bands and 32 bands, it was observed that when the neighborhood increases the Kappa decreases for all methods, due to the spectral contamination.

As in Table 1 it can be observed in Table 2 that if the number of spectral bands decrease, only the TS Kappa indices remain in excellent agreement (greater than 0.81).

It can be also observed that in the classification using 32 bands the TS classifier is better than the others and also comparable with the SAM classification with 209 bands.

The TS using 32 bands has in the worst case the Kappa 7.5% smaller than ED Kappa with 209 bands and 8.3% smaller than SID Kappa with 209.

Take into account Table 1 for 209 bands the contamination is worst in TS classification method.

4. FINAL REMARKS

This paper presents a new hyperspectral data classification method using the HMM modeling and Typical Sequences concept, derived from Information Theory. The classification method was tested with data obtained from an AVIRIS image and was compared with the classification performed by SAM, SID and ED. With the data set used, the classification performance with the proposed classification method, with only 32 sparse bands, achieved equivalent results in comparison with the classification done with other methods, using all 209 bands. This proposed TS classification method can be used as an alternative methodology for hyperspectral image classification or as a method for image classification of an Imaging System with reduced bands. So this method can be considered an alternative methodology for the classification of hyperspectral images with reduced bands, resulting in advantages on processing, transmission and storage of the data.

5. ACKNOWLEDGEMENT

This work was supported by the CAPES-DINTER Brazilian program.

6. REFERENCES


