Matching hyperspectral absorptions by weighted hamming distance

Baofeng Guo
School of Automation, Hangzhou Dianzi University, Hangzhou, China
Email: gbf@hdu.edu.cn

Introduction: In a typical task of hyperspectral data classification [1, 2], the data acquired from a hyperspectral sensor can be modelled as an electromagnetic reflectance function against different wavelengths or bands. Based on these hyperspectral reflectance curves (i.e. the spectral “signature”), a group of suitable features have to be extracted or selected before we can feed them into classifiers. Apparently, these features should be as effective as possible to include the fundamental characteristics in line with each type of the materials. In the past two decades, many feature representations have been investigated, such as the spectra curves [3], the subsets of spectral bands [4], the features based on manifold learning [5], the spatial-spectral features [6] etc.

Recently, absorption features have been attracting more and more attentions [7]. This methodology is justified by the spectrometry theory [8], namely mass always absorb incident light at some particular wavelengths or bands. The detailed values of these wavelengths or bands are related with the materials’ constituents [9]. Hence, for each type of material, a group of special dips or valleys, i.e. absorptions, will appear on the hyperspectral reflectance curve. In theory, this set of absorptions should correspond to the material’s chemical constituents exactly, and therefore can be served as a kind of identity.

But in practice, it is observed that the absorptions are connected with materials not only by its chemical constituents, but also by many other conditions, such as surface roughness and various environmental factors. Particularly in remote sensing, the absorptions, originated from materials’ chemicals, are often mixed with those features caused by atmospheric transmission and Sunlight’s spectrum. Moreover, the amounts or the levels of materials’ chemicals (e.g. water of vegetation) are often changing with different seasons or growing status. This makes it much more complicated to use the absorption features for hyperspectral classification.

In this letter, we address the aforementioned problems through two succeeding steps, namely a robust absorption feature extraction and an effective matching algorithm.

Improved absorption detection: To find out the absorption features, the spectra have to be extracted from a hyperspectral date cube, and then normalised to the range of [0, 1]. For example, a spectra is represented by a vector \( x = (X_1, X_2, \ldots, X_L) \), with \( X_i \in [0, 1] \) standing for the normalised radiance value at the \( i \)-th band, and \( L \) the total number of the bands. A standard peak detection algorithm, shown in Equation (1), can be applied to the normalised spectra \( x \) to retrieve all possible absorption valleys.

\[
F_i = \begin{cases} 
1, & \text{if } X_i < X_{i+1} \text{ and } X_i < X_{i-1}, i = 2, \ldots, L - 1 \\
0, & \text{otherwise}
\end{cases}
\]

where the binary vector \( F_i \) indicates whether an absorption occurs at the \( i \)-th band, and \( f = (F_1, F_2, \ldots, F_L) \) is an initial absorption vector corresponding to the spectral vector \( x \).

It is not appropriate to work directly at the initial absorption vector \( f \), for the detection algorithm shown in Equation (1) is very sensitive to sensors’ noise. It is seen from Equation (1) that this algorithm only considers its two neighbouring bands. When the central band is slightly lower than its neighbourhoods, it will be detected as an absorption band. But this spectral valley is more likely to be caused by random noise from sensors or interferences during light transmission, rather than a valid absorption originated from materials’ constituents. So a modified absorption detection algorithm is proposed, depicted as follows:

\[
F_i = \begin{cases} 
1, & \text{if } X_i - X_i' < \delta \text{ and } X_i' \in R_i, i = 1, 2, \ldots, L \\
0, & \text{otherwise}
\end{cases}
\]

where the parameter \( R_i \) is a neighbourhood of the \( i \)-th band and \( \delta \) is a threshold.

In the above algorithm, the new parameter \( R_i \) is introduced to avoid those absorption valleys with the smaller bandwidths. Usually these narrow absorption valleys are related with the sensor’s white noise. Here, by observing the obtained training samples and considering the diagnostic absorption-bandwidths, we choose a neighbourhood of two bands, i.e. \( R_i = [X_i - \delta, X_i + \delta] \). In this way, a band \( i \) can be considered as a valid absorption only when it’s radiance value \( X_i \) is smaller than the four neighbouring values on its two sides.

The second parameter of the above algorithm is the threshold \( \delta \), which is used to exclude those absorption valleys with the lower absorption intensity. The shallower absorption valleys are usually associated with the scattering of the adjacent pixels or the atmospheric transmission. Comparing to the direct reflection from the material, the scattering of the adjacent pixels or the atmosphere are relatively weak, and the corresponding absorption-valleys are shallower thereby. In this application, the threshold \( \delta \) and neighbourhood width \( R_i \) in Equation (2) are decided empirically. First, the training samples are investigated and the spectral valleys appearing on less than half of the samples are picked up. They are more likely to be caused by noise or interferences for their lower probability of occurrence. Then, the depth and the width of these noisy spectral valleys are examined. Finally, the threshold \( \delta \) and the neighbourhood width \( R_i \) are selected as \( \delta = 0.05 \) and 5 respectively, which can exclude 95% of the noisy spectral valleys. Apparently, the detailed values of \( \delta \) and \( R_i \) are closely related with the sensors’ specifications as well as application scenarios, and should be reconsidered if the data set is changed.

Weighted hamming distance: After the absorption detection, a real-valued hyperspectral curve has been converted to a binary feature vector, in which the elements corresponding to the absorption bands are assigned as value 1 and other absorption-absent elements are assigned the value 0. Then, the Hamming distance can be used to measure how many absorption features are different between the two spectral curves, such as follows:

\[
D_h(x_1, x_2) = \sum_{i=1}^{L} (x_1^i \oplus x_2^i)
\]

where \( x = [X^1, X^2, \ldots, X^L]^T \) is a \( L \)-dimensional binary absorption feature vectors, and \( \oplus \) stands for an exclusive OR operation.

The Hamming distance should be 0 for the samples that are from the same class (i.e. the within class distance), because theoretically they should possess identical absorptions. The Hamming distance should be a positive value if the samples are from different classes (i.e. the between class distance), and the detailed divergence represents how many different absorptions occur between the two samples.

In an ideal laboratory environment, the absorption bands regarding the same material would remain constant no matter the samples or measurements change. But in practice, especially in remotely sensing, the acquired hyperspectral data are subjective to the changes of the environmental variables (such as the neighbouring objects, climates etc), and the outputs of the absorptions for the same material may vary in different runs. To cope with this variability, a weighting scheme is proposed to tolerate this alteration. First, a probability \( p_i \) is used to estimate the rate of occurrence of an absorption at a specific band such as the follows:

\[
p_i = \frac{n_i}{N}, \quad i = 1, 2, \ldots, L
\]
where \( n_i \) is the number of the absorptions detected within a group of \( N \) training samples for a specific material at the \( i \)-th band, and \( L \) is the total number of spectral bands.

According to Equation (4), the probability \( p_i \) can be considered as a likelihood of absorption occurrence at the \( i \)-th band. Observing the definition in Equation (3), the Hamming distance is a summation of a group of individual discrepancy corresponding to each of the bands. Since different bands of the same material may have different absorption-likelihood due to the uncertain environments, the bands with the higher possibility of absorption should have more weights. In other words, the higher the \( p_i \) is, the more contribution the Hamming distance should be given to this band. Therefore, a weighted Hamming distance is proposed as follows:

\[
D_{wh}(x_1, x_2) = \sum_{i=1}^L \left[ p_i \cdot (X'_i \otimes X'_j) \right] \tag{5}
\]

Comparing to the conventional Hamming distance in Equation (3), the weighted Hamming distance takes the environmental influences into consideration. Based on the training samples, the weights are proposed to model all random interferences to the occurrence of an absorption as a probability. Because the weighted Hamming distance includes the prior knowledge in the training samples and accounts for the uncertainty of absorption (an inevitable issue in remote sensing), it may provide a better measurement of the divergence that the Hamming distance sums up to estimate the difference between the two binary absorption feature-vectors.

**Elastic matching:** Applying the weighted Hamming distance to classification, it is found that the absorptions’ positions of different samples may displace along the spectral axis although they are belonging to the same material. Figure 3 shows this phenomenon in details, in which ten samples of the “corn” are illustrated based on the Salinas data set. The shifts of absorption valleys are taking place at 625 nm, where the absorptions of the different samples are spreading out within the range from 600 to 650 nm. These displacements of absorptions are probably caused by the changing proportions of the constituents. For example, the hyperspectral curves of vegetable’s samples may show different levels of water or different levels of chlorophyll according to their different growing or health status. It may give valuable information for crop growth inspection. But in classification, the absorptions from different samples cannot be matched exactly anymore, even though they are from the same class. For this reason, the matching algorithm should have a capability to accommodate the natural variation of absorption position.

To avoid the aforementioned mismatching, an elastic matching algorithm is proposed to allow a certain level of tolerance between the shifting of the absorptions’ positions. The idea is to append a sliding window into the Hamming distance to check whether the two absorptions are from a pair of the displaced absorptions. First, the Hamming distance are rewritten as a form of inner production:

\[
D_h(x_1, x_2) = (x_1 - x_2)^T \cdot (x_1 - x_2) \tag{6}
\]

Then, a special window is applied to traverse the part of \((x_1 - x_2)\) in Equation (6). If there are more than two positive elements detected in the \((x_1 - x_2)\) within a narrow window, it is likely that they are a pair of corresponding absorptions (i.e. they are displaced from the same absorption), and the Hamming distance within this window should be assigned as a value of 0 rather than number 2.

Based on Equation (6), the searching of the corresponding absorption pairs can be carried out as follows.

\[
D_h(x_1, x_2) = \sum_{j=1}^{L-d} \frac{1}{M} \left[ \left[ w_{j}^T \cdot (x_1 - x_2) \right]^T \cdot \left[ w_{j}^T \cdot (x_1 - x_2) \right] \right] \tag{7}
\]

where \( w_j \) is a window vector with the same dimensionality as that of \( x_1 \) or \( x_2 \), and

\[
M = \max \left\{ \left| w_j^T \cdot (x_1 - x_2) \right|, 1 \right\} \tag{8}
\]

is a variable to normalise the result of the windowed matching (see the formula in the braces of Equation (7)). To implement the elastic matching, the window \( w_j \) is designed as a row vector, where only the elements within the window are assigned as the value 1 and the remaining elements are 0.

The window position is controlled by the parameter \( j \), and the window width is decided by the parameter \( d \). According to Equation (8), the variable \( \left| w_j^T \cdot (x_1 - x_2) \right| \geq 0 \) measures the total number of the mismatching within the window, and the summation of \( \sum_{j=1}^{L-d} \) of Equation (7) is equivalent to a traversing procedure. Thus, by moving the window vector \( w_j \) across the whole spectrum, we check whether an absorption-displacement happens. Finally, \( D_h(x_1, x_2) \) counts the total number of absorption displacement occurred in the whole spectrum.

By considering the matching scores made by the absorption displacement, a revised Hamming distance is proposed as follows:

\[
D_{ah}(x_1, x_2) = D_h(x_1, x_2) - D_h(x_1, x_2) \tag{9}
\]

where \( D_{ah}(x_1, x_2) \) is the weighted Hamming distance described in Equation (5). In the revised Hamming distance if there are two absorption occurs within the window, they are considered as the same absorption and the Hamming distance of the windowed segmentation is calculated as 1 rather than 2.

Through the above elastic matching, the absorption pairs within the window can be found out and the corresponding discrepancy is subtracted from the Hamming distance. The revised Hamming distance in Equation (9) takes the absorption shifting into consideration, and therefore reduce the likelihood of mismatching.

In the elastic matching, the parameter of window width \( d \) gives the flexibility for the correspondence matching. Figure 1 shows that the different samples from the same material may have slightly different absorption bands, but they are actually originated from the same absorption property. The conventional Hamming distance cannot differentiate this, and the resulting distance will overestimate the disparity between the two spectral samples. In the revised Hamming distance, the total number of the mismatching within the window is calculated from Equation (7), and is deducted from the final Hamming distance. The parameter \( d \) is adjustable and can be changed for different application scenarios to accommodate their different levels of the spectral displacement.

In our testing, the detailed values of \( d \) is decided empirically by observing the training samples. For example, a value of 5 can be chosen by looking at Figure 1, where the maximal displacement is 5 bands.

**Experiments and results:** To assess the performance of the proposed method, experiments have been carried out based the Salinas hyperspectral data set. The data was acquired by AVIRIS sensor over Salinas Valley, California. The AVIRIS is a 224-band hyperspectral sensor with the wavelength coverage from 400 to 2500 nm and spectral resolutions between 9 and 15 nm. The Salinas scene is characterised by the higher spatial resolution, i.e. about 3.7 meter/pixel, and the higher SNR (signal-to-noise ratio). It gives a better representation of spectral absorption, and becomes an ideal candidate to verify the aforementioned absorption-based feature extraction. The studied scene covers vegetation, soils, and fields, including broccoli/green/weeds, corn/weeds, lettuce/romaine (4 weeks), lettuce/romaine (5 weeks), lettuce/romaine (6 weeks), and lettuce/romaine (7 weeks).

In remote sensing applications, it is difficult to accrue a big amount of training data due to costly human labelling and strenuous space-ground coordination. Realistic simulations should be based on a small set of training data. Therefore in the experiments, the training set is randomly
The results, not included in this letter due to the pages limit, are coherent with the aforementioned conclusion.

For the scarcity of supervised data, popular deep-learning based methods are not applied to this research. On the other hand, support vector machines (SVMs) [3], with proven performance from many few-shot learning applications, and a back propagation neural network (NN) associated with MNF (maximum noise fraction) feature extraction are considered as the benchmark classifiers. A polynomial kernel function is used as the kernel function, and its parameters are chosen as $d = 3$ and $C = 35,000$ by a two folds validation procedure using only training data.

Table 1 compares the performances of the proposed method with the neural network-based method (NN) and the SVM-based method. It can be seen that the overall accuracy of the proposed methods is 73.13%, which is competitive to that of the SVM-based method (73.06%) and is higher than the NN-based method (71.86%). The Cohen’s kappa coefficients also confirm this result (0.67 vs. 0.66 and 0.65). Although the proposed method does not get a significantly higher classification accuracy than the state of the art method, it is still encouraging because it is achieved based on a much more compact feature set (binary absorption feature vector versus real value spectra). In the cases where the requirement for data storage or data communication is stringent, this method can be found useful.

Furthermore, it is found that the individual classification accuracies of the proposed method have a more uniformed distribution. In contrary, the SVM shows a poor performance for the class lettuce_4wk, which is thought to be caused by an insufficient sampling of the training data. For this issue, the proposed method exhibits more robust performance than those sampling-relying approaches. One possible reason is that the absorption features are more stable than the radiance that are easily influenced by the sunlight and the neighbouring scattering. Similar results can be found in the classification maps (Figure 3(a,b,c)).

Further experiments are carried out on the AVIRIS 92AV3C data set. The results, not included in this letter due to the pages limit, are coherent with the aforementioned conclusion.