Combining Spectral and Texture Features in Hyperspectral Image Analysis for Plant Monitoring

DOI:
10.1088/1361-6501/aad642

Document Version
Accepted author manuscript

Link to publication record in Manchester Research Explorer

Citation for published version (APA):

Published in:
Measurement Science and Technology

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester’s Takedown Procedures [http://man.ac.uk/04Y6B0] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.
Combining Spectral and Texture Features in Hyperspectral Image Analysis for Plant Monitoring

To cite this article before publication: Ali AlSuwaidi et al 2018 Meas. Sci. Technol. in press https://doi.org/10.1088/1361-6501/aad642

Manuscript version: Accepted Manuscript

Accepted Manuscript is “the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an ‘Accepted Manuscript’ watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors”

This Accepted Manuscript is © 2018 IOP Publishing Ltd.

During the embargo period (the 12 month period from the publication of the Version of Record of this article), the Accepted Manuscript is fully protected by copyright and cannot be reused or reposted elsewhere.

As the Version of Record of this article is going to be / has been published on a subscription basis, this Accepted Manuscript is available for reuse under a CC BY-NC-ND 3.0 licence after the 12 month embargo period.

After the embargo period, everyone is permitted to use copy and redistribute this article for non-commercial purposes only, provided that they adhere to all the terms of the licence https://creativecommons.org/licences/by-nc-nd/3.0

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions will likely be required. All third party content is fully copyright protected, unless specifically stated otherwise in the figure caption in the Version of Record.

View the article online for updates and enhancements.
Combining Spectral and Texture Features in Hyperspectral Image Analysis for Plant Monitoring

Ali AlSuwaidi, Bruce Grieve, and Hujun Yin
School of Electrical and Electronic Engineering, The University of Manchester, Manchester, M13 9PL, UK
E-mail: ali.bghalsuwaidi@postgrad.manchester.ac.uk, bruce.grieve@manchester.ac.uk, h.yin@manchester.ac.uk

March 2018

Abstract. A texture enhanced spectral analysis framework is proposed for classifying hyperspectral images of plants of different conditions. Differentiating different plant conditions is important to precision agriculture as it helps detect diseases and stresses and optimise growth control. Advanced machine learning techniques are used to identify distinctive features in the spectral domain of hyperspectral images. In addition, texture properties are explored in the sub-band images. The framework integrates these two levels of properties at both feature extraction and classifying decision stages. The main crux of the work lies in the use of the significant spectral and texture features and a decision fusion mechanism to enhance the image properties, thus improving classification accuracy. Two hyperspectral datasets, originated from proximal hyperspectral systems, were used in the evaluation and significant improvements in classification accuracy achieved.

Keywords: conventional SVM, decision fusion, feature fusion, feature selection, hyperspectral imaging, one-class SVM, spectral analysis, texture analysis
1. Introduction

The number of imaging modalities has increased with the evolution of imaging technology [1]. Hyperspectral imaging is one of these modalities that has triggered a considerable amount of interest in exploring innovative research missions [2]. It can be considered to be a natural extension of multispectral imaging in terms of individual bandwidth and spectrum range. Hyperspectral imaging usually senses a wider range of the spectrum with a large number (i.e. hundreds) of detailed narrow bands. Images in this emerging imaging branch may be captured using one of the following four configurations: point, area, line, and single shot scanning [3, 4]. Hyperspectral imaging has been utilised in several applications such as remote sensing, agriculture, military and mineralogy [5].

The characteristics of spectral (i.e. function of electromagnetic spectrum) and texture (i.e. function of inter-pixel intensity) information are the fundamental properties of hyperspectral images. Each pixel is represented by a unique spectral signature in the former, containing reflectance, absorbance, or transmittance at each wavelength. Moreover, the electromagnetic coverage of this signature can cover single or several spectrum regions (e.g. visible, infrared, microwave, etc.) [6]. The resolution of the spectral information is defined by the ability of the system to distinguish between two adjacent spectral bands. On the other hand, the spatial resolution (i.e. the clarity of the image) is determined by the smallest detected pixel in the latter. Dark pixels represent low intensity values, whilst bright pixels represent high intensity values.

Hyperspectral imaging has attracted scientific interests from various disciplines and a tremendous increment in developing hyperspectral-based analysis techniques has been noticed [7]. Several image analysis approaches have been proposed in recent years, for instance spectral, texture, and spectral-texture based, and used for analysing hyperspectral images. In spectral analysis approach, several hyperspectral-based indices have been proposed and developed to study plant properties, types, and conditions for agriculture and geosensing applications. The normalised difference vegetation index (NDVI) is a popular example of these indices and it is widely used in both proximal and remote sensing to describe vegetation health as well as density. Moreover, other indices such as simple ratio, normalised difference, and modified indices have been introduced to detect certain plant health and stress levels through the content of leaf pigments, especially green pigment (i.e. chlorophyll) [8, 9]. In addition, several exiting and new indices have been introduced to identify and detect plant diseases [10, 11].

Selecting the most informative subspace of the datasets is another useful approach in spectral analysis. The selection process is a crucial step for hyperspectral systems because a large amount of data is being generated using said systems [5]. Feature selection techniques tend to be a good option for selecting relevant spectral subspaces or subsets for the investigated problem. It is “a process that chooses an optimal subset of features according to a certain criterion” [12]. Usually, independent (i.e. using data characteristics) and dependent (i.e. using learning algorithms) criteria are used to evaluate the goodness of the selected subspace [12]. Finding the best informative subspace may not only lead to a better utilisation of data storage, but also improve classification performance and reduce computational complexity [13]. Our previous study reported better classification performance on the use of a systematic spectral selection over those of using empirical indices [14].

Texture analysis is another common image analysis technique where adjacent pixels and their intensities are analysed or modelled to gain insightful information about surface properties of the image [15]. Several texture analysis techniques have been existed in the literature such as statistical, structural, transform-based, and model-based [16]. These techniques have similar aims but different methodologies. In other words, each technique has a unique form to investigate the texture, and thus determine its properties.

Spectral-texture analysis approach has been introduced to take advantage of both spectral and texture analysis approaches - i.e. using different perspectives to analyse and investigate the problem. Previously, the spectral-based approaches were only utilised without considering the texture-based ones in the analysis due to the assumption that spectral features contain enough information and the mixing issue in moderate spatial resolution images [17, 18]. Recently, the ability of various imaging systems to acquire high spatial resolution images have been improved, thus easing the mixing issue. However, increased spatial resolution may result in increased spectral variability within a class but decreased variability between classes [18, 19]. Currently, several studies highlighted the importance of incorporating spectral and texture features and the effect of the combined features on classification accuracy [17, 18].

Recently, there has been a surge of interest in feature and decision fusion to improve the classification performance of hyperspectral datasets. On the one hand, feature fusion is used to combine the best discriminative data properties from single or multiple sources, thus improving local predictive performance. Several methods can be employed to fuse features
such as simple concatenation and softmax regression [20, 21]. On the other hand, the goal of decision fusion is about improving overall classification performance through combining all local decisions under one centralised rule. Hard and soft fusion are the main examples of decision fusion. Weighted or unweighted majority voting and mean are examples of the former, whilst posterior probabilities, linear and logarithmic opinion pool are examples of soft decision fusion [22].

This work focuses on using different analysis approaches and combining them under a centralised decision to classify plant stress and disease conditions. Classification performance of the combined decisions is then compared to those using selected wavelengths, selected texture parameters, and selected wavelength-texture parameters. The texture properties of the images in this work have been estimated using Markov random field (MRF), whilst the correlation-based feature selection (CFS) was used for the selection purposes. The MRF is one of the most powerful models for texture description and has been utilised in several applications [23]. For feature selection, CFS was chosen due to its capability to eliminate redundant features, applicability over a wide range of applications, and positive ability in classification performance. The dominant features form each domain were then selected and fused for classification. The datasets studied were captured in controlled environments, i.e. in dark chamber or room with constant light source, to prevent noises and make the illumination and measurement conditions reproducible.

The main contribution of this work lies in introducing a framework that unifies the performance of several properties under one decision to improve the overall performance. This framework consists of two stages: feature and decision fusions. The first stage, including selection, is based on our previous work introduced in [2, 24], while the goal of the second stage is to fuse texture features based on selected spectral bands (selected fusion). A scheme of boosting weak spectral features from textural information is also considered in the second fusion stage. In addition, conventional and one-class support vector machines (SVMs) are used for classification for a comparison. One-class SVM is found to be more suitable for unbalanced data cases, especially when data of one class (abnormal) is missing. Although the datasets used in this study were all captured in the laboratory conditions, the methodology can be applied to field environments when suitable enclosure can be applied to hyperspectral cameras (e.g. portable systems) to reduce noise.

The remainder of this paper is organised as follows. The background is reviewed in Section 2. Section 3 presents the materials and the proposed framework, including the description of the hyperspectral systems and the datasets. The experimental results and discussions are given in Sections 4 and 5 respectively, followed by concluding remarks in Section 6.

2. Background

This section outlines the machine learning and texture analysis techniques relevant to this work. It is divided into four sub-sections, starting with a description of feature selection, an overview of texture analysis, followed by an introduction of classifiers and an explanation of feature and decision fusions.

2.1. Feature Selection

The process of retaining relevant subset of features to the problem investigated and eliminating redundant and irrelevant features is termed as “feature selection” [12, 25, 26]. Feature selection can not only reduce the dimensionality of the dataset (e.g. hyperspectral images), but also can gain better classification performance. Several feature selection algorithms have been proposed in the literature and a good number of them have been reviewed in [27]. Feature selection methods are generally classified, based on the evaluation criteria, into three models: filter, wrapper, and embedded [12, 25, 26]. The main differences of these models are: the use of a mining algorithm in subspace evaluation (in wrapper and embedded), computational time required (filter model is the shortest), and ease of implementation (e.g. wrapper and embedded are easy to implement). The primary goal of all of these models is to maximise a evaluation criterion \( J(.) \), thus defining a relevant subset of the features. This can be formulated as:

\[
J(S_i) \leq J(S_o), \forall S_i, S_o \subseteq X, i = 1, 2, ..., n
\]

where \( S_o \) denotes the optimal subset, \( S_i \) represents the \( i-th \) subset of \( n \) total subsets generated in the process of feature selection, and \( X \) denotes a specific dataset.

In this study, we employed a correlation-based feature selection algorithm (filter based) for the following reasons: 1) good ability in eliminating redundant features, 2) no predefined stopping criterion is needed, 3) applicable in a wide range of applications, and 4) easy to implement [28]. The concept of information theory (i.e. Shannon’s entropy \( H(x) \) and information gain \( I(x, y) \)) is used in this algorithm to measure the correlation instead of using simple correlation such as Pearson’s. In addition, the information-based symmetrical uncertainty (SU) is used to measure average feature-to-class \( (r_{cj}) \) and average feature-to-feature \( (r_{ff}) \) correlations in order to minimise the information gain bias introduced to
Combining Spectral and Texture Features in Hyperspectral Image Analysis for Plant Monitoring

the features [28]. Shannon’s entropy, information gain, symmetrical uncertainty, and the evaluation criterion of CFS are expressed as:

\[ H(x) = - \sum_{j=1}^{m} P(x_j) \log_2 P(x_j) \]  

(2)

\[ I(x,y) = H(x) - H(x|y) \]  

(3)

\[ SU = \frac{I(x,y)}{H(x) + H(y)} \]  

(4)

\[ Merit_s = \frac{N_{rf}}{N + (N + N(N-1))_{rf}} \]  

(5)

where \( N \) represents the total number of features and \( Merit_s \) denotes the heuristic evaluation criterion that determines the significant features, and process stops if no improvement is achieved after five consecutive iterations. Note that the entire spectral signature (of \( N \) wavelengths) is used to select a relevant subset of wavelengths.

2.2. Texture Analysis

Texture contains useful information about pixels structural arrangement and their relationship [15]. Extracting textural features from original texture image is essential for applications. Texture analysis techniques can be broadly categorised into statistical, structural, transform-based, and model-based [16]. The first technique uses the distribution of grey levels to describe the properties of texture and the co-occurrence matrix is a common example of this category. The second technique employs simple texture properties such as lines and edges to describe regular textures. The third technique transforms texture into another domain and then uses the transformed characteristics to describe the properties of the texture. 2-D Fourier transform and wavelet transform are two examples of this category. The last technique uses models and estimated parameters of the models to define texture properties. Markov random field (MRF) is a widely used example in this category.

Texture analysis has been employed in various applications such as pattern recognition and image segmentation [29, 30, 31]. Security screening, personal identification, and automated immigration control are examples, in which texture analysis is used to extract unique iris signatures. Analysis of texture has also been used as a post-processing technique to enhance the representation and thus improve classification performance.

MRF, a model-based technique, is used in this study to analyse texture. It is a natural extension of Markov chain model [32]. The corresponding intensities of neighbouring pixels are used to generate a probabilistic model. It is also regarded as an undirected graph model and it is more natural for modelling textures [33]. The structure of neighbouring pixels is used to define the order of MRF as shown in Figure 1 and it can be represented by the sum of local interactions \( V \). The local interactions of the first MRF order can be expressed as:

\[ V = \beta_0 + \beta_1(x_{i,j-1} + x_{i,j+1}) + \beta_2(x_{i-1,j} + x_{i+1,j}) \]  

(6)

where \( V \) denotes the local interaction sum and \( \{\beta_0, \beta_1, \beta_2\} \) are the MRF parameters. An MRF is said to be isotropic, if the perpendicular parameters for a specific order are the same, and anisotropic if the parameters are different. Positive parameters govern clustering of pixel intensities in a specific direction (e.g. \( \beta_1 \) in horizontal and \( \beta_2 \) in vertical directions) while the negative values of these parameters cause repulsion.

The equivalent Gibbs distribution can be used to describe the MRF mathematically [34]. Let \( P(x) \) denotes a Gibbs distribution for realisation \( x \), \( N \) represents a neighbouring system, \( \Omega \) denotes a finite lattice, and \( C \) represents all possible cliques, i.e. subset of a lattice consists of single and/or set of pixels which are neighbour to each other, then the distribution can be represented as:

\[ P(x) = \frac{1}{Z} e^{-U(x)/T} \]  

(7)

where \( T \) stands for temperature; \( U(x) \) represents the energy function that depends only on clique potential \( V_C \) on the lattice, which can be written as:

\[ U(x) = \sum_{c \in \Omega} V_C(x) \]  

(8)

\( Z \) denotes a normalising constant, also termed as partition function, and is defined as:

\[ Z = \sum_x e^{-U(x)/T} \]  

(9)

Several algorithms have been proposed in the literature to estimate the parameter of MRF, such as the maximum likelihood (ML) and the least square (LS) estimate [23]. The LS estimate is simpler and yields lower computational requirements compared to the ML, and is more preferable in practice. In addition, there is no closed form solution for the ML estimate of parameters (i.e. the assumption of an alternative function such as pseudo likelihood (MPL) is needed) and the result is not guaranteed [23, 33]. For the LS estimate, the parameters over a finite lattice \( \Omega \) can be estimated using:

\[ \hat{\beta} = \left[ \sum_{m \in \Omega} z_m z_m^T \right]^{-1} \left[ \sum_{m \in \Omega} \sum_{i,j} z_m x_{i,j} \right], m = 1, 2, \ldots, M \]  

(10)

where \( x_{i,j} \) is the centre pixel and \( z_m \) denotes the neighbouring pixels and can be represented as:

\[ z_m = \text{col}[x_{i+u,j+v}], (u,v) \neq 0, (u,v) \in N \]  

(11)
where \( u, v \) represent the location of the neighbouring pixels horizontally and vertically, respectively and \( c \) stands for column. It is worth noting that LS is not consistent for non-causal neighbour sets [23, 33].

2.3. Conventional and One-Class SVM

Support vector machine (SVM) is a non-parametric classifier for solving binary classification problems [35]. It is widely preferred due to its good generalisation ability with few training samples. The training samples are used to define the optimal linear hyperplane that maximises the decision boundary, hence minimising misclassification. Moreover, the generalisation of SVM can also be extended to non-linear cases via kernel tricks. Several kernels have been introduced such as radial basis function (RBF): \( K(x, y) = e^{-\gamma \|x - y\|^2} \), and polynomial: \( K(x, y) = ((x, y) + 1)^n, n \in \mathbb{R}^+ \) [35]. In addition, many methods have been introduced in the literature to extend SVM to multi-class problems [36]. Parallel SVM is one example, where one class is against all of the other classes. Another example is hierarchical tree-based (either balanced branches or one against all).

Conventional SVM can be expressed as the following quadratic equation:

\[
\min_{w \in \mathbb{R}^d, \xi_i \in \mathbb{R}^+} \frac{\|w\|^2}{2} + C \sum_{i=1}^{N} \xi_i
\]

subject to: \( y_i (w \cdot x_i + b) \geq 1 - \xi_i \)

where \( w \) is the weight vector, \( \xi_i \) represents a non-zero slack variable, \( y_i \in \{-1, 1\} \) denotes the class label, \( b \) is a bias, and \( C \) stands for a regularisation constant and is used in the non-separable case to penalise the samples that lie on the other class side. In this case, both the bias and weight vectors are adapted to solve the decision function \( f(x) = \text{sign}(w \cdot x + b) \) and \( \xi \in \{\pm1\} \).

On the other hand, one-class SVM is a modified version of the conventional SVM and it is used in extremely unbalanced-data cases. In other words, one class (called normal) is well sampled while the other classes (abnormal) are severely under-sampled, if not absent entirely. The lack of abnormal samples is usually due to the difficulty of obtaining them (e.g., high measurement cost) or the low frequency of the abnormal events [37]. The data description is an important domain where one-class SVM has been utilised to describe how abnormal samples are deviated from the normal samples used in the training process [38]. The model is generated to define the boundaries of the normal class. One-class SVM has been used in industrial applications [39], medical monitoring [40], and remote sensing [37] especially in high-integrity systems where abnormality raises a flag for a potential problem. One-class SVM can be expressed as:

\[
\min_{w \in \mathbb{R}^d, \xi_i \in \mathbb{R}^+} \frac{\|w\|^2}{2} + C \sum_{i=1}^{N} \xi_i - b
\]

subject to: \( (w \cdot \phi(x_i)) \geq b - \xi_i, \forall \xi_i \geq 0 \)

where \( \phi(x_i) \) is the transformed vector \( x_i \). It should be mentioned that \( C = \frac{1}{\rho} \), where \( \rho \in \{0,1\} \) and \( \rho \) is the lower bound of the support vectors, as well as the upper bound of the outliers, used to determine the flexibility of the decision boundary. Moreover, values of the decision boundary \( f(x) = \text{sign}(\omega \cdot \phi(x) - b) \) lie in \( \{\pm1\} \) where \( +1 \) denotes normal class (i.e., within the boundary) and \( -1 \) represents the abnormal. In addition, the linear boundaries of one-class SVMs can also be extended to non-linear cases through kernel tricks (i.e., transforming the data into a higher dimensional space).

2.4. Feature and Decision Fusion

Feature fusion is an active area of interest in machine learning. The process of feature fusion combines the most dominant features, either with different characteristics from a single source or multiple sources, to achieve the best possible classification accuracy locally [41]. Feature concatenation is the simplest method of feature fusion; however, the result is not guaranteed since it does not consider the differences in the feature space structures. Serial and parallel combinations are also two examples of feature fusion [42]. In the serial combination, the weighted features (of different characteristics or sources) are combined...
Combining Spectral and Texture Features in Hyperspectral Image Analysis for Plant Monitoring

6

The issue of decision fusion has received considerable attention in machine learning due to the ability of such fusion to improve global classification performance. It is the process of combining various classifiers in a weighted or unweighted sense to minimise the risk of selecting the least performing classifier [43]. Several theories on decision fusion have been proposed and we have simplified them into two categories; hard and soft fusions. In the former, the central decision of all of the classifiers can be achieved using weighted or unweighted versions of majority voting or averaging. However, the overall classification decision in the latter can be obtained using the posterior probabilities of each classifier without relying on their labels [22]. Note that the unweighted majority voting of the hard fusion, employed in our experimental procedure, can be expressed as:

\[
\hat{y} = \max_{y} \sum_{r} y_i
\]

where \(\hat{y}, y, r\) represent the fused class label, class label, and the total number of classifiers, respectively.

Several feature and decision fusion methodologies have been proposed in the literature to improve the classification accuracy of the recognition system locally and globally [21, 44]. For example, a softmax regression-based technique, called marginalized kernel, has been proposed to fuse spectral, texture, and local-self similarity descriptors of the QuickBird imagery linearly. The fusion process is done after estimating object-to-class and class-to-class similarities and then the fused features are used for classification. Each feature vector is encoded using a bag-of-visual-words method to unify all feature spaces (no normalisation is required) [21]. On the other hand, the decisions of Gabor features, gradient binary patterns, grey level co-occurrence matrix, and colour features are merged hierarchically to classify outdoor scenes such as forest, coast, mountain, and open country scenes [44].

In this work, both spectral and textural features were assessed individually, due to their differences in physical meanings and statistical properties, to select the distinctive features. Moreover, one-class classifier was used to extend the analysis of the framework to unbalanced situations - i.e. using only one class for training. This extension has been considered because not all abnormal samples (e.g. all diseases or stress types) are available in the real-world cases. For the fusion cases, the selected spectral and textural features were combined with and without considering the superiority of one feature over another. The former case shows better discrimination ability compared to the latter.

3. Materials and Proposed Methods

The materials and methods relevant to this work are presented in this section. First, the specifications of the HSI systems are provided. Second, the datasets are described. Last, the proposed framework is described and its schematic presented.

3.1. Hyperspectral Imaging Systems

Two HSI systems were used to capture the images: one is located at the University of Manchester, and the other at the Bonn University. For simplicity, we termed them UoM and Bonn systems, respectively. The former is based on a low-noise Peltier-cooled digital system (Hamamatsu, model C4742-95-1ER, Hamamatsu Photonics K.K., Hamamatsu, Japan). It generates effective pixels of 1024×1344. The spectrum range is 400 - 720 nm with spectral resolution of 10 nm, resulting in 33 narrow spectral bands. Moreover, a fast liquid crystal tunable filter is used to control the spectral transmission electronically and an infrared blocking filter is mounted in front of the lens to prevent leakage. Further information on this system can be found in [45].

On the other hand, the Bonn system is a line scanning system that operates over a visible and near infrared regions (400 to 1000 nm) with a spectral and spatial resolution of 2.8 nm and 0.29 mm respectively. Furthermore, this system is surrounded by six lamps (provided by the Analytical Spectral Devices Inc.) to provide homogenous illumination. More information about the orientation and pre-processing can be found in [11, 46]. Note that both systems have radiometric resolution of 12-bits, spectrally normalised using white reference, and are operated in a controlled environment to prevent noise.

3.2. Datasets

Two HSI datasets were considered in our experiment: UoM and Bonn datasets, captured in controlled environment (dark room or chamber) by the UoM and Bonn systems, respectively [46, 47]. A powerful and constant light source was used in the UoM case, while six lamps were surrounding the Bonn system to lighten the sample plate. The leaf samples in both datasets were flattened on the sample plate to maximise the reflectance and then the scene images were captured. The flat field for white reference surface and dark noise images (e.g. blocking the camera by a cloth or using...
Combining Spectral and Texture Features in Hyperspectral Image Analysis for Plant Monitoring

The unhealthy condition was further classified into Cercospora and Rust diseases. Both diseases were obtained from infected sugar leaves [46]. Some of the healthy sugar plants were then sprayed (using hand sprayer) by these diseases and covered by plastic bags for 48 hours to ensure diseases inoculation. The Leaves in the Bonn dataset were captured on five different days (days 7, 9, 11, 13, and 15) [48]. The unhealthy group was represented by 320 samples, while the healthy group was represented by 160 samples.

3.3. Proposed Framework and Extensions

The proposed classification framework can be described in two stages: feature fusion and decision fusion. The feature fusion stage is implemented according to our procedure used in [2, 24]. It consists of five steps: extraction of spectral signature, estimation of texture parameters, selection of significant features (spectral and textural), serial fusion (i.e. simple concatenation), and classification. The spectral information and the parameters of texture are then fused based on spectral relevance (i.e. dependent approach), selecting textural parameters based on spectral relevance, and based on relevance of the texture itself (i.e. independent approach). It should be mentioned that equations (2) to (4) are used to measure average feature-to-feature and average feature-to-class correlation values, and then these values are used (in equation (5)) to obtain the optimal subsets of wavelengths and texture parameters. Moreover, the selected texture parameters are used 'as is' and as mean over the selected wavelengths. In the decision fusion stage, all of the merged versions of the spectral-texture information (four in this case) are fused using unweighted majority voting (Equation (14)), thus forming a centralised decision of the classification (in other words, the outputs of the schematics in [2, 24] are combined).

Two extensions (selected fusion [2, 24] and a new one termed boosting fusion) have been made to the core of our proposed framework to enhance the overall classification performance. In the selected fusion, the unweighted majority voting is used to fuse spectral and texture information and then another voting stage is used to form the final decision. This extension has been made for two reasons: 1) spectral and texture properties have different structure, thus serial fusion might not be the best, and 2) relevancy has to be considered individually, as it might be not related.

On the other hand, the boosting fusion, shown in Figure 3, is proposed due to the superiority of spectral information over texture parameters. The concept of boosting has been adopted in this extension. Boosting focuses on misclassified samples, updates their weightings, and combines them to generates a strong centralised classification rule [49]. In this extension, misclassified samples of the selected wavelength are passed on to the feature-texture fusion stage, in which the selected wavelengths and texture parameters (normal and averaged versions) are fused through majority voting. This step is employed in this extension since previous studies conform the superiority of the spectral features over other features.
Combining Spectral and Texture Features in Hyperspectral Image Analysis for Plant Monitoring

4. Experimental Results

The experiments assessed the usefulness of the proposed framework and its extensions for analysing and classifying plant hyperspectral datasets under stressed and diseased conditions. The extended version of one-class SVM, proposed in [39], was used to generate the result of unbalanced situations. The decision function of the one-class SVM was further calibrated to class probability using non-decreasing (isotonic) regression. The calibrated result was then used to determine the parameters of the SVM as well as the threshold that distinguishes normal from abnormal samples. Further details on calibration can be found in [39] and the result can be expressed as a novelty score

\[ z(x) = b - w \Phi(x_i) \]  

The UoM and Bonn datasets were used in all of the experiments. In the conventional classifier case, 50% of the samples were used for training with a 10-fold cross validation and the remaining samples were used for testing. 60% of the normal samples were used for training and the remaining 40% and all of the abnormal samples were used for testing and validation in the unbalanced situation where the samples of controlled and healthy conditions considered as normal. The classification performance was represented as the average of 100 iterations.

4.1. Feature Fusion vs. Decision Fusion

Several texture parameters (estimated by the MRF model) were used for both of the classifiers and up to the third order was reported (high orders did not improve the performance). Tables 1 and 2 illustrate the averaged classification rate of 100 runs for both datasets.

What stands out in the tables is that spectral information is more dominant in the classification than texture properties. The average classification rate of any fused version is often better than individual ones, especially the texture-based, in both datasets. Furthermore, fusing the local decisions via majority voting (last case in Table 1) enhances the classification accuracy of the UoM dataset for the first order only, while no improvement for the Bonn dataset. This might be due to the sparse nature of the estimated texture parameters and their dimensionality (2 \times \text{order} \times \text{number of wavelengths}). Although the prediction performances of the conventional and one-class classifiers are almost the same, the latter setting is more applicable when only normal samples are available. Fusing a number of local decisions has been found better than fusing all decisions in the literature [50]. A few of manually selected decisions was used to improve the classification accuracy and the results are displayed in Table 3. Note that only the first order of the texture parameters were used to generate the result.

Closer inspection of Table 3 shows that the classification performance of fusing a few decisions was almost the same as fusing all decisions in the UoM case, while 5% improvement in the Bonn case. Fusing a few decisions tended to be more robust as shown in the Bonn case. Further analysis revealed that this was applicable for estimated parameters from first order MRF only. In addition, different combinations led to different performance. Finding a systematic way to combine the decisions may lead to better performance.

4.2. Selected Fusion

The decisions on selected spectral signatures and selected estimated texture parameters were fused via majority voting, resulting in four spectral-texture decisions shown in Table 4 (cases 1 - 4). These four decisions were then merged to form final decision (the case combined in the table), thus generating an overall classification rate. The estimated parameters from the first and the second order MRF from each dataset were used with conventional SVM only. The results indicate that the first voting stage outperformed the second one (combined).

Inconsistency in the classification accuracy (increase/ decrease) of feature fusion has been observed in table. Furthermore, increasing the order of the parameters led to better discrimination in the second majority voting stage. Further analysis (not reported here) revealed that further increase in the order led to a negative effect on the first majority voting stage and a positive one on the second stage. In addition, the classification accuracy of a few decisions (e.g. merging number 1 and 2) outperformed the overall accuracy. In contrast, no improvement in classification was observed compared to serial fusion with majority voting. This might be due to the sparsity nature of the estimated texture parameters.

4.3. Boosting Fusion

The selected spectral signatures from both datasets were used in this experiment with both conventional and one-class SVM in the boosting fusion. The misclassified samples were then passed on to the
Combining Spectral and Texture Features in Hyperspectral Image Analysis for Plant Monitoring

Figure 3. Schematic of boosting extension. The decision of misclassified samples (i.e spectrally selected) is fused via majority voting with the decisions of the selected texture (all versions) using the index of the misclassified samples. The label of misclassified samples are then replaced (through the same index) by the label of the fused decisions to generate the overall classification accuracy.

Table 1. Average classification rates of selected wavelengths (case No. 1), all selected texture parameters (cases No. 2 - 5), fused spectral-texture (cases No. 6 - 9), and fusion of cases 6 - 9 (case No. 10) on UoM using both SVM settings.

<table>
<thead>
<tr>
<th>No.</th>
<th>Approach</th>
<th>Conventional SVM</th>
<th>One-class SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1st Order</td>
<td>2nd Order</td>
</tr>
<tr>
<td>1</td>
<td>Sel. Wavelengths</td>
<td>91.68 (0.011)</td>
<td><strong>91.73 (0.011)</strong></td>
</tr>
<tr>
<td>2</td>
<td>Sel. Texture (dependent)</td>
<td>75.61 (0.016)</td>
<td>82.27 (0.018)</td>
</tr>
<tr>
<td>3</td>
<td>Sel. Texture (independent)</td>
<td>79.92 (0.018)</td>
<td>79.25 (0.016)</td>
</tr>
<tr>
<td>4</td>
<td>Avg. Sel. Texture (dependent)</td>
<td>52.74 (0.025)</td>
<td>72.67 (0.018)</td>
</tr>
<tr>
<td>5</td>
<td>Avg. Sel. Texture (independent)</td>
<td>54.98 (0.025)</td>
<td>72.29 (0.020)</td>
</tr>
<tr>
<td>6</td>
<td>Fused 1 and 2</td>
<td>92.07 (0.012)</td>
<td>78.39 (0.017)</td>
</tr>
<tr>
<td>7</td>
<td>Fused 1 and 3</td>
<td>91.45 (0.016)</td>
<td>71.82 (0.011)</td>
</tr>
<tr>
<td>8</td>
<td>Fused 1 and 4</td>
<td>91.44 (0.013)</td>
<td>91.10 (0.016)</td>
</tr>
<tr>
<td>9</td>
<td>Fused 1 and 5</td>
<td>91.81 (0.011)</td>
<td>90.23 (0.016)</td>
</tr>
<tr>
<td>10</td>
<td>MV 6 to 9</td>
<td><strong>93.90 (0.011)</strong></td>
<td>78.77 (0.017)</td>
</tr>
</tbody>
</table>

Sel.: selected, Avg.: Averaged, and MV: Majority Voting.

Several texture parameters that were estimated form the MRF model were used in this experiment and Table 5 shows the averaged classification rate of 100 runs for both datasets.

Several studies have shown the potential of
Table 2. Average classification rates of selected wavelengths (case No. 1), all selected texture parameters (cases No. 2 - 5), fused spectral-texture (cases No. 6 - 9), and fusion of cases 6 - 9 (case No. 10) on Bonn using both SVM settings.

<table>
<thead>
<tr>
<th>No.</th>
<th>Approach</th>
<th>Conventional SVM</th>
<th>One-class SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1&lt;sup&gt;st&lt;/sup&gt; Order</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt; Order</td>
</tr>
<tr>
<td>1</td>
<td>Sel. Wavelengths</td>
<td>92.55 (0.013)</td>
<td>92.70 (0.015)</td>
</tr>
<tr>
<td>2</td>
<td>Sel. Texture (dependent)</td>
<td>66.95 (0.022)</td>
<td>66.67 (0.000)</td>
</tr>
<tr>
<td>3</td>
<td>Sel. Texture (independent)</td>
<td>66.48 (0.027)</td>
<td>66.53 (0.004)</td>
</tr>
<tr>
<td>4</td>
<td>Avg. Sel. Texture (dependent)</td>
<td>62.69 (0.052)</td>
<td>59.55 (0.033)</td>
</tr>
<tr>
<td>5</td>
<td>Avg. Sel. Texture (independent)</td>
<td>58.63 (0.064)</td>
<td>63.36 (0.030)</td>
</tr>
<tr>
<td>6</td>
<td>Fused 1 and 2</td>
<td>82.05 (0.021)</td>
<td>66.86 (0.002)</td>
</tr>
<tr>
<td>7</td>
<td>Fused 1 and 3</td>
<td>90.31 (0.017)</td>
<td>68.58 (0.007)</td>
</tr>
<tr>
<td>8</td>
<td>Fused 1 and 4</td>
<td>95.77 (0.010)</td>
<td>94.74 (0.012)</td>
</tr>
<tr>
<td>9</td>
<td>Fused 1 and 5</td>
<td>95.77 (0.009)</td>
<td>94.74 (0.012)</td>
</tr>
<tr>
<td>10</td>
<td>MV 6 to 9</td>
<td>90.85 (0.016)</td>
<td>67.85 (0.009)</td>
</tr>
</tbody>
</table>

Sel.: selected, Avg.: Averaged, and MV: Majority Voting.

Table 3. Average classification rates of selected decisions (many vs. all) on UoM and Bonn datasets using conventional SVM.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Average Classification Rate % (std)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UoM</td>
</tr>
<tr>
<td>MV 6 and 7</td>
<td>91.13 (0.012)</td>
</tr>
<tr>
<td>MV 6 and 8</td>
<td>92.32 (0.014)</td>
</tr>
<tr>
<td>MV 6 and 9</td>
<td>93.12 (0.014)</td>
</tr>
<tr>
<td>MV 7 and 8</td>
<td>92.02 (0.014)</td>
</tr>
<tr>
<td>MV 7 and 9</td>
<td>91.74 (0.016)</td>
</tr>
<tr>
<td>MV 8 and 9</td>
<td>92.44 (0.014)</td>
</tr>
</tbody>
</table>

MV: Majority Voting.

boosting, feature and decision fusions. What stands out in this table is that the improvement of the boosting fusion is almost 1% to 2% for the conventional SVM and 1% to 5% for the one-class SVM compared to the best possible performance of our previous framework. Moreover, this extension tended to markedly improve the overall classification rate over individual decisions. Furthermore, the classification performances of both classifiers were almost the same for the Bonn dataset but slightly different for the UoM dataset. In addition, a consistent improvement in classification rate has been noticed (apart from the first order in the UoM dataset) and this highlight the advantage of the boosting fusion.

5. Discussions

Fusion among spectral information and among texture information have been found useful. These have been also reported in the literature [2, 24, 42, 51]. The current study finds that spectral information is more dominant than texture parameters as shown in Tables 1 and 2. This may be due to the sparsity nature of the estimated parameters of texture that were generated using the LS estimate. This has highlighted the usefulness of fused features (spectral and texture) compared to using individual properties. As mentioned in the literature review, feature fusion tends to improve classification performance [42] and the experimental results of the current study agree with the previous studies.

In this study, the proposed framework with the boosting fusion (i.e. fusing misclassified samples only) has been noticed to improve classification accuracy compared to feature selection and other fusion schemes such as selected wavelengths, selected texture parameters, and fused spectral-textures properties. Moreover, the use of various decision combination schemes in this study indicates the advantage/disadvantage of such schemes compared with those using single decisions. In other words, careful selection of a decision fusion scheme will lead to better classification accuracy and thus outperform the results of individual decisions.

Another finding was the usefulness of one-class classifier compared to conventional classifiers (illustrated in Tables 1, 2, and 5). This finding is consistent with our previous work [2, 47]. The most interesting finding is the ability of the one-class SVM to detect stress and disease conditions using normal samples only. Moreover, the calibrated output of the one-class SVM in this framework led to better parameter utilisation compared to uncalibrated ones. It should be noted that both labels (normal and abnormal) were available in our study (i.e. laboratory environment) while this may not be the case if the experiment was conducted in the field. In addition, only two disease types were considered in our study, whilst several disease types may exist in real cases.

Future work will further explore different features...
Table 4. Average classification rates of the selected extension on UoM and Bonn datasets using conventional SVM.

<table>
<thead>
<tr>
<th>No.</th>
<th>Approach</th>
<th>UoM Dataset</th>
<th>Bonn Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1st Order</td>
<td>2nd Order</td>
</tr>
<tr>
<td>1</td>
<td>MV 1 and 2&lt;sup&gt;a&lt;/sup&gt;</td>
<td>71.45 (0.020)</td>
<td><strong>76.26 (0.020)</strong></td>
</tr>
<tr>
<td>2</td>
<td>MV 1 and 3&lt;sup&gt;a&lt;/sup&gt;</td>
<td><strong>75.25 (0.021)</strong></td>
<td>72.36 (0.019)</td>
</tr>
<tr>
<td>3</td>
<td>MV 1 and 4&lt;sup&gt;a&lt;/sup&gt;</td>
<td>49.42 (0.024)</td>
<td>66.72 (0.018)</td>
</tr>
<tr>
<td>4</td>
<td>MV 1 and 5&lt;sup&gt;a&lt;/sup&gt;</td>
<td>52.39 (0.026)</td>
<td>66.69 (0.019)</td>
</tr>
<tr>
<td>5</td>
<td>Combined&lt;sup&gt;b&lt;/sup&gt;</td>
<td>48.73 (0.024)</td>
<td>55.10 (0.021)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Based on Table 1 numbering and <sup>b</sup>Based on Table 4 numbering. MV: Majority Voting.

Table 5. Average classification rates of the boosting extension on both datasets using both SVM settings.

<table>
<thead>
<tr>
<th>order</th>
<th>Conventional SVM</th>
<th>One-class SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UoM</td>
<td>Bonn</td>
</tr>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>94.61 (0.010)</td>
<td>97.03 (0.012)</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>97.08 (0.007)</td>
<td>97.05 (0.011)</td>
</tr>
<tr>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
<td>97.14 (0.008)</td>
<td>97.23 (0.010)</td>
</tr>
<tr>
<td>4&lt;sup&gt;th&lt;/sup&gt;</td>
<td>97.06 (0.009)</td>
<td>97.12 (0.012)</td>
</tr>
<tr>
<td>5&lt;sup&gt;th&lt;/sup&gt;</td>
<td>97.56 (0.008)</td>
<td>98.15 (0.008)</td>
</tr>
<tr>
<td>6&lt;sup&gt;th&lt;/sup&gt;</td>
<td>97.96 (0.006)</td>
<td>98.84 (0.007)</td>
</tr>
<tr>
<td>7&lt;sup&gt;th&lt;/sup&gt;</td>
<td>98.07 (0.007)</td>
<td>98.55 (0.008)</td>
</tr>
<tr>
<td>8&lt;sup&gt;th&lt;/sup&gt;</td>
<td><strong>98.16 (0.007)</strong></td>
<td><strong>98.89 (0.007)</strong></td>
</tr>
</tbody>
</table>

6. Conclusions

This paper presents a feature and decision fusion framework for classifying hyperspectral images of plants under stressed and diseased conditions. Two extensions have been used in the proposed framework with the primary goal of evaluating their classification performance, and consistency in relation to available datasets. The findings indicate the following. (1) It is significant to fuse spectral information and texture properties in studying plant types and conditions. This will not only help detect plant conditions but also help implement new strategies to reduce effect of these conditions. (2) It is beneficial to use one-class classifier when only data of certain conditions are available. (3) Decision fusion helps improve overall classification accuracy. Employing such fusion will prevent selecting the least performing classifier (i.e. unstable decision). (4) Feature-decision fusion has good impact on local and global classification performance. Using feature and decision fusion techniques tends to be more powerful than individual decision or classifier. (5) Fusing spectral features of misclassified samples with their textural features has certain influence on classification accuracy. Overall, the current study highlights the importance of the proposed framework in studying plant types and conditions.

Acknowledgments

The authors would like to thank G. Johnson and H. Herrmann for the provision of Arabidopsis plants and K. Amano for helping the data collection with the UoM HSI system. They are also grateful to A.-K. Mahlein for the provision of the Bonn dataset.

References

Combining Spectral and Texture Features in Hyperspectral Image Analysis for Plant Monitoring


[34] S. Geman and D. Geman, “Stochastic relaxation, gibbs
distributions, and the bayesian restoration of images,”
IEEE Transactions on Pattern Analysis and Machine
classification: A survey of support-vector-machine-based
advanced techniques,” IEEE Geoscience and Remote
remote sensing images with support vector machines,” IEEE Transactions on Geoscience and Remote
[37] F. de Morsier, D. Tuia, M. Borgeaud, V. Gass, and J. P.
Thiran, “Semi-supervised novelty detection using SVM
entire solution path,” IEEE Transactions on Geoscience
and Remote Sensing, vol. 51, no. 4, pp. 1939 – 1950,
April 2013.
[38] L. Hamel, “Novelty detection,” in Knowledge Discovery
with Support Vector Machines. John Wiley and Sons,
[39] L. Clifton, D. A. Clifton, Y. Zhang, P. Watkinson,
[40] M. A. F. Pimentel, D. A. Clifton, L. Clifton, and
L. Tarassenko, “A review of novelty detection,” Signal
[41] P. Ghamisi, J. Plaza, Y. Chen, J. Li, and A. J. Plaza,
“Advanced spectral classifiers for hyperspectral images:
[42] U. G. Mangai, S. Samanta, S. Das, and P. R. Chowdhury,
“A survey of decision fusion and feature fusion strategies for pattern classification,” IETE Technical Review,
341349, 2006.
[47] A. AliSuwaidi, B. Grieve, and H. Yin, “Feature-ensemble-
based novelty detection for analyzing plant hyperspectral
datasets,” IEEE Journal of Selected Topics in Applied
Earth Observations and Remote Sensing, vol. 11, no. 4,
pp. 1041–1055, April 2018.
H. Goldbach, and B. Grieve, “Supplemental blue LED lighting array to improve the signal quality in hyperspectral imaging of plants,” Sensors, vol. 15, no. 6,
[49] T. M. Cover and J. A. Thomas, Elements of Information
Theory (Wiley Series in Telecommunications and Signal