Survey of Learning based Hyperspectral Image Classification in Remote Sensing Applications

M. K. Pradhan, Assistant Professor (Computer Science), Deptt. of Agricultural Statistics, IGKV, Raipur, India
S. Minz, Professor (Computer Science), School of Computer and Systems Sciences, JNU, New Delhi, India

ABSTRACT

Classification and segmentation are the major stages in the remote sensing applications to identify the relevant objects from the images captured by the acquisition system. Hyperspectral Image (HSI) classification is the prominent research area to provide the information for land cover and isolate the objects from the background. The inclusion of multiple features from the multiple algorithms improves the classification accuracy. This paper reviews the traditional research studies on the HSI classification. This paper organizes the survey into five groups namely; sparse-representation, kernel, active learning, collaborative representation and the Support vector Machine (SVM) boosted learning models. The brief literature review in this paper highlights the processes involved in each model, its advantages in detail. Dimensionality, low-quality background, computational time, representation errors are the major limiting factors in HSI classification. This paper presents the clear description of the mathematical models and associated parameters in each model. The survey addresses the issues in the sparse model generation with the noise effects, parameter selection, and weight adjustments. The comprehensive review highlights that the multi-label segmentation problem induces the difficulties in the image clustering with various spectral limitations. An adaptive change of rule with the increase in data dimension is the major issue in the HSI classification.

Keywords

Active Learning, Classification, Hyperspectral Image, Joint Representation, Kernel-based Models, Learning Strategies, Sparse-Spectral Representation

I. INTRODUCTION

Hyperspectral Images (HSI) allow the deep analysis of object of interest and keep the inventories effectively. Numerous advancements in the signal processing and the exploitation algorithms improve the spectral resolution. The major challenges in the HSI classification are increased in spatial and spectral variability require compromise in classification accuracy. High dimensionality, noise occurrence and the trade-off between the various parameters (acquisition time, dynamic range, size, weight, power, and cost) are the major challenging issues in HSI classification. Traditionally, sparse models and Compressive Sensing (CS) framework addresses these issues and enable the accurate target detection. The application of CS theory on the surface is limited by the various factors such as quantization method, noise, background signal effects and the calibration errors. Hence, the researchers shift into the boundaries of knowledge collection regarding the signal structure. Statistical Learning Theory (SLT) is the major driving factor to handle the problems imposed in remote sensing applications. The embedding of numerical evaluations on the non-linear classifiers with the prior knowledge is the basic idea of the SLT. Besides, the integration of spatially-based and the manifold-based regularization allow the SLT to generate the sparse models for relevant feature sub-space. Representation of smooth land cover using the spectral information within class variability requires the spatial filtering with compromise between the user's preferences and the knowledge of the problem[1-5].

During the dimensionality reduction phase, the selection of feature vectors and the arbitrary features leads to the discarding of discriminative information. Sub-optimal approaches based on the spatial filtering rely on two problems such as restriction of filter parameters and the integration of thousands of spatial filters together. Hence, the new models are required to address these types of problems. An extraction of noticeable objects with limited resources usage and the background ignorance addresses the sub-optimal problems and such extraction refers saliency detection. The balance between the spectral count and the classification performance is the major requirement in the HSI classification. A good trade-off depends on the salient bands extracted from the HSI. The absence of specific definition about the salient objects, some methods are not extended. Besides, the inappropriate band difference measurement in saliency detection is the major problem in HSI classification. The exploitation of manifold with the Semi-Supervised Learning (SSL) methods considered as the major process to project the data into the reduced feature space. The absence of consideration of relationship between the clusters and classes in SSL algorithms and the independence of data cluster features by clustering considered the major problems in the SSL algorithms[6-12].

The incorporation of spatial and spectral information through the kernel models improve the classification accuracy. Composite kernels and the Multiple Kernel Learning (MKL) on the basis of Support vector Machine (SVM) classifiers require the convex combinations that induce the limitations. Bounding of kernels by the convex combinations and the optimization of learning parameters are the major issues in the MKL approaches. The generalized composite kernels are considered as the alternative to the MKL and they are dependent on the weight parameters adjustment for spatial-spectral information integration ratio. The built of the flexible model requires the suitable kernel parameters and their selection is the critical problem. The rapid increase of repetitive training and texting steps consume increased the operational cost. Prediction of optimal weights for biasing kernels, F-norm error, high-redundancy and the prediction of initial basis kernels are the major criteria for the kernel-based functions[13-16].

The most powerful tools in HSI classification are the kernel-based models in which the data from the original input space is mapped into the high dimensional space. Image pixel representation is done by the linear combination of various regularization norms such as l_1 and l_0. Sparse approximation and the direct assignment of the test sample to the class generates the representation errors in the kernel-based models. To alleviate this, the collaborative representation is evolved and that utilizes all the training data simultaneously. The performance of kernel-based classifiers depends on the quantity and quality of the training data. The expensive and time-consuming data labelling process produced the small set of the labelled data [17-20]. The unique identification of candidate pixels and requesting the labels require the Active Learning (AL) that picks the samples from the unlabeled data for gain maximization. Divergence, Margin Sampling (MS) and entropy-based query search are the major variations in the AL approaches[21-26]. The proper AL should satisfy the following features: minimum training samples, quick convergence of learning process and useful informative samples prediction[21-26].

Representation methods illustrated above consider the sparse patterns corresponding to the coefficients of input signals. An integration of various features like sparse, spatial and texture in the joint representation model is the major limiting factor. The representation with the norm-based measures consumes more number of iterations that leads to time complexity. To reduce that, the Simultaneous Orthogonal Matching Pursuit (SOMP) is introduced in the research to share the representation coefficients with the features. But, practically the sharing is not possible for unequal dimensions and the less similarity[27-31]. The utilization of discriminant between the important and less important samples rather than the distance measure in SVM formulation[32-35] induces the discrimination problems. For that, the explicit utilization of spatial information is the major requirement in the active learning process. The brief view of introductory section states that the numerous methods available for HSI classification.

The paper is organized as follows. Section 2 describes the overview of major categories of learning models-based HSI classification with table. Section 3 discusses conclusion.

II. LEARNING-BASED HYPERSPECTRAL IMAGE CLASSIFICATION TECHNIQUES

Smooth land cover description with the spectral information utilization within the class variability is the major issue in the hyperspectral image classification. Dimensionality and the low-quality images are the major limiting factors of the HSI classification. Several methods are available based on filters, transform, ensemble classifiers, sparse representation to perform the HSI classification. This paper presents the detailed survey of these methods with the mathematical formulations and highlights the major issues in those methods in five sections as follows:

1. Sparse representation
2. Kernel-based models
3. Active Learning (AL) models
4. Collaborative representation
5. SVM-boosted models

A. *Sparse Representation*

The improvement of classifier performance depends on the exploration of various aspects such as linear regression formulation and sparsity models. The selection of active atoms from the whole variables is represented by the regression formulation refers sparsity. The spatial contextual information is the necessary parameter in the HSI classification and the development of multinominal regression formulation addresses this issue and encodes the spatial information effectively. The 3-Dimensional Discrete Wavelet Transform (3D-DWT) based feature descriptors and the structured sparse logistic regression modeling[29] are the two major stages in the sparse representation. In general, the 1D-wavelet transform is defined by

$$(W_{\psi f})(a,b) = \left(f(x), \psi_{a,b}(x)\right) = \int f(x)\psi_{a,b}(x)dx \quad (1)$$

Where, $\psi_{a,b}(x) = |a|^{-1/2}\psi_{\left(x-\frac{b}{a}\right)}$

a-frequency region.

b – Time of the signal

For multi-scale analysis, the input function is recovered from the combination of wavelet and the scale functions formulates the feature space. The application of 3D-DWT model on the hyperspectral cube supports the spectral-spatial signatures thoroughly. Next stage like sub-space creation is performed by the linear sparse regression formulations. The selection of relevant features through the regression formulation improves the classification performance effectively. Sparse group lasso provides the sparsity among the group and individual feature levels that facilitate the relevant feature selection. Logistic regression models with the conditional probability address the classification issues as follows:

$$\log\left(\frac{p_w\left(\frac{y_i}{x_i}\right)}{1-p_w\left(\frac{y_i}{x_i}\right)}\right) = x_i w \quad (2)$$

Where. $p_w\left(\frac{y_i}{x_i}\right)$ – Conditional probability between input (x_i) and output variables (y_i)

w – Sparse coefficient

High sparse correlation existence between the pixels lie in the small neighborhood requires the Joint Sparsity Modeling (JSM)[27]. The projection of data into the feature space is the major stage for classification. The kernel function is used to project the data into the feature space and it is represented as

$$k(x_i, x_j) = (\varphi(x_i), \varphi(x_j)) \quad (3)$$

The kernel-based sparse representation of input samples extend the pixel-wise sparsity in feature space. The relationship between the pixels centered on the neighborhood (T) and the sparse vectors is represented through the sparisty matrix (S) formulation as

$$X = [x_1, x_2, \dots, x_T] = [A\alpha_1, A\alpha_2, \dots, A\alpha_T] = AS \quad (4)$$

Where, $S = [\alpha_1, \alpha_2, \dots, \alpha_T]$

The extension of sparse vectors into feature space is represented by

$$X_\varphi = [\varphi(x_1), \varphi(x_2), \dots, \varphi(x_T)]$$

$$= [A_\varphi\alpha_1, A_\varphi\alpha_2, \dots, A_\varphi\alpha_T] = A_\varphi S' \quad (5)$$

With the formulations defined from (1) through (5), the OMP, Multi-Scale Adaptive Sparse Representation (MASR)[28] and the Class level Joint Sparse Representation (JSRC)[30] are evolved in research to handle the classification issues. The built of robust classifiers with the unlabeled samples is the major limitation in sparse models.

B. *Kernel-based models*

The format of the kernel and the parameters govern the kernel function are used to control the learning performance of the classifiers. The standard SVM classifier employs the single kernel that selects the fixed kernel function and parameters optimally. The mining of information from the high dimensional data requires Multi-Kernel Framework (MKF) with the convex combination. The mathematical formulation of MKF is the extension of equation (3) as follows:

$$k(x_i, x_j) = \sum_{m=1}^{M} d_m k_m (x_i, x_j) \quad (6)$$

Subject to the constraints of

$$d_m \geq 0, and \sum_{m=1}^{M} d_m = 1$$

Where, M- Number of candidate basis kernels

d_m – weight for the m^{th} kernel

The adoption of Gaussian kernels as the candidate basis kernels in MKL depicts the discovery of relevant information. The Gaussian kernel formulation is expressed as

$$k(x_i, x_j) = \exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2}) \quad (7)$$

Where, σ – Bandwidth

The kernel distance measure and bandwidth are directly proportional to each other. In MKL framework, the relative weights are fixed. The generalized composite kernel formation [15] modifies the MKL formulation with the consideration of cross information of spectral-spatial as follows:

$$k(x_i, x_j) = [K^{\omega}(x_i^{\omega}, x_j^{\omega}), K^{s}(x_i^{s}, x_j^{s}), K^{\omega s}(x_i^{\omega}, x_j^{s}), K^{s\omega}(x_i^{s}, x_j^{\omega})]^T \quad (8)$$

Where, $K^{\omega s}(x_i^{\omega}, x_j^{s})$ and $K^{s\omega}(x_i^{s}, x_j^{\omega})$ are the cross-information kernels

The provision of balance between the spectral and spatial information equally helps to create the composite kernel effectively. The trade-off between the computational and classification accuracy is the next limiting factor for classification problem. The selection of crucial kernels based on the optimized weight [14] is the extension of MKL framework. The incorporation of contextual knowledge of local regions improves the classification accuracy by using the Joint Sparse Representation-based Classification (JSRC)[30]. Consider the unlabeled hyperspectral pixelx_i, the different types of features extracted on different perspectives. The assumption to perform the JSRC is the position of nonzero coefficients are same and the representation coefficients are not identical. With these assumptions, the additional information is preserved for classification. The mathematical formulation of JSCR with the assumptions is expressed as

$$A = \arg\min \sum_{s=1}^{S} w^s \|x^s - D^s \alpha^s\|_2$$

Subject to the constraint $\|A\|_{row,0} \leq k$ (9)

Where, $A = [\alpha^1, \alpha^2, \dots, \alpha^S]$ =Feature vectors

w^s – Weight assigned to each vector

The label is determined after the balance between the similarity and diversity of the different features as follows:

$$label(x) = \arg\min A \quad (10)$$

Based on the semantic information, the lack of bridging between the features limit the learning ability and development of parallel computing. The creating effort of training sets, regressive formulations-based classification was complex that leads to generation of active learning methods.

C. *Active Learning models*

The process of the developed program to take the control on the inputs and training set to produce more discriminative results than the random sampling refers active learning[24]. Consider the set of training samples (m) and the corresponding region of uncertainty is defined as $R(S^m)$. The iterative process of selecting most informative samples from the huge dimensionality dataset is the major role in active learning. The assumption for the iterative AL process[23] is that the classifier must learns the conditional ability in terms of probabilistic function called $P(Y|x)$. The expected error between the initial labelled pool and the large size unlabeled pool is defined as

$$E(P) = \int L\left(P(Y|x), \hat{P}(Y|x)\right) P(x) dx \quad (11)$$

Where, L –loss function

The entropy formulation minimizes the expected error through the greedy algorithm and this facilitates the better objects pre-selection. The expansion of training set by selecting most relevant tagged images through the Social Active Learning for Image Classification (SALIC)[21] improves the performance. The basic formulation of SALIC requires the reduction in complexity and the minimization of mistakes in unreliable oracle. With the Platt's sigmoidal function, the oracle confidence is mathematically defined as

$$P(S|T) = \frac{1}{1+\exp(AT+B)} \quad (12)$$

Where, S – Random variable corresponding to the selection event

T – Random variable corresponds to probability of examined concept

A, B – Learned parameters from the cross validation.

The design of selective view strategy in response to the application is does not depend on the single view learning framework. Hence, the AL framework is extended to Multi-View (MV) that requires the intelligent selection of the query samples for cost minimization. Here, the combinations of outputs from the individual views provide the reliable prediction. The operations involved in MV-AL framework[25] are listed as follows:

1. Decomposition of hyperspectral image cube into different scales using DWT
2. Utilize the spatial low-frequency sub-bands (LLL, LLH1 to LLHJ) where, each is assigned to dedicated learner.
3. Train the SVM classifiers separately with J+1 views
4. Apply SVM classifier on unlabeled samples and check the query criteria
5. Apply the weighted voting scheme for the actual prediction.

The accuracy of classification is further improved by using the most selective features based on the class-specific model[22]. The selection of class-specific weights from the different weights and the redundancy removal improves the classification accuracy. Based on the label of training samples, the position coordinates are predicted. The weights corresponding to the kernels are to be optimized in class- specific models. The analysis of characteristics of the mixed pixels plays the major role in the HSI classification. The image signatures are observed as the linear combination of previously known spectral signatures that requires the collaborative models.

D. Collaborative models

Post-partitioning and pre-partitioning of the input samples are the major differences in sparse representation and the collaborative representation. In post processing, the linear combination of all the training samples approximates the testing samples with the weight factor. Alternatively, the training data is split up into nearest neighbor samples in pre-processing. The Tikhonov matrix formation governs the collaborative models. The steps for collaborative representation models[18] are listed as follows:

1. Selection of kernel and associated parameters
2. Concatenate the spectral and spatial features
3. Computation of Tikhonov matrix
4. Computation of the weight factor
5. Assign the class label

Linear representation of training samples in the kernel-induced space is governed by the weight factor (α'). The l_2 norm regularization for the weight factor is expressed as

$$\alpha' = \arg\min \|\varphi(y) - \varphi\alpha^*\|_2^2 + \lambda\left\|\Gamma_{\varphi(y)}\alpha^*\right\|_2^2 \quad (13)$$

Where, $\varphi(.)$ – Mapping function

$\Gamma_{\varphi(y)}\alpha^*$ - Tikhonov matrix

The matrix formulation to govern the regularization is expressed as

$$\Gamma_{\varphi(y)} = \begin{bmatrix} \|\varphi(y) - \varphi(x_1)\|_2 & 0 \\ 0 & \|\varphi(y) - \varphi(x_n)\|_2 \end{bmatrix} \quad 14)$$

Where, $\|\varphi(y) - \varphi(x_1)\|_2 = [k(y,y) + k(x_i,x_i) - 2k(y,x_i]^{1/2}$

The employment of closed form solution recovers the weight factor described as

$$\alpha' = \left(K + \lambda^2 \Gamma_{\varphi(y)}{}^T \Gamma_{\varphi(y)}\right)^{-1} k(.,y) \quad (15)$$

With these formulations, the class labels are assigned as

$$class(y) = \arg\min \|\varphi_l \alpha'_l - \varphi(y)\|_2 \quad (16)$$

The above formulations are helpful in the integration of spatial information regarding the neighbor locations corresponding to the feature space. The computational complexity and the parameter selection are the complex issues in the collaborative models. The utilization of single regularization parameter instead of the more parameters in the matrix formulation reduces the computational load and the parameter selection complexity[17]. The objective of the target at sparse representation handles the single target in one time based on l_1-norm minimization. Alternatively, the aim of collaborative representation is to consider all the background samples to isolate the target from the background images. Hence, they are (sparse+collaborative) are combined to achieve the better classification accuracy[19]. The inclusion of either discriminant or the correlation information in the Weighted Collaborative Representation Classifier (WCRC) and the adaptive WCRC [20] improves the classification accuracy by distance information. The lack of prior computation of projective matrix causes the computational complexity in the classification. High-dimensional origin spaces and the non-adaptive kernel variants are the major issues in collaborative models. To alleviate these issues, the collaborative models are integrated with the SVM classifiers.

E. SVM-boosted models

The provision of user hints regarding the informative pixels in the image assure the high-accurate acquisition in AL-based methods. Reduction of re-utilization of previously collected ground truth associated with the images obtained from the sensor requires the suitable classifier. The SVM classifier acts as the base classifier for AL-methods. The weighed variant of SVM classifier plays the major role in the regression formulation and AL methods. The formulation of weighted SVM classifier [34] variant compared to standard SVM classifier are described hereafter. Let us consider the training set $(x_i, y_i)_{i=1}^{n}$ with the n samples. The sample weight assigned to each sample is expressed as $w = (w_i)_{i=1}^{n}$. The primal problem in classification is described by using the SVM formulation expressed as

$$\min_{v,b}\left(\frac{1}{2}\|v\|^2 + C\sum_{i=1}^{n}\omega_i\xi_i\right) \quad (17)$$

Subject to the constraints of

$$y_i(v^T x_i + b) \geq 1 - \xi_i, \quad i = 1, 2, 3, \dots, n$$

Where, v – Vector for hyperplane

b – Associated bias

ξ_i – Magnitudes of error allowance

C – Penalty parameter.

The penalty parameter in (17) describes the trade-off between the margin maximization and the error minimization. With these formulations, the decision function for SVM classifiers expressed as:

$$f(x) = \sum_{i=1}^{n} y_i\alpha_i x_i^T x + b \quad (18)$$

The upper bound of coefficients in (18) define the influence of support vectors and increases the flexibility of SVM-boosted method. Probabilistic output of SVM and the uncertainty in SVM classification leads to generation of multi-feature ensemble SVM classifiers [33]. Simple projection of samples into the high dimensional feature space is the major issues in the Elastic Net Regularizers (ELR) and Logistic Regression (LR). To overcome the issues, the Dual Coordinate Descent (DCD) –Projection method [32] is developed with the same iterative steps and convergence properties. The DCD-based projection improves the generalization performance for extreme sparsity models. With this capability, the time required for the large datasets handling is minimum compared to the standard SVM-boosted model. The AL model with the SVM-based training [35] comprises following steps:

1. Input the training set to the SVM-boosted models
2. Compute the Support Vectors (SV)
3. Check whether the SV satisfies the spectral-selection criteria or not
4. Check whether the SV satisfies the spatial selection criteria or not
5. Apply the non-dominated sorting and select the relevant samples
6. Assign the labels to the relevant samples that satisfy the criteria.

The Table I shows the brief discussion of learning models for HSI classification with their merits and demerits.

Table 1 Information about Different Learning based HSI Classification

Techniques	**Author & Ref**	**Year**	**Performance**	**Advantages**
Sparse Representation Models				
3D-DWT Structured Sparse Logistic Regression	Qian *et al.* [29]	2012	The utilization of 3D-DWT decomposes the hyperspectral data cube into the different scales, orientations and frequencies. Sparse representation/modeling are applied on the captured features such as geometrical and spectral-spatial improved the data analysis	1. Better classification accuracy 2. Minimum empirical loss 3. Low prediction cost
Kernel Sparse Representation	Chen *et al.* [27]	2013	With an assumption of representing the test pixel linearly with the training samples, the kernel sparse representation vector is computed. The direct utilization of vector improves the classification performance	1. Better classification performance than linear version 2. Effective creation of dictionary invariant
MASR	Fang *et al.* [28]	2014	The consideration of complementary uncorrelated information on the regions is the base of the MASR. An adaptive sparse strategy considers the spatial information on multiple scales to improve the classification performance.	1. Correlation usage improved the classification accuracy 2. Easy decision making
Multi-feature Classification	Zhang *et al.* [30]	2016	Simultaneous representation of pixels with the multiple features against the class level space constraint through joint-sparse form preserved the spatial information. The utilization of additional complementary information and the kernalized models handled the non-linear data	1. Easy handling of unequal dimensions 2. Efficient non-linear data separable
Sparsity Based Model	Huang *et al.* [31]	2017	A robust sparsity model for HSI classification performs well by incorporating appropriate priors for noise free and degradation when the data is corrupted byGaussian noise and sparse noise. Futher, it is extended to spatial features.	1. Improves the processing speed.
Kernel-based Image Classification				
Representative Multiple Kernel Learning	Gu *et al.*[13]	2012	The extraction of most variations from the multiple kernels-based space with RMKL is performed. Prior identification of kernels to be preserved and their weights based on statistical significance results in good classification accuracy	1. Satisfactory computational performance 2. Redundancy removal
Generalized composite kernel	Li *et al.* [15]	2013	The construction of generalized composite kernels provided the operational flexibility for combination of spectral-spatial information without any weight parameters.	1. Flexibility in operation 2. Optimization problem was convex.
Multiple Data Dependent kernel	He and Li[14]	2014	Optimization of multi-kernels combination evaluated the degree of agreement between the kernel and learning task. Then, the optimization of data-dependent kernel measured the class separation.	1. Improved classification performance on few-labeled samples scenario
Weighed Multi-feature image classification	Zhang *et al.* [16]	2016	Simultaneous acquiring of representation vector for the features of spectral, spatial and texture by imposing l_1 norm regularization. The assigned weights are different for different features since their contribution was unequal.	1. Less memory requirement 2. Fast operation
Active Learning Models for Spectral-Spatial Image Classification				

Iterative-based window search	Satzoda *et al.* [24]	2016	The employment of Haar-like features and the Adaboost classifiers detect either fully or partially visible target region of vehicle detection applications. The combination of active-learning framework and the symmetry-based analysis improved the vehicle detection performance	1. Better detection 2. Reduced false alarm rate 3. Optimized computational level
DAEM	Polewski *et al.*[23]	2016	Based on Reny entropy regularization, the active and semi-supervised learning combination enabled the synergy effect. An active learning approach based on the entropy minimization of unlabeled object probabilities is discussed	1. Expected Error Reduction (EER) 2. Minimum computational complexity
SALIC	Chatzilari *et al.* [21]	2016	The utilization of probabilistic approach supported the joint maximization of informativeness and confidence. Under the noisy context examination, the contextual-based indication of true contents of images is performed.	1. High learning speed 2. Minimum computational complexity 3. Less mistakes
3D-RDWT	Zhou *et al.* [25]	2016	The incorporation of 3D-RDWT generated the multiple views and they are integrated into the multiview active learning framework. The 3D-RDWT based spatial features provided the sufficient view and less sensitive to additive noise.	1. Strong and diverse views 2. Query efficiency improvement
CS-SMKL	Liu *et al.* [22]	2016	An adoption of extended multiattribute profiles represent the spatial and spectral information effectively. The utilization of l_1 norm regularization constraint supported the automatic learning of compact feature set	1. Effective retaining of useful features 2. Enhanced discriminability
Superpixel-AL	Guo [26]	2017	Use of superpixel as an adaptive window allows to extract spatial features that enables to extract informative and discriminative features of each object.	1. Effective classification on difficult classes.
Collaborative Representation-based classification				
WCRC	Timofte and Gool [20]	2014	Consideration of weight-based on the classification confidence for samples and the variance of feature channels. The formulation of l_1 and l_2 regularized least square decomposition provided the algebraic solution effectively	1. Simplicity in operation 2. High speed
Refined Sparse Unmixing	Iordache *et al* [17]	2014	An exploitation of low number of endmembers in the large size library of real images. The adoption of either multi-task or the simultaneous sparse regression framework improved the unmixing results	1. Dimensionality reduction of end members 2. Robust solutions to the joint sparse representation problem
CSCR	Li *et al* [19]	2015	Sparse representation encourages the competition among the atoms and the collaborative representation utilizes all the atoms for target detection	1. Achievement of optimal and sub-optimal performance parameters 2. Better performance without tuning of regularization parameters
KCRT	Li *et al.* [18]	2015	Projection of data on the high dimensional kernel space with non-linear mapping function improved the class separation. Besides, the spatial information is incorporated into the neighboring locations.	1. Widely acceptable classifiers in remote-sensing applications 2. High efficient capture of spectral-spatial variations
SVM-Boosting Strategies for Spectral-Spatial Classification				
SVM-boosted AL	Matasci *et al.* [34]	2012	An adoption of classifier trained by the images with same spectral properties. Selection of pixels	1. Better understanding 2. Integration of instant

			to be labeled and the update of weights assigned to them enhanced the classification	weights conveyed the importance of domain of interest
Multifeature SVM	Huang *et al.* [33]	2013	Construction of SVM combining multifeatures (spectral-spatial) at both pixel and object levels is discussed. GLCM, morphological and the complexity index are the major features in multifeature SVM.	1. Accurate classification results 2. Discrimination enhancement
Logistic Regression	Balamurugan [32]	2013	The alternate optimization approach comprises elastic net regularized linear SVM and the Logistic Regression (LR) to overcome the simple projection problem.	1. Fast generalization performance 2. Few passes of data
Parzon Window	Pasolli *et al.* [35]	2014	A new active learning that defines the remote sensing images intrinsically both in spectral-spatial domains. The explicit computation of Euclidean distance and the parzon window initialization with the spatial entropy are performed.	1. Reduction of manual sample labeling 2. Less computational time

III. CONCLUSION

The inclusion of multiple features from the multiple algorithms improved the classification accuracy. This paper reviewed the traditional research studies on the HSI classification. This paper organized the survey into five groups namely, sparse-representation, kernel, active learning, collaborative representation and the Support vector Machine (SVM) boosted learning models. The brief literature review in this paper highlighted the processes involved in each model, advantages and disadvantages in detail. Dimensionality, low-quality background, computational time, representation errors are the major limiting factors in HSI classification. This paper presented the clear description of the mathematical models and associated parameters in each model. The survey addressed the issues in the sparse model generation with the noise effects, parameter selection, and weight adjustments. The comprehensive review highlighted that the multi-label segmentation problem induces the difficulties in the image clustering with various spectral limitations. An adaptive change of rule with an increase in data dimension is the major issue in the HSI classification.

REFERENCES

[1] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Processing Magazine,* vol. 31, pp. 45-54, 2014.

[2] W. Li and Q. Du, "Gabor-filtering-based nearest regularized subspace for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 7, pp. 1012-1022, 2014.

[3] W. Li, S. Prasad, and J. E. Fowler, "Decision fusion in kernel-induced spaces for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, pp. 3399-3411, 2014.

[4] R. M. Willett, M. F. Duarte, M. A. Davenport, and R. G. Baraniuk, "Sparsity and structure in hyperspectral imaging: Sensing, reconstruction, and target detection," *IEEE signal processing magazine,* vol. 31, pp. 116-126, 2014.

[5] Y. Zhang and S. Prasad, "Locality preserving composite kernel feature extraction for multi-source geospatial image analysis," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 8, pp. 1385-1392, 2015.

[6] B. Du, L. Zhang, L. Zhang, T. Chen, and K. Wu, "A discriminative manifold learning based dimension reduction method for hyperspectral classification," *International Journal of Fuzzy Systems,* vol. 14, pp. 272-277, 2012.

[7] J. Li, H. Zhang, L. Zhang, X. Huang, and L. Zhang, "Joint collaborative representation with multitask learning for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, pp. 5923-5936, 2014.

[8] S. Prasad, W. Li, J. E. Fowler, and L. M. Bruce, "Information fusion in the redundant-wavelet-transform domain for noise-robust hyperspectral classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, pp. 3474-3486, 2012.

[9] Z. Shao, L. Zhang, X. Zhou, and L. Ding, "A novel hierarchical semisupervised SVM for classification of hyperspectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, pp. 1609-1613, 2014.

[10] Y. Y. Tang, H. Yuan, and L. Li, "Manifold-based sparse representation for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, pp. 7606-7618, 2014.

[11] Q. Wang, J. Lin, and Y. Yuan, "Salient Band Selection for Hyperspectral Image Classification via Manifold Ranking," *IEEE transactions on neural networks and learning systems*, vol. 27, pp. 1279-1289, 2016.

[12] B. Waske and J. A. Benediktsson, "Fusion of support vector machines for classification of multisensor data," *IEEE transactions on geoscience and remote sensing*, vol. 45, pp. 3858-3866, 2007.

[13] Y. Gu, C. Wang, D. You, Y. Zhang, S. Wang, and Y. Zhang, "Representative multiple kernel learning for classification in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, pp. 2852-2865, 2012.

[14] Z. He and J. Li, "Multiple data-dependent kernel for classification of hyperspectral images," *Expert Systems with Applications*, vol. 42, pp. 1118-1135, 2015.

[15] J. Li, P. R. Marpu, A. Plaza, J. M. Bioucas-Dias, and J. A. Benediktsson, "Generalized composite kernel framework for hyperspectral image classification," *IEEE transactions on geoscience and remote sensing*, vol. 51, pp. 4816-4829, 2013.

[16] E. Zhang, X. Zhang, L. Jiao, H. Liu, S. Wang, and B. Hou, "Weighted multifeature hyperspectral image classification via kernel joint sparse representation," *Neurocomputing*, vol. 178, pp. 71-86, 2016.

[17] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, pp. 341-354, 2014.

[18] W. Li, Q. Du, and M. Xiong, "Kernel collaborative representation with Tikhonov regularization for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, pp. 48-52, 2015.

[19] W. Li, Q. Du, and B. Zhang, "Combined sparse and collaborative representation for hyperspectral target detection," *Pattern Recognition*, vol. 48, pp. 3904-3916, 2015.

[20] R. Timofte and L. Van Gool, "Adaptive and weighted collaborative representations for image classification," *Pattern Recognition Letters*, vol. 43, pp. 127-135, 2014.

[21] E. Chatzilari, S. Nikolopoulos, Y. Kompatsiaris, and J. Kittler, "SALIC: Social Active Learning for Image Classification."

[22] T. Liu, Y. Gu, X. Jia, J. A. Benediktsson, and J. Chanussot, "Class-Specific Sparse Multiple Kernel Learning for Spectral–Spatial Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, p. 7351, 2016.

[23] P. Polewski, W. Yao, M. Heurich, P. Krzystek, and U. Stilla, "Combining Active and Semisupervised Learning of Remote Sensing Data Within a Renyi Entropy Regularization Framework."

[24] R. K. Satzoda and M. M. Trivedi, "Multipart Vehicle Detection Using Symmetry-Derived Analysis and Active Learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, pp. 926-937, 2016.

[25] X. Zhou, S. Prasad, and M. M. Crawford, "Wavelet-Domain Multiview Active Learning for Spatial-Spectral Hyperspectral Image Classification."

[26] Jielian Guo, Xiong Zhou et al. " Superpixel-Based Active Learning and Online Feature Importance Learning for Hyperspectral Image Analysis", *IEEE Journal of Selected Topics in Applied Earth*

[27] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Transactions on*

Geoscience and Remote sensing, vol. 51, pp. 217-231, 2013.

[28] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral–spatial hyperspectral image classification via multiscale adaptive sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, pp. 7738-7749, 2014.

[29] Y. Qian, M. Ye, and J. Zhou, "Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, pp. 2276-2291, 2013.

[30] E. Zhang, L. Jiao, X. Zhang, H. Liu, and S. Wang, "Class-Level Joint Sparse Representation for Multifeature-Based Hyperspectral Image Classification."

[31] S. Huang, H. Zhang and A. Pizurica, " A Robust Sparse Representation Model for Hyperspectral Image Classification", *Sensor*, September 2017.

[32] P. Balamurugan, "Large-scale elastic net regularized linear classification svms and logistic regression," in *2013 IEEE 13th International Conference on Data Mining*, 2013, pp. 949-954.

[33] X. Huang and L. Zhang, "An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery," *IEEE transactions on geoscience and remote sensing*, vol. 51, pp. 257-272, 2013.

[34] G. Matasci, D. Tuia, and M. Kanevski, "SVM-based boosting of active learning strategies for efficient domain adaptation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, pp. 1335-1343, 2012.

[35] E. Pasolli, F. Melgani, D. Tuia, F. Pacifici, and W. J. Emery, "SVM active learning approach for image classification using spatial information," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, pp. 2217-2233, 2014.