A Survey on Remote Sensing Scene Classification Algorithms

DEBABRATA GHOSH, NAIMA KAABOUCH
Department of Electrical Engineering
University of North Dakota,
Grand Forks, ND, USA 58203
debabrata.ghosh@my.und.edu
naima.kaabouch@engr.und.edu

Abstract: - Scene classification has been widely utilized in various remote sensing applications. Successful image classification depends on several factors, such as availability of data, complexity of available data, availability of ancillary data, expertise of an analyst, availability of suitable classification algorithms, etc. There is no single best classification method that would be suitable for all applications. This paper aims at highlighting the present-day practices of scene classification by summarizing the major scene classification categories available in the literature. Research shows that high-level classification outperforms the other classification methods for almost any kind of data, however, at the cost of high computation. Further research is needed to improve classification accuracy and at the same time reduce computational complexity in order to make a classification method more suitable for real time applications.

Key-Words: - Nearest Neighbor, Support Vector Machine, Artificial Neural Network, Decision Tree

1 Introduction

Scene classification is an important and challenging task in various application areas such as biomedical imaging, video surveillance, vehicle navigation, remote sensing, etc. Remote sensing images with a wide range of spatial, temporal, spectral, and radiometric resolutions have been significantly utilized to map the land-cover/ land-use information on the Earth's surface [1-9]. The intent of the classification process is to categorize a digital image into one of several distinct and exclusive classes or categories. This categorized data may then be used to produce thematic maps of the details present in an image. A classification unit could be a single pixel, a group of neighboring pixels, or the whole image. The classes of a classification process may be predefined by an analyst (in the case of supervised classification) or automatically clustered into sets of prototype classes (in the case of unsupervised classification). Irrespective of an analyst's importance in constructing the training classes, they need to be independent, discriminatory, and reliable.

Scene classification usually consists of seven steps. The first step for successful classification is selection of suitable sensor data. If classification is performed on a local level, it is preferred to have data with high spatial resolution. However, if classification in global level is required, data with coarse spatial resolution are helpful [10]. Selecting sufficient number of training samples is the second step for scene classification. Training samples are collected from the respective area of study. The third step for scene classification is image preprocessing, which includes radiometric processing, geometric correction, atmospheric correction, noise removal, etc. Extracting appropriate features from data is the fourth step for scene classification. This feature may be spectral signature, vegetation indices, textural or contextual details, local or global image descriptors, etc. The next step is selection of a suitable classification algorithm. Post-process a classified image is the sixth step for image classification. Up-scaling and raster to vector conversion are the two most common post-processing techniques applied to a classified map. Performance assessment is the final step for scene classification. Evaluation of a classification algorithm could be quantitative or qualitative. As algorithms are becoming more accurate in recent years, quantitative evaluation becomes more important. Classification accuracy seems to be the most appropriate performance evaluation approach. Otherwise, several other assessment criteria are available: stability of algorithm, robustness to noise, computational resources, etc.

In the literature, several advanced scene classification algorithms have been developed [11-20]. Even though the state of art clearly indicates improved accuracy of the recently developed classification methods, scene classification still remains a challenge because of several factors such

as complexity of the scene, availability of the sensor data, availability of scene classification algorithms, etc. Thus, majority of the recent works is focused specifically on dealing with the challenges of the scene classification methods. A comprehensive up-to-date review of the newly emerged classification algorithms is mostly overlooked. Furthermore, none of the existing surveys discusses all the major classification categories and eventually fails to categorize the most recent scene classification algorithms. The regular emergence of new classification algorithms in present days requires a comprehensive review work, which could be a valuable guide for selecting a suitable classification procedure for a specific application.

2 Introduction

There are several ways of grouping the existing scene classification algorithms. Grouping could be based on analyst's contribution in classification methods, or based on parameters on data used, or based on pixel information used, or based on knowledge available from ancillary data, or based on image attributes used. Based or analyst's role, scene classification can be supervised and unsupervised classification. Based on parameters on data used, scenes can be classified as parametric and non-parametric classification. Based on pixel information, scene classification can be per-pixel, sub-pixel, per-field, and contextual classification. Based on availability of knowledge, images can be classified as knowledge-free and knowledge-based classification. Finally, based on attributes, images can be classified using low-level and high-level classification. For convenience, this paper classifies the existing algorithms into per-pixel, sub-pixel, per-field, contextual, knowledge-based, and high-level classification. Categories of scene classification algorithms are described briefly in the following subsection.

2.1 Per-pixel Classification

In per-pixel classification each pixel is assigned to a class by considering the spectral similarities with the different classes [21]. Per-pixel classification can be parametric or non-parametric. In parametric classification it is assumed that the probability distribution of each of the classes is known. Usually, parameters like mean vector and covariance matrix are obtained from the training data. However, the assumption of normal probability distribution of each class is often violated for complex landscapes. Moreover, insufficient training samples may lead to a singular covariance matrix. The most commonly used parametric classifier is the maximum likelihood classifier (MLC). Unlike parametric classification, non-parametric classification is neither based on any assumption nor uses statistical parameters. This classifier assigns pixels to classes based on pixel's position in discretely positioned feature space [22]. Some of the most commonly used non-parametric classifiers are nearest neighbor (NN), support vector machine (SVM), artificial neural network (ANN) based classifiers, and decision tree-based classifiers.

2.1.1 Maximum Likelihood Classification

MLC is one of the most common methods for thematic mapping using aerial images. Initially, the statistical probability of an unknown pixel X with multidimensional values (d dimensions) being a member of each class is computed using a likelihood function. It is assumed that the training samples of each class form the Gaussian (normal) distribution. Using multivariate normal density function in d dimension the likelihood function for the pixel X being a member of class c can be expressed by LC (X) as:

$$L_C(X) = \frac{1}{(2\pi)^{\frac{d}{2}}|S_C|^{\frac{1}{2}}} e^{\left\{-\frac{1}{2}(X-\overline{X_C})^T S_C^{-1}(X-\overline{X_c})\right\}} \quad (1)$$

Where S_C is the variance-covariance matrix of class C produced from the training data, $|S_C|$ is the determinant of S_C. $\overline{X_C}$ is the mean vector of class C produced from the training data. Finally, X is assigned to class c if $L_C(X) > L_i(X)\ \forall i \neq c$ or X is labeled as 'unknown' if the probabilities are below a predefined threshold [23]. When there is large number of classes which needs to be classified, MLC requires high degree of computation. Also, if the sample size for each class is not large enough then this method may have large variance and hence be unreliable [24]. Furthermore, if the classes have heavy overlapping, maximum likelihood generates classified map with poor separation among the classes.

An extension of MLC is Bayesian classification, which states that the probability of an unknown pixel belonging to class c, given the feature vector X is expressed by:

$$P(c|X) = \frac{P(X|c)P(c)}{P(X)} \quad (2)$$

Where $P(X|c)$ is the likelihood function, $P(c)$ is the priori probability of class c, and $P(X)$ is the

probability that X is observed. As P(X) is independent of c the aim is to maximize P(X|c)P(c) in order to maximize P(c|X). Again, if P(c) has uniform distribution, maximizing P(c|X) means maximizing P(X|c), i.e. maximizing the likelihood function. Two of the noble contributions include land cover classification using cloud-contaminated high-resolution images proposed by Salberg et al. [25], MLC based on bivariate Gaussian mixture model proposed by Liu et al. [26]. These methods are discussed briefly as below:

Salberg et al. [25] proposes a two-stage classification technique for cloud and snow contaminated multitemporal high resolution remote sensing data. The first stage consists of identifying pixels containing snow and clouds, and the second stage uses the information to model a missing observation and subsequently performs a pixel-level fusion of the cloud/snow map, and the actual sensor data. Finally a trained classifier, which can handle missing data, is used to classify the land cover. Missing data is modeled as:

$$x = rv + (1 - r)w \quad (3)$$

Where x is the feature vector corresponding to a pixel, r is a response indicator of the missing data mechanism. If r is 1 the feature vector corresponds to land cover observation v, and when it is 0 the feature vector corresponds to cloud or snow observation w. Let the input data matrix be $X = [x_1, x_2, \dots, x_N]$, and the corresponding missing data indicator matrix be $R = [\rho_1, \rho_2, \dots, \rho_N]$. Each x in the matrix X is a d-dimensional vector (considering there are d number of spectral bands used), and each ρ in the matrix R indicates the missing data in corresponding x. Considering missing data to be "missing as random" (MAR), interaction between R and the observed and missing features of X could be expressed as $P(R|X, Z, \theta) = P(R|\chi_o, \theta)$. χ_o is the set of observed features from X. After the missing data model is formed, in the learning phase EM algorithm (expectation-maximization algorithm) is used to estimate the mean vector and the covariance matrix from the training data for each land cover class. After this learning phase, a cloud/snow map is computed using each of the input images. For testing, sequence of three Landsat ETM+ images are used. Other than MLC, performance is also evaluated using k-nearest neighbor (kNN) and Parzen classifiers. It is observed that the classification performance is improved using pixel-level image fusion. For MLC performance is improved from 62.8% to 74.9% and for kNN classifier performance is improved from 63.4% to 77.4%.

Liu et al. [26] develops a MLC that can handle subject variability. Conventional supervised method for estimating the unknown parameters of the MLC algorithm assumes that the feature values of different classes have similar distribution for different subjects. However, this assumption is not feasible in many real situations. Thus, in order to consider the differences across subjects, the proposed method formulates a statistical model using the intensity values and the mean intensity values of each subject as bivariate random variable. Assuming Y is the intensity value of each pixel denoted by y and K is the class of each pixel denoted by k the estimator $\hat{K}$ of K is obtained by maximizing the probability of each class as:

$$\hat{K} = \arg\max_k p(y, z|K = k) \quad (4)$$

Here y and z are expressed as two groups of random variables representing intensity value of a pixel and the mean intensity values of each subject respectively. Assuming simplified development with background and a single object, and assuming y and z belong to a bivariate Gaussian distribution, the above equation can be rewritten as:

$$\hat{K} = \arg\max_k \frac{1}{2\pi|\Sigma_k|^{1/2}} \exp[-\frac{1}{2}(x - \mu_k)\Sigma_k^{-1}(x - \mu_k)^T] \quad (5)$$

Where μ_k is the mean vector and Σ_k is the covariance matrix for the k^{th} class.

Evaluation is performed using several simulated image datasets. The qualitative and quantitative comparison between the proposed method and the classical supervised MLC reveals that the proposed method outperforms the classical method when there are considerable differences between subjects.

2.1.2 Nearest Neighbor Classification

Nearest neighbor based algorithms are simple but effective methods used in statistical classification. Categorizing unlabeled samples is based on their distance from the samples in training dataset. Let a set of n labeled training samples be given as $S = \{x_1, x_2, \dots, x_n\}$, where $x_i \in \mathbb{R}^d$. According to the nearest neighbor classification rule, an unlabeled sample t is assigned to the class of $x_i \in S$ if x_i happens to be the nearest neighbor of t. Usually Euclidean distance is used as a measure of nearest neighbor. On the other hand, according to kNN classification a set of k nearest neighbors is

computed for an unlabeled sample t instead of a single nearest neighbor. Then, the test sample is assigned to the class that occurs most frequently among the k-nearest training samples. If the ranges of the data in each dimension vary considerably, this can affect the accuracy of the nearest neighbor based classifications. Thus, both the training and testing data need be normalized [27].

2.1.3 Support Vector Machine Classification

SVM is an efficient supervised binary classification technique. Given a set of labeled training samples, SVM outputs an optimal hyperplane which can categorize unlabeled samples. An optimal hyperplane is the one which has largest distance from the nearest training sample of each class. Let a set of n training samples with their labels be given as $S = \{x_i, y_i\}, 1 \leq i \leq n, y_i \in \{-1,1\}$ where $x_i \in \mathbb{R}^d$. For linearly separable case, hyperplane can be expressed as:

$$w.x + b = 0 \quad (6)$$

Where w is a normal vector to the hyperplane, b is the offset for the hyperplane. The problem of finding the optimal hyperplane could be expressed as the following convex optimization problem:

Minimize $\frac{1}{2}\|w\|^2$ subject to $y_i(w.x + b) \geq 1$

Using Lagrange multipliers α the above optimization can be further expressed in a dual formulation as:

Maximize $L = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} y_i y_j x_i . x_j \alpha_i \alpha_j$

Subject to $\sum_{i=1}^{n} \alpha_i y_i = 0$ and $\alpha_i \geq 0$

For nonlinearly separable case, training samples are mapped into a higher dimensional space where linear separation is possible. Thus a transformation function Φ is necessary. The high time consuming computation $\Phi(x_i).\Phi(x_j)$ can be avoided by using suitable kernel function [28]. Accordingly $K(x_i, x_j) = \Phi(x_i).\Phi(x_j)$.

Thus for nonlinear SVM, objective is to maximize $L = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} y_i y_j K(x_i, x_j)\alpha_i \alpha_j$ such that $\sum_{i=1}^{n} \alpha_i y_i = 0$ and $\alpha_i \geq 0$

Once this maximization problem is solved, α_i and b can be used to classify test point x [29]. The decision function can be expressed as:

$$\text{sign}[\sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b]$$

SVM classification methods have often found to provide higher accuracies compared to other methods, such as MLC, ANN-based classifications. SVM classifiers always deliver unique solutions, since the optimality problem is convex. Some of the significant contributions in SVM classification include cluster-assumption based active- learning for classifying remote sensing images proposed by Patra et al. [30], fusion of texture and SIFT-based descriptors for remote sensing image classification proposed by Risojević et al. [31], image classification based on linear distance coding proposed by Wang et al. [32]. These algorithms are presented briefly as follows:

Patra et al. [30] develops a reliable active-learning based classification for remote sensing images. Collecting labeled samples is time consuming and costly. Also, redundant samples slow down the training process. Thus, training set needs to be kept as small as possible to avoid redundancy, and at the same time, patterns with the largest amount of information need to be included in the training set. The proposed active learning method is implemented in the learning phase of the SVM classifier. The SVM classifier is first trained with a small number of labeled samples. Each unlabeled sample is given an output score based on how likely or unlikely it is a member of a class. These output scores are plotted into a histogram. Thus, the most ambiguous samples generate output scores located in the valley region of the histogram. A threshold is chosen to determine which unlabeled samples should be considered. This technique is not strongly affected by the initial training samples chosen and it is simple in terms of computational complexity. Thus, it has important advantage in remote sensing applications.

Risojević et al. [31] proposes a hierarchical fusion of local and global descriptors in order to classify high resolution remote sensing images. They suggest use of a Gabor filter bank at S scales and K orientations. An Enhanced Gabor texture descriptor (EGTD) is developed based on cross correlation between the spatial-frequency sub-bands of Gabor image decomposition using the following formula:

$$\upsilon_{mm'n} = \sqrt{\iint_{\Omega} |\Delta W_{mm'n}(x,y) - \psi_{mm'n}|^2 dxdy} \quad (7)$$

Where

$$\Delta W_{mm'n}(x,y) = \left| \frac{W_{mn}(x,y)}{\mu_{mn}} - \frac{W_{m'n}(x,y)}{\mu_{m'n}} \right| \quad (8)$$

$W_{mn}(x,y)$ and $W_{m'n}(x,y)$ are the responses of a Gabor filter at orientation n and scales m and m′, respectively. The opponent features, which measure the cross correlations between the filter responses at different scales can be expressed as:

$$\psi_{mm'n} = \iint_{\Omega} \Delta W_{mm'n}(x,y)dxdy \quad (9)$$

$$\text{Means } \mu_{mn} = \iint_{\Omega} |W_{mn}(x,y)|dxdy \quad (10)$$

The responses of the filters at same scale but different orientations are also correlated. Finally, $\upsilon_{mm'n}$ (and values obtained by using correlation of filter responses at same scale but different orientations) values are arranged into SK(S+K) dimensional column vector to get the EGTD, which is mainly texture oriented. Since it averages the wavelet coefficients over the image domain, it is unable to encode the local information from an image. Thus, EGTD is good for classifying images, which are mainly texture- based. On the other hand, Bag of Words (BoW), which works using SIFT-based local descriptor, is good at classifying images with distinctive structures. Since EGTD and SIFT encode complementary information of an image, a hierarchical fusion for local and global descriptors is used. First, level-0 classifiers are trained for both the descriptors. The confidence scores returned by the level-0 classifiers are then transformed into posterior probabilities and later concatenated to obtain mid-level representation. Mid-level descriptors are then used as input to the level-1 classifier (metalearner), which outputs the final prediction. The SVM with linear, χ^2, and radial basis function (RBF) based kernels are used as metalearners. It is observed from the evaluation that in the absence of any metalearner, i.e., simple concatenation of the descriptors performs poorly compared to the cases when metalearners are used. It is also observed that linear classifiers perform equally well when compared to the classifiers using kernel.

Wang et al. [32] develops a linear distance coding (LDC) based classification method. Bag of Words (BoW) based classifier uses the three-step method: extraction of local features of an image, generating codebook and then quantize/encode local features accordingly, finally pooling all the codes together to generate a global image representation. However, because of the quantization process, the information loss is inevitable in such a feature extraction-coding-pooling based method. Naive Bayes Nearest Neighbor (NBNN) method tackles this information loss by avoiding the quantization/coding process. Instead, it uses image-to-class distance, which it calculates based on local features. Since, spatial context of images needs to be explored more effectively for better performance of a classifier, Spatial Pyramid Matching (SPM) is often used as coding-pooling based methods. However, SPM strictly requires that the involved images exhibit similar spatial layout. The proposed method uses the advantages of both BoW and NBNN, and at the same time relieve the strict spatial layout requirement for SPM. In this method each local feature is transformed into a distance vector, whose each element represents certain class-specific semantics. Since image representation produced by LDC is complementary to the one produced by original coding-pooling method, their combination can result in performance improvement of a classifier. Performance is evaluated using both Locally-constrained Linear Coding (LLC) and Localized Soft-Assignment Coding (LSA) as the linear coding method. LLC and LSA are individually used as coding methods, whereas max-pooling is always employed. Original coding-pooling based image representation, LDC based image representation, and their concatenation are used for evaluation. It is observed that the concatenated representation outperforms the other two in terms of classification accuracy.

2.1.4 Artificial Neural Network-based Classification

ANN is a computational model inspired by the biological neural network. It could be considered as a weighted directed graph in which nodes are neurons and edges with weights are connection among the neurons. Each artificial neuron computes a weighted sum of its n input signals $x_j, j = 1,2,3,\dots,n$ and generates an output, based on certain activation functions, such as piecewise linear, sigmoid, Gaussian, etc. Use of a sigmoid activation function can be mathematically expressed as:

$$y = \frac{1}{(1+\exp^{-\sum_j \frac{w_j x_j}{p}})} \quad (11)$$

Where w_j is the connection weight associated with the j^{th} input, and p is the slope parameter.

ANN consists of one input layer, one output layer, and depending on the application it may or may not have hidden layers. The number of nodes at the output layer is equal to the number of information classes, whereas the number of nodes at the input is equal to the dimensionality of each pixel. Feed-forward ANN with the back propagation

learning algorithm is most commonly used in ANN literature. In the learning phase, the network must learn the connection weights iteratively from a set of training samples. The network gives an output, corresponding to each input. The generated output is compared to the desired output. The error between these two is used to modify the weights of the ANN. The training procedure ends when the error becomes less than a predefined threshold. Then, all the testing data are fed into the classifier to perform the classification.

For very high dimensional data, the learning time of a neural network can be very long, and the resulting ANN can be very complex [33]. Consequently, several ANN-based classification algorithms have been proposed in literature, aiming to minimize the complexity. Both [34, 35] suggest use of adaboost algorithm, i.e., building a strong classifier, using linear combination of several weak classifiers. Both of them use a back propagation learning algorithm. A two-layer-ANN with a single hidden layer is used in [34] as a weak classifier. It uses 50 nodes in the input layer and 25 nodes in the hidden layer. The proposed algorithm works by using weak classifier in a number of iterations (t) and by maintaining a distribution of weights for the training samples. Initially the training samples are assigned equal distribution. However, in subsequent iterations weights of poorly predicted training samples are increased. Finally, the weak classifier finds a weak hypothesis h_t which is suitable for the distribution of the samples at that iteration. A confidence score α_t for the weak hypothesis is also calculated. The final classification result for an input vector is based on the sign of f(x) as:

$$\text{sign}\big(f(x)\big) = \text{sign}(\textstyle\sum_t \alpha_t h_t(x)) \qquad (12)$$

AVIRIS data is used in [34] to compare the performance of the proposed algorithm with MLC. It is evaluated that the maximum likelihood-based classifier requires 4,554 parameters for learning, whereas the proposed algorithm requires only 975 parameters for learning but still the proposed method outperforms the maximum likelihood-based classifier. The INRIA human database is used in [35] for evaluation. Three different combinations of weak classifiers are tested by varying the number of nodes in the hidden nodes from 1 to 3. When more hidden nodes are used, the accuracy of the proposed classifier is seen to be better. Comparing the proposed classifier with global linear SVM, global kernel SVM, and cascade linear SVM based classifiers show that the proposed algorithm performs better than the others.

In order to address the complexity of the ANN in case of high dimensional data, feature reduction mechanisms have also been investigated in literature. Majhi et al. [36] proposes a low complexity ANN for recognition of handwritten numerals. The proposed method uses an image database of handwritten numerals. Three hundred ninety six data for each numeral is used for training purposes. At first, binary images of the numerals are converted into gray scale images. Then gradient and curvature values are computed for each image, and subsequently 2,592 dimensional gradient feature vectors and 2,592 dimensional curvature feature vectors are generated. A principal component analysis (PCA) technique is used to compress the data and generate gradient feature vectors and curvature feature vectors of dimensions 66 and 64, respectively. These feature vectors are extended to trigonometric terms and fed to a low complexity single layer classifier. Each numeral with 100 data is used for testing purposes. The classification accuracy obtained using gradient feature vectors and curvature feature vectors are 98% and 94% respectively. Also, it is evaluated that the performance of the proposed algorithm is comparable to the modified quadratic discriminant function (MDQF) based classifier, but it offers low complexity.

2.1.5 Decision Tree-based Classification

A supervised classifier which requires less complicated training compared to the ANN is based on a decision tree. A decision tree breaks up a complex decision into multiple simpler decisions so that the final solution resembles the desired solution. Decision tree is a hierarchical structure consisting of nodes and directed edges. Each node is an attribute of an observation that needs to be classified, whereas each edge represents a value the attribute can take. The root node is the attribute, which best divides the training data, whereas each leaf node is assigned a class label. Hunt's algorithm is the most commonly used method for building a decision tree. Hunt's algorithm recursively partitions the training data until all the members of each partition belongs to the same class label, or there are no more attributes remaining for partitioning [37]. Selecting the best split (also known as attribute selection) is a challenging task while building a decision tree and consequently several measures are proposed in literature. The goodness of a split can be measured quantitatively by several metrics, such as information gain, information gain ratio, Gini index etc. While using a large dataset, a decision tree

representation can be significantly complex and, hence, classification may suffer from substantial complexity. As a result, a number of pruning methods are employed to reduce the size of the decision tree by removing sections of the tree, which are insignificant in classifying observations. Two of the significant contributions in decision tree based classification are discussed as follows:

Pal et al .[38] proposes the use of a univariate decision tree classifier with error-based pruning (EBP). They use four different attribute-selection-measure metrics to verify that the classification accuracy is not affected by the choice of attribute-selection- measure metric. The accuracy of the decision tree classifier is measured while using different pruning methods, such as reduced error pruning (REP), pessimistic error pruning (PEP), error-based pruning (EBP), critical value pruning (CVP), and cost complexity pruning (CCP). It reveals that the EBP outperforms the other pruning methods. They also perform a comparative evaluation between ANN-based classification and the proposed decision tree-based classification. Accuracy and processing time are recorded for both the ANN-based classifier and the decision tree-based classifier, using ETM+ and InSAR datasets. It shows that for both the datasets, the decision tree-based classifier performs better than the other in terms of both classification accuracy and processing time.

Thangaparvathi et al. [39] proposes a modification to the RainForest algorithm, which was developed to address the scalability issue when a large dataset is used. The data structure used in this proposed method IAVC set and IAVC group is the improved version of AVC set (attribute-value class) and AVC group used in the RainForest algorithm. The IAVC group of a node is the set of IAVC sets of all attributes at that node. The IAVC set, on the other hand, captures class distribution information as well as distribution of record IDS. Rather than working on the actual dataset, this algorithm works on this aforementioned data structure. At a specific time, only one IAVC group of a certain node is kept in the main memory. This method requires a single pass over the dataset instead of a pass over the dataset at each level of the decision tree construction. Thus, time complexity of the proposed method reduces to $O(R)$from the RainForest algorithm's $O(R[L+1])$ where L is the number of levels of the decision tree, and R is the number of records in the database. Two different datasets are used for training purposes. Experimental evaluation, using datasets with sizes ranging from 100,000 to 1,000,000, shows that the proposed method outperforms the RainForest algorithm in the time dimension.

Several other decision tree-based classifiers have been proposed, which use a variation of the Hunt's algorithm as the decision tree induction method. [40, 41] use the classification based on the ID3 algorithm. [42] uses the C4.5 decision tree classifier and [43] utilizes the CART based decision tree.

2.2 Sub-pixel Classification

In principle the per-pixel classification methods assign each pixel into one category based on the assumption that a given pixel can belong only to one class. However, in certain situations, the geometric resolution of the imaging sensor might not be high enough to guarantee that the radiance measurement associated with a pixel is the contribution of single information class in the scene [44]. Rather, the radiance measurement of a pixel could be contribution of multiple classes. This is very common in medium and coarse spatial resolution remote sensing images. In order to overcome this mixed-pixel problem, proportional membership of each pixel measurement to each information class is considered. This method is known as sub-pixel classification. Significant contributions in sub-pixel classification includes a fuzzy rule classifier for classifying satellite images proposed by Prabhu et al. [45], classification of very high resolution satellite images using fuzzy logic proposed by Jabari et al. [46]. These two methods are briefly discussed as follows:

Prabhu et al. [45] recommends a fuzzy rule classifier for classifying satellite images into barren land, vegetation area, building area, and road area. Initially, the satellite images are Gaussian filtered to reduce the noise contents and then they are converted to Lab color space. Lab color approximates the human vision more accurately than RGB color. These pre-processed satellite images are segmented into building and vegetation areas using intensity constraint method. Then a modified graph cut is used to segment the barren land areas. The modified graph cut considers spatially neighboring pixels in addition to the two pixels while making a cut. After that some thresholding techniques are used to segment the road areas. Once the segmentation is over, features such as mean, variance, and entropy are extracted from those segmented areas. Finally, a fuzzy rule classifier is used to classify scenes into barren lands, building areas, road areas, and vegetation areas. The proposed fuzzy rule classifier has two steps. In the first step, the classifier generates the fuzzy rules

using an indiscernibility matrix of rough set theory and gives those rules in the rule base of the classifier. In the second step, a triangular membership function is used to convert a test data into fuzzified value. Then the fuzzified input is matched with the fuzzy rules. Subsequently, a fuzzy score is generated to classify a scene into different areas. The triangular membership function that is used in the proposed method is:

$$f(x) = \begin{cases} 0 & \text{if } x \le a \\ \frac{x-a}{b-a} & \text{if } a \le x \le b \\ \frac{c-x}{c-b} & \text{if } b \le x \le c \\ 0 & \text{if } x \ge c \end{cases} \quad (13)$$

Where a, b, c are the vertices of the triangular function. a is the lower boundary and c is the upper boundary where membership degrees are zero. b is the center where membership degree is 1. The average classification accuracy is measured as 92%. Also, it is evaluated that the modified graph cut segmentation increases the accuracy of the classification algorithm almost 6% compared to when a normal graph cut segmentation is used.

Jabari et al. [46] proposes a segmentation and fuzzy rule-based classification for very high resolution satellite images. Classifying very high resolution images is very challenging since there are uncertainties in the position of the object borders. Thus, a fuzzy-rule based classification exhibits more promising solution to this challenging task. Initially, the input image is segmented into shadows, vegetation, and roads using eCognition software. Then triangular and trapezoidal fuzzy functions are used to assign membership values to those segmented regions.

Triangular fuzzy function:

$$\mu_A(x) = \begin{cases} 1 - \frac{|a-x|}{\lambda} & \text{for } 0 \le |a-x| \le \lambda \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

Trapezoidal fuzzy function:

$$\mu_A(x) = \begin{cases} \min\left[2 - 2\left(\frac{|a-x|}{\lambda}\right), 1\right] & \text{for } a-\lambda \le |a-x| \le a+\lambda \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

Instead of crisp threshold, a lingual variable is used by utilizing the aforementioned fuzzy functions. Several parameters (e.g. brightness, density, normalized difference vegetation index, NIR ratio, standard deviation etc.) are used for setting up the fuzzy rules in order to classify the segmented objects. Since the building areas have more complexity compared to the others, they are not properly segmented in the initial segmentation step. Thus, a second level segmentation is performed using FbSP (Fuzzy-based segmentation parameter) optimizer. This segmentation is performed as a supervised manner and thus training data is required. A fuzzy rule is set up using four different parameters (rectangular fit, elliptical fit, area, and shadow neighborhood) in order to classify the segmented regions into buildings. Finally, a contextual check is performed in order to classify some of the unclassified segments if they resemble specific classes. Evaluation is performed using two different datasets: GeoEye and QuickBird imagery. The proposed fuzzy method achieves overall accuracies of 82% and 90% compared to 68% and 42% for crisp method using GeoEye and QuickBird datasets. Evaluation shows that the accuracy for building classification is low compared to other classes. However, the accuracy achieved is still higher than using crisp method.

2.3 Per-field Classification

Traditional automated classification methods consider spectral distribution of the pixels and each pixel is assigned to the most similar spectral class. However, due to intraclass spectral variation, the classification result using a very high spatial resolution image often generates a 'noisy' map. An alternative technique is the per-field classification, which classifies fields (known as 'parcels') as opposed to pixels as independent units. Per-field classifier takes the advantages of GIS data and integrates it with raster data. Feasibility of the integration process using the image processing software packages has helped to gain interests in recent research studies on per-field scene classification. Two of the important per-field classification algorithms are briefly presented as below:

Caprioli et al. [47] proposes a per-field classification by integrating very fine spatial resolution satellite images with topographic data. Contextual information is first extracted using integrated raster data-GIS data system. This approach requires a priori information about the boundaries of fields in the image. Boundaries of the fields are then digitized and registered to the image. Some properties of the pixels lying within boundaries of the field are then used to characterize those fields. Finally, classification based on those fields is performed by developing decision rules

with map algebra. IKONOS multispectral images having 4m spatial resolution are used as raster data, whereas a digital topographic map of scale 1:5000 is used as vector data. The data are integrated in a geographical information system (ARC/INFO). The overall accuracy of their classification procedure is measured as 79%.

Conrad et al. [48] develops a per-field classification of crops using SPOT and ASTER data with an aim to acquire accurate information on crop distribution and crop rotation. At first, field boundaries are derived using very high resolution SPOT 5 data (ground sampling distance 2.5m) and subsequently classified into fields and non-fields using Definiens Developer 7. Later, each field-parcel is classified based on bi-temporal, medium spatial resolution ASTER data (ground sampling distance 15-30m). A pixel-based basic cover type classification is then obtained assuming the tasseled cap indices greenness and brightness as representative of vegetation density and soil wetness, respectively. Within each field-parcel, bi-temporal information of vegetation density and soil wetness is analyzed using ground truth samples and cropping calendar to establish a rule-base for the final classification. Some of the ground truth samples are used for rule generation, while others are used for accuracy assessment. The rule-base implementation is also developed using Definiens Developer 7. The overall accuracy of their per-field classification method is evaluated as 80%, with an exception of very low accuracy for the class having multiple crops. This class has different spatial and spectral characteristics and hence an adequate number of training samples is essential for better discrimination of this class. The use of tasseled cap indices greenness and brightness makes this classification very flexible and transferrable.

2.4 Contextual Classification

Majority of conventional classification techniques only considers pixel's spectral information. Their results often generate a salt and pepper appearance, which leads to misclassification. In remote sensing data, adjacent pixels are mostly correlated both because imaging sensors acquire significant portions of energy from adjacent pixels and because ground cover types generally occur over a region that is large compared to the pixel size [49]. Thus, instead of considering a pixel in isolation as used in per pixel classification, a contextual classification method considers spatially neighboring pixels, which derive more complete contextual information for an accurate classification. Notable contributions in contextual classification include a frequency domain based classification proposed by Farinella et al. [50] and a contextual classification based on Markov random fields proposed by Kuo et al. [51]. Brief descriptions of these algorithms are presented as follows:

Farinella et al. [50] proposes a scene categorization method which builds holistic representation of a scene by exploiting features extracted on discrete cosine transform (DCT) domain. Two local features are used to represent each 8x8 spatial block of an image: local dominant orientation (LDO) of the block and strength of the dominant orientation. These local features are extracted using the corresponding 8x8 DCT block. The LDO is computed as the ratio of sum of DCT coefficients corresponding to the vertical and horizontal frequencies. The local variance of each DCT block (the AC energy) is related to the strength of the LDO. The strength of each block indicates how much the corresponding LDO needs to be taken into account. A holistic representation of an image is then built by analyzing the distribution of these LDOs. kNN algorithm is used to choose the representation parameters like number of bins to consider, threshold on strength to consider only significant orientations etc. Finally, the proposed representation is coupled with term frequency-inverse document frequency (TF-IDF) weighting scheme [52] to select most discriminative LDOs between different classes of a scene. Once the holistic representation is made, a probabilistic model is used for classification. The classifier is trained offline considering a dataset is properly collected. Then a newly observed scene is classified using a simple decision function. For simplicity the algorithm focuses on classification of natural and artificial classes. A basic assumption made here is that the difference between the logarithms of the class conditional density functions is linear in the vector f representing the images using TF-IDF-LDO.

$$\log\big(\mathrm{p(f|Artificial)}\big) - \log\big(\mathrm{p(f|Natural)}\big) = w_0 + w_1 f_{\hat{\theta}_1} + \cdots + w_d f_{\hat{\theta}_d} \quad (16)$$

Where $w_0, w_1, \ldots\ldots, w_d$ are the TF-IDF weights for the d number of bins, and $f_{\hat{\theta}_i}$ is the ith element of the d-dimensional feature vector representing an image. $\{\theta_1, \ldots\ldots, \theta_K\}$ is the K LDOs and $\{A_1, \ldots\ldots, A_K\}$ is the K AC energies extracted from the image.

$$f_{\hat{\theta}_i} = \frac{N(\hat{\theta}_i)}{SN} \quad (17)$$

Where $N(\hat{\theta}_i) = \sum_{A_k \in \Theta_i} \log(A_k)$

$\hat{\theta}_i \in [-90,90]$
$\Theta_i = \{A_k | \hat{\theta}_i < \theta_k \leq \hat{\theta}_{i+1}, A_K > \zeta, k = 1, \ldots, K\}$
$SN = \sum_{n=1}^{d} N(\hat{\theta}_n)$ is the normalization constant
ζ is the threshold used to discard marginal orientations

For a binary classification, an observation (f) is assigned to one of the two possible classes (natural class or artificial class) based on the following function:

$$g(f) = w_0 + w_1 f_{\hat{\theta}_1} + \cdots + w_d f_{\hat{\theta}_d} \quad (18)$$

f is assigned to the class natural if the value of the above function is negative, otherwise f is assigned to the class artificial. Evaluation is performed using number of bins equal to 32 for the LDO representation, and the strength threshold equal to 10% of the maximal A_K extracted from an image under consideration. An average accuracy of 92% is achieved using LDO representation, whereas using TF-IDF-LDO representation gives 94% average accuracy. Evaluation shows that the proposed classification technique can also be applied for classification of other classes.

Kuo et al. [51] proposes a nonparametric contextual classification for hyperspectral images. Since, classification of a high dimensional data requires higher number of training samples, the input images are initially reduced in dimension by using some feature extraction technique. The spatial information of an input image is then obtained on the basis of Markov random field (MRF) and finally combined with nonparametric density estimation obtained by kNN or Parzen. A decision rule to determine which class a given d dimensional pixel belongs to is formulated based on MAP estimation which combines both spectral and spatial information and it could be expressed as:

$$u(s) = \arg\max[p\{u(s)|u(\partial s)\}p\{X(s)|u(s)\}] \quad (19)$$

Where X is a multivariate image having d dimensional pixels, s is a pixel to be classified, and u denotes the field that contains the classification of each pixel in X. The first term indicates the joint probabilities of the class of s and its spatial neighbors ∂s and it is modeled by MRF. Whereas, the second term indicates the class conditional density function, which is modeled by kNN and Parzen density estimations and denoted as Bayesian contextual classification_kNN (BCC_kNN) and Bayesian contextual classification_Parzen (BCC_Parzen) respectively. Evaluation is performed using three different datasets. Discriminant analysis feature extraction (DAFE), and nonparametric weighted feature extraction (NWFE) are used in order to reduce data dimensionality. The performance of the proposed classification method is compared to SVM, kNN, Parzen, BCC_Gaussian classifications. Testing reveals that the BCC_Parzen along with DAFE outperforms the other classification methods.

2.5 Knowledge-based Classification

In addition to the high resolution aerial and satellite images, several ancillary data has readily been available these days. They include digital elevation model, soil map, housing and population density, road network, temperature, and precipitation [10]. These kinds of data can be used in a classification process to generate highly accurate classified map. Scene classification methods which integrate knowledge gained from ancillary data into the classification process are known as knowledge-based classification. Important contributions in knowledge-based classification include modeling vegetation in urban areas from high-resolution aerial imagery proposed by Iovan et al. [53], knowledge-based classification for finding the unstructured road boundaries proposed by Chern [54]. Their methods are discussed as below:

Iovan et al. [53] proposes a vegetation modeling method using multiview high resolution color infrared aerial images. Initially, digital surface model (DSM) is computed using multiview matching algorithm. Cadastral data is used to obtain digital terrain model (DTM). A normalized digital surface model (nDSM), containing the height of above ground objects, is computed as the difference between DSM and DTM. A SVM with linear kernel is then used in order to extract vegetation. For all pixels in the training dataset, the feature vector contains four characteristics, namely, the reflectance values of each pixel in the infrared (IR), red (R), green (G), and blue (B) bands. Local height variance is then computed in those vegetation areas corresponding to the DSM as:

$$V = \sum \frac{(x_{ij} - M)^2}{(n-1)} \quad (20)$$

Where x_{ij} is the height value of pixel (i, j) on DSM, n equals to the number of pixels in a 11 x 11 sliding window, and M is the mean value of the moving window. The height variance values are separated into low and high level values to separate trees from vegetation areas. Once trees are detected, individual tree crown is delineated using a two-step approach: seed point detection and region growing

(RG) algorithm. Subsequently, the crown diameter is measured. Three-height is also computed using the nDSM. Finally, first and second order texture measures are computed to classify different species of trees. Evaluation shows that the SVM based classifier outperforms normalized difference vegetation index (NDVI) with a vegetation extraction accuracy of almost 99% compared to 88% of NDVI. Use of normalized DSM attains a tree detection accuracy of 97%. The proposed RG algorithm shows 78% accuracy in tree crown delineation. Finally, the texture analysis achieves an accuracy ranging from 96% to 100% using feature vectors computed at different color spaces and also using different texture measures.

Chern [54] proposes a knowledge-based region classification for locating unstructured road areas. In order to reduce the processing time, input images are downsampled first. After that an unsupervised color classification method is used to classify pixels of similar color and brightness. A region growing technique is then used to cluster local pixels of similar color into small regions. During this process the color mean value of each region is updated. The small regions are then merged into larger homogeneous regions by applying Fischer's criterion. If μ_1 and μ_2 are the mean values of two small regions and υ_1 and υ_2 are their variance values, the Fischer between these two regions is $|\mu_1 - \mu_2|/\sqrt{\upsilon_1{}^2 + \upsilon_2{}^2}$. The Fischer distance between two regions is expressed as:

$$\text{Fisher}_{\text{dist}} = \sqrt{\text{RFischer}^2 + \text{GFischer}^2 + \text{BFischer}^2} \quad (21)$$

Where $\text{RFischer}, \text{GFischer}$ and BFischer are the Fischer values in the three color channels. If the Fischer distance between two neighboring regions is smallest among all others, then those two regions can be merged. Once the scene is segmented into regions using the above method, a set of rules are proposed based on the knowledge of human judgment of "road" and "off-road" in various road scenes. Using these rules, the segmented regions are mapped into seven categories. Then the borders of the appropriate neighboring regions are considered to find the edge points for the unstructured roads. For each edge point detected in this step, a score is assigned to the corresponding edge using the scene with segmented regions as:

$$\text{edge\_score} = \max(d_1, d_2) \quad (22)$$

Where $d_1 = \Sigma$ |r, g, b diff. between the edge's left and right region|

$d_2 = \Sigma$ |r, g, b diff. between the edge's upper and lower region|

$$r = \frac{R}{(R+G+B)}, g = \frac{G}{(R+G+B)}, b = \frac{B}{(R+G+B)}$$

The boundary edge points are traced and then the merit of each traced edge-link sequence is evaluated based on edge score. A path searching algorithm is used to acquire possible edge-link sequences (road border candidates). Finally, the candidate with smallest deviation from simple approximation curve is regarded as road border. More than hundred road scenes are tested using this proposed method with satisfactory results.

Yi et al. [55] proposes a knowledge-based classification for urban mapping. Their method uses three modules: primitive extraction, explicit knowledge representation, and contextual reasoning. Initially, object oriented segmentation is achieved by combining nearest neighbor classification with canny edge detection and garbo filter. Then semantic networks are used to represent the knowledge for different pattern understanding tasks. These networks are implemented in ERNEST [56]. The nodes in the network represent the concepts expected to appear in the scene with respect to the object, whereas the links indicate the relationship between the concepts. Data structures called attributes are used to describe the properties of concepts. The contextual reasoning integrates a bottom-up and a top-down search algorithm within the semantic network. A goal concept is first selected and then the concepts on lower levels are expanded until the concept on the lowest level is reached. The correspondences of this concept with an extracted segment along with its attributes are calculated. Analysis is then done in a bottom-up approach. Several true color aerial images are used for evaluating the comparative performance of the proposed method and maximum likelihood-based method. The classification result shows that the salt-an-paper appearance in the maximum likelihood-based classified map is significantly improved in the proposed method based map.

2.6 High-level Classification

The conventional classification methods take into account only the physical features on the input images. High-level classifications consider not only the physical attributes, but also the pattern formation of the input data. Thus, high level classification approaches are based on human visual perception which uses both high and low levels classification in

order to identify the patterns according to the semantic meaning of the input data. As a result, these type of classifiers exhibit improvements in the performance accuracies when compared to the traditional classifiers. Significant contributions include scene classification based on an improved standard model feature proposed by Huang et al. [57], network-based high level classification proposed by Silva et al. [58]. These two algorithms are presented as below:

Huang et al. [57] proposes a scene classification algorithm based on improved version of standard model feature (SMF), which works on the basis of human visual cognition system. Even though local descriptors perform satisfactorily in scene classification, most of them suffer either from weak robustness or poor selectivity. Basic SMF, on the other hand offers higher robustness and selectivity. Feature extraction using SMF has four steps namely S1, C1, S2, and C2 [59]. Output vectors of SMF consist of minimum distance between C2 output and learned prior patches. However, patch extraction/learning procedure in basic SMF is completely random, which causes computational complexity of the SMF model. Inspired by the fact that humans usually pay more attention on salient features, an energy function is used to represent patch saliency instead of choosing patches randomly. Since Gabor filter responses are salient, they are used in local maximum operation in the first stage of SMF.

$$G(x,y) = \exp\left(-\frac{({x_0}^2+\gamma^2{y_0}^2)}{2\sigma^2}\right) x\cos(\frac{2\pi}{\lambda}x_0) \qquad (23)$$

Where $x_0 = x\cos\theta + y\sin\theta$ and $y_0 = -x\sin\theta + y\cos\theta$. Filter orientation is θ, aspect ratio is γ, effective width is σ, and wavelength is λ. Based on the idea that the sum of local mask area represents energy density, a local energy measurement is conducted after the S1 stage and the result is used for patch extraction/learning. The procedure of computing local energy distribution is time-consuming and hence dynamic programming is used. Finally, a SVM-based classifier is trained based on the features extracted using the aforementioned improved SMF model. Evaluation is performed using both a public database having 13 categories of natural scenes, and a self-built database. Experimental results show that the proposed model outperforms SIFT and original SMF, particularly in the presence of occlusion and disorder. The proposed model achieves 70% accuracy in the occlusion and disorder database and 81% in the illumination change and view angle change database.

Silva et al. [58] proposes a network-based hybrid classification technique, which combines both low and high level classification. The low level classifier is implemented by a classification technique presently available (kNN, Bayesian decision theory, neural network, decision tree, committee machines), whereas the high level classifier is realized by the extraction of features of the underlying network constructed from the input data. The low level classifier classifies the input data by its physical features or class topologies. The high level classifier measures the compliance of the input data to the pattern formation. The training phase of the high label classifier maps labeled training datasets into a graph/network using a network formation technique. The graph is composed of vertices (□) and edges (E), where each vertex represents one training data. Edges are created between vertices using a combination of ϵ-radius (used for dense regions) and kNN (used for sparse regions) network formation technique. The neighborhood of a training vertex x_i is given by:

$$\mathcal{N}_{\text{tarining}}(x_i) = \begin{cases} \epsilon - \text{radius}\left(x_i, y_{x_i}\right), & \text{if } \left|\epsilon - \text{radius}\left(x_i, y_{x_i}\right)\right| > k \\ \text{kNN}\left(x_i, y_{x_i}\right), & \text{otherwise} \end{cases} \qquad (24)$$

Where y_{x_i} is the class label of the training instance x_i, $\epsilon - \text{radius}\left(x_i, y_{x_i}\right)$ returns the set $\left\{x_j, j \in \mathcal{V}: d\left(x_i, x_j\right) < \epsilon \wedge y_{x_i} = y_{x_j}\right\}$, $\text{kNN}\left(x_i, y_{x_i}\right)$ returns the set containing the k nearest vertices having the same class as x_i. Using this graph formation technique each class will have a unique graph component. In the classification phase the compliance of each unlabeled instance is measured to each class pattern using number of complex topological features. Combining the membership of a test instance with respect to a given class obtained using low label and high level classifiers gives the membership using a hybrid classifier as:

$$M_i^{(j)} = (1-\lambda)T_i^{(j)} + \lambda C_i^{(j)} \qquad (25)$$

Here $M_i^{(j)}$ indicates the membership of test instance x_i with respect to the class j. $T_i^{(j)}$ is the membership produced by low level classifier, whereas $C_i^{(j)}$ is the membership information yielded by a high level classifier, and $\lambda \in [0,1]$ is the compliance term. A test instance receives the label

from the class j that maximizes the above hybrid membership function. Performance is evaluated using five well-known low label classifiers (Bayesian network, weighted kNN, fuzzy C4.5, multilayer perceptron, fuzzy multiclass SVM) on eight distinct datasets. Evaluation shows that classification accuracy is always less when only low label classifiers are used compared to the higher classification accuracy when hybrid classifiers are used.

3 Conclusion

Scene classification is an important task in the field of computer vision. Classification algorithms can be per-pixel, sub-pixel, per-field, contextual, knowledge-based, and high-level. Success of a classification method depends on several factors. Per-pixels classification methods are mostly used in practice. However, they suffer from mixed pixel problem, particularly for medium and coarse spatial resolution data. Sub-pixel classification methods can overcome the problem associated with mixed pixels in medium and coarse spatial resolution data. For fine spatial resolution data, the mixed pixel problem is reduced. However, the presence of object shadows and intraclass spectral variation makes this type of data particularly unsuitable for per-pixel classification methods. Per-field and contextual classification methods are mostly suitable for fine spatial resolution data. When ancillary data is available, knowledge-based methods are optimal as scene classifiers. High-level classification methods, which have highest correlation to the human visual cognition system, seem to offer best result for almost any type of data, but at the cost of computational complexity. Thus, success of a scene classification method depends on several factors and there is no single best classification method. Furthermore, the continuous advent of new classification methods in recent years makes it really difficult to choose a suitable classification method for a specific purpose. Hence, this paper aims at providing a guide for selecting appropriate classification method by giving brief knowledge about different classification methods. This survey also highlights the evaluation and achievements of each of those classification methods.

References:

[1] S. Bernabe, P. Reddy Marpu, A. Plaza, M. Dalla Mura, and J. Atli Benediktsson, "Spectral-Spatial Classification of Multispectral Images Using Kernel Feature Space Representation," *Geoscience and Remote Sensing Letters, IEEE,* vol. 11, pp. 288-292, 2014.

[2] L. Di and V. P. Singh, "An Entropy-Based Multispectral Image Classification Algorithm," *Geoscience and Remote Sensing, IEEE Transactions on,* vol. 51, pp. 5225-5238, 2013.

[3] G. Moser and S. B. Serpico, "Multitemporal region-based classification of high-resolution images by Markov random fields and multiscale segmentation," in *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*, 2011, pp. 102-105.

[4] W. Ying, L. Xiaoyun, W. Zhensong, and C. Wufan, "Multispectral remote sensing image classification algorithm based on rough set theory," in *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*, 2009, pp. 4853-4857.

[5] H. Chih-Cheng, H. Purnawan, and K. Bor-Chen, "Multispectral image classification using rough set theory and the comparison with parallelepiped classifier," in *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*, 2007, pp. 2052-2055.

[6] W. Van and J. Mo, "Multispectral landsat image classification using fuzzy expert systems," in *Automation Congress, 2004. Proceedings. World*, 2004, pp. 21-26.

[7] L. Sanghoon and M. M. Crawford, "Multitemporal classification of image series with seasonal variability using harmonic components," in *Geoscience and Remote Sensing Symposium, 2003. IGARSS '03. Proceedings. 2003 IEEE International*, 2003, pp. 3353-3355 vol.5.

[8] F. Melgani, S. B. Serpico, and G. Vernazza, "Fusion of multitemporal contextual information by neural networks for multisensor image classification," in *Geoscience and Remote Sensing Symposium, 2001. IGARSS '01. IEEE 2001 International*, 2001, pp. 2952-2954 vol.7.

[9] R. Hsuan and I. C. Chein, "A generalized orthogonal subspace projection approach to unsupervised multispectral image classification," *Geoscience and Remote Sensing, IEEE Transactions on,* vol. 38, pp. 2515-2528, 2000.

[10] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *International Journal of Remote Sensing,* vol. 28, pp. 823-870, 2007/03/01 2007.

[11] D. Cerra, R. Muller, and P. Reinartz, "A Classification Algorithm for Hyperspectral Images Based on Synergetics Theory," *Geoscience and Remote Sensing, IEEE Transactions on,* vol. 51, pp. 2887-2898, 2013.

[12] C. H. Genitha and K. Vani, "Classification of satellite images using new Fuzzy cluster centroid for unsupervised classification algorithm," in *Information & Communication Technologies (ICT), 2013 IEEE Conference on*, 2013, pp. 203-207.

[13] T. Yu-Chang, F. Kuo-Tai, and C. Kun-Shan, "A Parallel Differential Box-Counting Algorithm Applied to Hyperspectral Image Classification," *Geoscience and Remote Sensing Letters, IEEE,* vol. 9, pp. 272-276, 2012.

[14] B. Baassou, M. He, S. Mei, and Y. Zhang, "Unsupervised hyperspectral image classification algorithm by integrating spatial-spectral information," in *Audio, Language and Image Processing (ICALIP), 2012 International Conference on*, 2012, pp. 610-615.

[15] S. Yang and C. Zhao, "A fusing algorithm of Bag-Of-Features model and Fisher linear discriminative analysis in image classification," in *Information Science and Technology (ICIST), 2012 International Conference on*, 2012, pp. 380-383.

[16] O. Gordo, E. Martinez, C. Gonzalo, and A. Arquero, "Classification of Satellite Images by means of Fuzzy Rules generated by a Genetic Algorithm," *Latin America Transactions, IEEE (Revista IEEE America Latina),* vol. 9, pp. 743-748, 2011.

[17] C.-C. Hung, S. Kulkarni, and B.-C. Kuo, "A new weighted fuzzy c-means clustering algorithm for remotely sensed image classification," *Selected Topics in Signal Processing, IEEE Journal of,* vol. 5, pp. 543-553, 2011.

[18] Q. Chen and Q. Yin, "Extract agricultural land use information from remote sensing image based on object-oriented classification algorithm," in *Remote Sensing, Environment and Transportation Engineering (RSETE), 2011 International Conference on*, 2011, pp. 4983-4986.

[19] D. Zhao, W. Zou, and G. Sun, "A fast image classification algorithm using support vector machine," in *Computer Technology and Development (ICCTD), 2010 2nd International Conference on*, 2010, pp. 385-388.

[20] N. Ghoggali and F. Melgani, "Automatic Ground-Truth Validation With Genetic Algorithms for Multispectral Image Classification," *Geoscience and Remote Sensing, IEEE Transactions on,* vol. 47, pp. 2172-2181, 2009.

[21] G. Yan, J. F. Mas, B. H. P. Maathuis, Z. Xiangmin, and P. M. Van Dijk, "Comparison of pixel-based and object-oriented image classification approaches—a case study in a coal fire area, Wuda, Inner Mongolia, China," *International Journal of Remote Sensing,* vol. 27, pp. 4039-4055, 2006/09/01 2006.

[22] B. R. Kloer, "Hybrid parametric/non-parametric image classification," in *Technical Papers, ACSM-ASPRS Annual Convention,* 1994, pp. 307-316.

[23] T. M. Lillesand, R. W. Kiefer, and J. W. Chipman, *Remote sensing and image interpretation*: John Wiley & Sons Ltd, 2004.

[24] R. A. Gopinath, "Maximum likelihood modeling with Gaussian distributions for classification," in *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, 1998, pp. 661-664.

[25] A. B. Salberg, "Land Cover Classification of Cloud-Contaminated Multitemporal High-Resolution Images," *Geoscience and Remote Sensing, IEEE Transactions on,* vol. 49, pp. 377-387, 2011.

[26] L. Xin and I. S. Yetik, "A Maximum Likelihood Classification method for image segmentation considering subject variability," in *Image Analysis & Interpretation (SSIAI), 2010 IEEE Southwest Symposium on*, 2010, pp. 125-128.

[27] M. P. Sampat, A. C. Bovik, J. K. Aggarwal, and K. R. Castleman, "Supervised parametric and non-parametric classification of chromosome images," *Pattern Recognition,* vol. 38, pp. 1209-1223, 2005.

[28] L.-B. Ren, H. Chang, and Y. Yi, "An Improved Binary Tree SVM Classification Algorithm Based on Bayesian," in *Information Processing, 2009. APCIP 2009. Asia-Pacific Conference on*, 2009, pp. 178-181.

[29] D. Barbella, S. Benzaid, J. M. Christensen, B. Jackson, X. V. Qin, and D. R. Musicant, "Understanding Support Vector Machine Classifications via a Recommender System-Like Approach," in *DMIN*, 2009, pp. 305-311.

[30] S. Patra and L. Bruzzone, "A Fast Cluster-Assumption Based Active-Learning Technique for Classification of Remote Sensing Images," *Geoscience and Remote Sensing, IEEE Transactions on,* vol. 49, pp. 1617-1626, 2011.

[31] V. Risojevic and Z. Babic, "Fusion of Global and Local Descriptors for Remote Sensing Image Classification," *Geoscience and Remote Sensing Letters, IEEE,* vol. 10, pp. 836-840, 2013.

[32] W. Zilei, F. Jiashi, Y. Shuicheng, and X. Hongsheng, "Linear Distance Coding for Image Classification," *Image Processing, IEEE Transactions on,* vol. 22, pp. 537-548, 2013.

[33] R. Ablin and C. H. Sulochana, "A Survey of Hyperspectral Image Classification in Remote Sensing."

[34] Q. Sami ul Haq, L. Tao, and S. Yang, "Neural network based adaboosting approach for

hyperspectral data classification," in *Computer Science and Network Technology (ICCSNT), 2011 International Conference on*, 2011, pp. 241-245.

[35] Y. Ren and B. Wang, "Fast human detection via a cascade of neural network classifiers," in *Wireless, Mobile and Multimedia Networks (ICWMNN 2010), IET 3rd International Conference on*, 2010, pp. 323-326.

[36] B. Majhi, J. Satpathy, and M. Rout, "Efficient recognition of Odiya numerals using low complexity neural classifier," in *Energy, Automation, and Signal (ICEAS), 2011 International Conference on*, 2011, pp. 1-4.

[37] D. Pop, C. Jichici, and V. Negru, "A combinative method for decision tree construction," in *Symbolic and Numeric Algorithms for Scientific Computing, 2005. SYNASC 2005. Seventh International Symposium on*, 2005, p. 5 pp.

[38] M. Pal and P. M. Mather, "A comparison of decision tree and backpropagation neural network classifiers for land use classification," in *Geoscience and Remote Sensing Symposium, 2002. IGARSS'02. 2002 IEEE International*, 2002, pp. 503-505.

[39] B. Thangaparvathi, D. Anandhavalli, and S. Mercy Shalinie, "A high speed decision tree classifier algorithm for huge dataset," in *Recent Trends in Information Technology (ICRTIT), 2011 International Conference on*, 2011, pp. 695-700.

[40] T. Srinivasan, M. Sathish, V. G. Krishna, and V. Krishnamoorthy, "TC-ID3: A TESTCODE Based ID3 Classifier for Protein Coding Region Identification," in *Computational Intelligence for Modelling, Control and Automation, 2006 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference on*, 2006, pp. 110-110.

[41] F. Yang, H. Jin, and H. Qi, "Study on the application of data mining for customer groups based on the modified ID3 algorithm in the e-commerce," in *Computer Science and Information Processing (CSIP), 2012 International Conference on*, 2012, pp. 615-619.

[42] A. Taherkhani, "Recognizing sorting algorithms with the C4. 5 decision tree classifier," in *Program Comprehension (ICPC), 2010 IEEE 18th International Conference on*, 2010, pp. 72-75.

[43] Z. Zili, Q. Qiming, G. Junping, D. Yuzhi, Y. Yunjun, W. Zhaoqiang, *et al.*, "CART-Based Rare Habitat Information Extraction For Landsat ETM+ Image," in *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International*, 2008, pp. III-1071-III-1074.

[44] F. Bovolo, L. Bruzzone, and L. Carlin, "A novel technique for subpixel image classification based on support vector machine," *Image Processing, IEEE Transactions on*, vol. 19, pp. 2983-2999, 2010.

[45] S. PRABHU and D. TENSING, "EFFICIENT TECHNIQUE FOR THE CLASSIFICATION OF SATELLITE IMAGES USING FUZZY RULE CLASSIFIER," *Journal of Theoretical & Applied Information Technology*, vol. 53, 2013.

[46] S. Jabari and Y. Zhang, "Very High Resolution Satellite Image Classification Using Fuzzy Rule-Based Systems," *Algorithms*, vol. 6, pp. 762-781, 2013.

[47] M. Caprioli and E. Tarantino, "Accuracy assessment of per-field classification integrating very fine spatial resolution satellite imagery with topographic data," *Journal of Geospatial Engineering*, vol. 3, pp. 127-134, 2001.

[48] C. Conrad, S. Fritsch, J. Zeidler, G. Rücker, and S. Dech, "Per-field irrigated crop classification in arid Central Asia using SPOT and ASTER data," *Remote Sensing*, vol. 2, pp. 1035-1056, 2010.

[49] R. Khedam and A. Belhadj-aissa, "A general multisource contextual classification model of remotely sensed imagery based on MRF," in *Remote Sensing and Data Fusion over Urban Areas, IEEE/ISPRS Joint Workshop 2001*, 2001, pp. 231-235.

[50] G. M. Farinella and S. Battiato, "Scene classification in compressed and constrained domain," *Computer Vision, IET*, vol. 5, pp. 320-334, 2011.

[51] B.-C. Kuo, C.-H. Chuang, C.-S. Huang, and C.-C. Hung, "A nonparametric contextual classification based on Markov random fields," in *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009. WHISPERS'09. First Workshop on*, 2009, pp. 1-4.

[52] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information processing & management*, vol. 24, pp. 513-523, 1988.

[53] C. Iovan, D. Boldo, and M. Cord, "Detection, Characterization, and Modeling Vegetation in Urban Areas From High-Resolution Aerial Imagery," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 1, pp. 206-213, 2008.

[54] M.-Y. Chern, "Knowledge-based region classification for rural road area in the color scene image," in *Networking, Sensing and Control, 2004 IEEE International Conference on*, 2004, pp. 891-896.

[55] X. Yi, X. Jia, and D. Tien, "A Knowledge Based Classification for Urban Mapping Using High Resolution Remote Sensing Data," in *Digital Image*

Computing Techniques and Applications, 9th Biennial Conference of the Australian Pattern Recognition Society on, 2007, pp. 586-591.

[56] H. Niemann, G. F. Sagerer, S. Schroder, and F. Kummert, "Ernest: A semantic network system for pattern understanding," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 12, pp. 883-905, 1990.

[57] H. Kaiqi, T. Dacheng, Y. Yuan, L. Xuelong, and T. Tieniu, "Biologically Inspired Features for Scene Classification in Video Surveillance," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 41, pp. 307-313, 2011.

[58] T. C. Silva and Z. Liang, "Network-Based High Level Data Classification," *Neural Networks and Learning Systems, IEEE Transactions on*, vol. 23, pp. 954-970, 2012.

[59] T. Serre, L. Wolf, and T. Poggio, "A new biologically motivated framework for robust object recognition," 2004.