International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 8 (2015) pp. 21185-21197

http://www.ripublication.com

Classification of Images Using Active Learning

E. Shehnaz and D. Venkataraman
Dept of Computer Science and Engineering
Amrita School of Engineering
Amrita Vishwa Vidyapeetham (Amrita University)
shehnu27@gmail.com and d_venkat@cb.amrita.edu

Abstract

Nowadays image classification has become an evolving area. Various methods are used for different classification. In image classification, there are many different machine learning algorithms used. In the database the retrieve images which are stored that is used to find the resemblance in the query image, then the CBIR allows the user to represent a query image. From image database it will retrieve all the images which belongs to a particular category and occurs problem in the search category. To achieve higher image accuracy within less execution time, classification of images is an intricate process which is essential to classify, organize and access them using an efficient, faster and easy way. The main motivation is to examine the performance of algorithm and check whether that algorithm is better suited for classification. By finding the accuracy of the classification, time and cost complexity can be low. The features can be extracted for both training and testing process in different set of images. The visual features of image such as texture, shape, color, etc. is a technique used by CBIR in which the user will search the image from large image database and represent the image in the form of a query according to the request of the user. The main objective is to classify images using active learning. By analyzing the active learning contribution in CBIR, different classification strategies are explained and compared.

Keywords: Image Classification; Content based image retrieval; Feature Extraction; Active Learning; Support Vector Machine; Fuzzy Neural Networks; Bayesian Classifier; Fuzzy Logic.

Introduction

Content-Based Image Retrieval (CBIR) system has developed and attracted to many researchers nowadays. From image database it will retrieve all the images which

belongs to a particular category and occurs problem in the search category. Image classification is one of major problems in image processing [1].

In pattern recognition and classification, SVM have the ability to process. In this classifier, there will be a set of points that is belonging to any of the two classes, then a linear SVM will find the hyperplane. In this, the hyperplane leaves the highest possible fraction of points of the similar class on the similar side, that will increase the distance of any of the two classes from the hyperplane. Accordingly, the hyperplane will reduce the risk of misclassifying samples of the query set. When we apply traditional machine learning algorithms it is difficult for the database on image classification because there will be a large number of images and that will describe an image [1,2].

In a wide variety of problem space, Fuzzy logic has been used. In the fuzzy set there can be a set of the elements which have degrees of membership. We can assume that those element can be a partial member or full member. i.e, Assigning the membership value to an element such as it can be 0, 1 or any value in-between. Membership function is nothing but, in a fuzzy set the degree of an element's membership can be defined in mathematical function [17]. The motivation of this work and to solve the problem of image classification in fuzzy logic is to compare the results of supervised and fuzzy classification.

Fuzzy neural networks is a learning machine where it helps for finding the parameters of both fuzzy sets and fuzzy rules (fuzzy system). In many ways Neural networks can be called as collective models, neural nets, neuromorphic systems, connectionist models, parallel distributed processing models and artificial neural networks. A Neural Network is an interconnected elements, units or nodes [15,17]. The functionality of this network is loosely based on the animal neuron. From a set of training patterns, the processing ability of the network is stored in the inter-unit connection strengths, or weights, obtained by a process of adaptation [21].

In a supervised learning algorithm, Naive Bayes classifiers can be trained in a efficient manner. In several applications, by using the method of maximum likelihood we have to calculate the parameter estimation for naive Bayes models. An advantage of naive Bayes is to estimate the parameters such as means and variances of the variables will require a small amount of training that is necessary for classification. Assuming only the variances of the variables should be determined not the entire covariance matrix because of the independent variables [22].

The k-nearest neighbor classifier (k-NN) is a supervised learning classifier, in a given dataset based on closest training set elements it classify test cases. An object being allotted to the class among its k nearest neighbors where the object will be classified with the help of many votes of its neighbors. KNN classifier becomes popular because of their simplicity and good performance [24]. It will be expensive when a number of training set is given or dimension of training set is high in order to classify a given pattern to search the entire training set will be the major defect in KNN classifier.

In this, there are various active learning methods have been explained. Nowadays, there are different machine learning algorithms that are used for image classification. In addition to it, there are various methods to extract color, shape and texture [1]. This

helps us to approach several methods for image classification based on shape, color and texture is presented. From a database on image classification, traditional machine learning algorithm will be different because there will be a large number of images and that will describe an image.

Classification

A classification technique is a well ordered approach from an input data set for building classification models. The relationship between the attribute set and class label of the input data that will fit to identify the model in the learning algorithm. To predict the class labels of records that fits the input data in which the algorithm should be generated from this model. The main objective of the learning algorithm is to build models with good generalization capability. This classification makes the models more accurate and it helps to predict the class labels of previously unknown records. To build a classification model a training set is used to applied to the test set which consisting of number of records with unknown class labels. Based on the counts of test records correctly and incorrectly predicted by the model, the performance of a classification model is evaluated. Using a performance metric such as accuracy, those information will be convenient when comparing the performance of different models and by summarizing with a single number.

Block Diagram

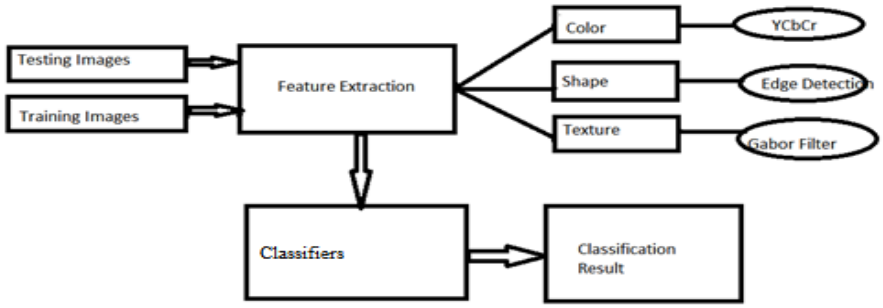

Figure 1: Block Diagram of Active Learning

Active Learning

Active learning is an active research area within the machine learning algorithm which acts as a strong impact to process data and it will be done before classification. The training data will be possible when the classifications depend on the most informative pixels and will be effective. Contribution of both the goals will obtain the respective labels and would be most informative when unlabeled data has the capability to provide active learning. By using active learning methods, the problem

of interactive sampling of training pixels will be used for classification of images with the help of SVM. All supervised learning algorithm is an active learning method.

Feature Extraction

Feature means characteristics of object. Feature extraction refers to dimensionality reduction of the object. It plays an important role in image processing. Features are classified into three types in image processing, i.e low, middle and high. Low level features are color, texture and middle level feature is shape and high level feature is semantic gap of objects. There are different techniques for extracting the features. The extraction of color feature can be done by using Color Histogram and Color correlogram [17]. With the help of wavelet Transform, texture and shape can be extracted. The fundamental step for image classification is feature extraction. It can identify and extract useful feature for classification. Features can be composed by using natural images. One of the basic feature is color and it explains a particular color space or model like RGB, LUV, HSV etc.

Methods of Classification

Classification Using Support Vector Machine

SVM are supervised learning algorithms that is used for analyzing data and for recognizing patterns. The representation of SVM model can be mapped as points in space, so that different category samples can be divided by a clear gap. The mapping of new samples can be done into the same space which is based on a category that may belong to side of the gap. In addition for mapping the inputs into high dimensional spaces, the kernel trick can perform a non-linear classification instead of performing linear classification [1].

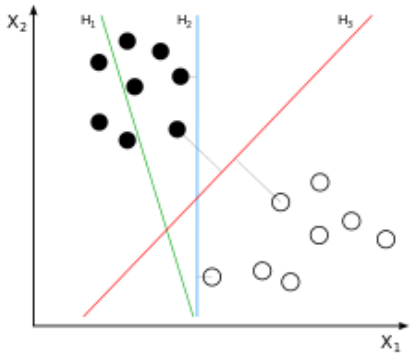

Figure 2: Linear SVM

In Fig: 2, Using classification, SVM builds a sequence of hyperplanes in a high dimensional space. In any of the classes, if the hyperplane has the largest distance to the nearest data point then there will be a better seperation if the margin is larger and

the generalization error is lower. If the sets are not linearly seperable in the space then the problem will occur in a finite dimensional space. To make the seperation easier, we should map the lower-dimensional space into a higher-dimensional space. Instead of using the variables, the dot products can be ensured and computed by mapping the SVM, by defining the kernel function to suit the problem. In the higher-dimensional space, the hyperplane can be as the collection of points in which the dot product with a vector in the space is a constant [1,2]. A linear SVM will find the hyperplane, when there will be a set of points that must belonging to any of two classes. To increase the distance of any of the two classes in the hyperplane, those hyperplane leaves the highest possible fraction of points of the similar class on the similar side. Accordingly, the hyperplane will reduce the risk of misclassifying samples of the query set. By choosing the best hyperplane that will separate irrelevant and relevant vectors for increasing the size of the margin between the two classes [1].

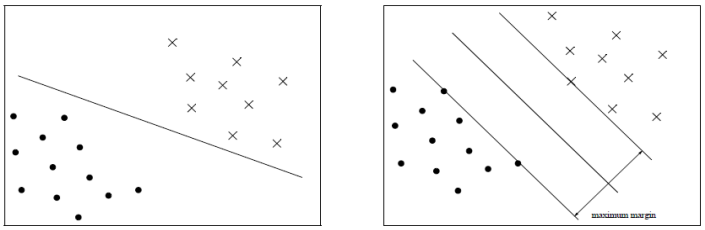

Figure 3: Optimal separating hyperplane

Given a set of training samples (xi, yi)1≤i≤N where xi ∈ Rn and that is belonging to class labeled by yi ∈ {−1, 1}. Then w and b can be written as,

$$y_i(w.x_i+b)>0, i=1,\ldots,N \quad (1)$$

If the hyperplane satisfies, then the set is said to be linearly separable. Rescale b and w so that we get,

$$\min y_i(w\cdot x_i+b)\geq 1, i=1.,N \quad (2)$$

$1\leq i\leq N$

The point which is closest to the hyperplane,

$$y_i(w\cdot x_i+b)\geq 1 \quad (3)$$

Optimal separating hyperplane (OSH) is nothing but to find the maximum distance to the closest point is $1/\|w\|$, then

$$\text{minimize}\tfrac{1}{2}w.w \quad (4)$$

$2/\|w\|$ is known as the margin. For maximizing the margin, the hyperplane will be separated by using OSH. If the margin is smaller then there will be more difficult circumstances and if the margin is larger then there will be better generalization [2].

Classification using Fuzzy Logic

Fuzzy logic distribute with logic term like think logically that is rather than fixed and exact. This variables may have a truth values that ranges between 0 and 1 when comparing with traditional binary sets. The intension of partial truth is the value of truth ranges between fully true and fully false. The propositions may have a value of truth or falsity which are permitted by classical logic. The notion can be absolute, immutable, mathematical truth. It is the part of representing and reasoning the notion of truth where possible outcomes can be aggregated into a dimensional spectrum [15]. In fuzzy logic, knowledge-based systems can be applied from human reasoning capabilities that requires an inference morphology. In theory, human cognitive processes such as thinking and reasoning is used for capturing the uncertainties and provides a mathematical strength. For representing the meaning of fuzzy concepts, lack of knowledge representation for conventional approach is required [17]. For dealing with the representation of common sense knowledge approach based on first order logic and classical probability theory, it will not provide an appropriate conceptual framework and those knowledge is applied for both lexically imprecise and non categorical. The need for a conceptual framework which can address the issue of uncertainty and lexical imprecision is motivated in large measure for the development of fuzzy logic [15,17].

Classification using Naive Bayes

Families of a probabilistic classifiers are classified as naive Bayes in machine learning which are independent between the features. In a learning algorithm, it requires the number of features must be linear to the number of parameters which are highly scalable. As it is a supervised learning algorithm, this classifiers can be trained efficiently for some of the probability models. In this algorithm, conditional probability is calculated. It formulate posterior, prior, maximum likelihood and neighbours for estimation. To estimate the means and variances that are used for classification will requires a less number of training samples is the advantage of naive Bayes [22]. For calculating conditional probabilities Bayes theorem allows from the probability theory. To reflect the influence of one event on the probability of another event, conditional probabilities are used. The term generally used in Bayes theorem are prior probability and posterior probability. The prior probability of a event is the original probability that is obtained before any additional information is obtained. Using Bayes' theorem, this can be written

$$p(C|F1,\ldots,Fn)=\frac{p(C)\,p(F1,\ldots,Fn|C)}{p(F1,\ldots.Fn)} \quad (5)$$

where, $p(C|F1,\ldots,Fn)$ is a conditional model.

C is a class variable or classes.

F is a features from F1….Fn.

Consider the numerator of that fraction but not denominator because it does not depend on class and features, so it becomes constant. Then the numerator will be equal to the joint probability model,

$$p(C,F1,\ldots,Fn) \quad (6)$$

By rewriting the above equation as follows,

$$p(C,F1,\ldots,Fn)=p(C)\ p(F1|C)\ldots\ p(Fn|C,F1,\ldots,Fn-1) \quad (7)$$

Assume that each feature Fi is conditionally independent from other feature Fj.

$$p(Fi|C,Fj,Fk,Fl)=p(Fi|C), \quad (8)$$

Thus, the joint probability model can be written as

$$p(C|F1,\ldots,Fn)\propto p(C,F1,\ldots,Fn)$$

$$\propto p(C)\prod i=1np(Fi|C) \quad (9)$$

The conditional probability distribution over the class variable C is:

$$p(C|F1,\ldots,Fn)=1/Zp(C)\prod p(Fi|C) \quad (10)$$
$$i=1$$

where Z=p(F1,…,Fn) is a evidence for scaling factor on the features F1,…,Fn is a constant. By combining the classifier with the decision rule, then the rule will become maximum a posteriori or MAP decision rule [22]. Based on applying Bayes' theorem with strong independence assumptions, a naïve Bayes classifier is a probabilistic classifier is performed. When represented as a Bayesian network, a naïve Bayes classifier has the structure depicted in Fig: 4.

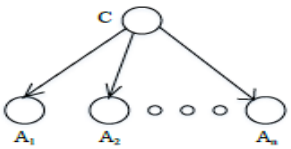

Figure 4: Structure of the Naive Bayesian Network

Classification using Fuzzy Neural Networks

Fuzzy neural networks is a learning machine where it helps for finding the parameters of both fuzzy sets and fuzzy rules (fuzzy system). From neural network theory a data-driven learning method can be trained based on an underlying fuzzy system [15,17]. In many ways Neural networks can be called as collective models, neural nets, neuromorphic systems, connectionist models, parallel distributed processing models and artificial neural networks. A Neural Network is an interconnected elements, units or nodes. The functionality of this network is loosely based on the animal neuron. From a set of training patterns, the processing ability of the network is stored in the inter-unit connection strengths, or weights, obtained by a process of adaptation [21] is shown in Fig: 5.

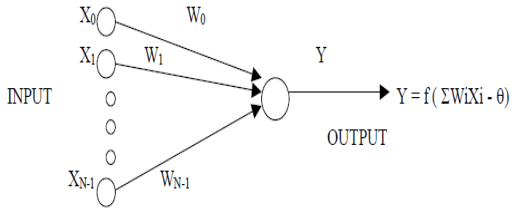

Figure 5: Structure of Neural Networks

Architecture Structures of Neural Networks

Neural networks are different in their structures or topology but not only different in their learning processes. This network is classified into recurrent and nonrecurrent. It has broadly divided the network architectures into the following three classes [21].

Single-Layer Perceptrons (Feed Forward Networks)

It was the first and simplest learning machines that are trainable in the neural network.

The perceptron refers the class of two-layer feed forward networks [21].

1. From the inputs, first-layer units must have fixed function with fixed connection weights.
2. When the first layer links to the second layer of outputs whose connection weights are learnable.

With the set of training patterns, the comparison of actual outputs with desired outputs is involved when the model of training in perceptrons is supervisory. An output layer of neurons can be projected from an input layer of source nodes. While training the patterns, the LMS algorithm can be used for calculation. It is a perceptron that have only one layer of variable weights and one layer of output neurons. Single-layer referring to the output layer of computation nodes.

Multi-Layer Perceptrons (Feed Forward Networks)

This network consists of a set of sensory nodes that form the input layer, one or more hidden layers and an output layer of computation nodes [21]. Some of the characteristics of a multi-layer perceptron are:

1. The activation function which is nonlinear is included from the model of each neuron in the network.
2. This network includes one or more layers of hidden neurons but not the input or output layers of the network.
3. By determining the synapses of the network, it exhibits a high degree of connectivity.

With the error back propagation algorithm, Multi-layer feed forward networks will be trained in supervised manner. The different layers of the network: a forward pass and a backward pass are used under the concept of error back propagation algorithm. The effect propagates through the network layer by layer in the forward pass, when an input vector is applied to the sensory node of the network. Lastly, the actual response of the network will be produced from a set of outputs [21]. During the forward pass, the synaptic weights of the networks are all stable. The output layer starts during the

backward pass, when error signals passes leftward through the network and the local gradient for each neuron will be recursively computed. An error-correction rule should be adjusted from the synaptic weights of the network. The structure of an unadorned multilayer perception network is shown in Fig: 6.

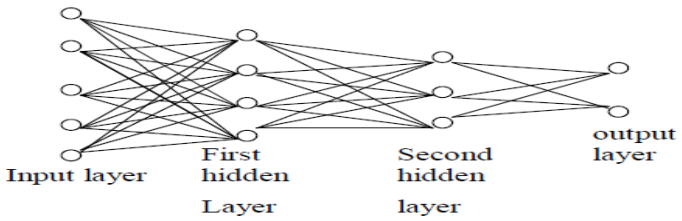

Figure 6: Structure and features of MLP

Back Propagation Algorithm

The one of the simple training algorithms collectively termed gradient descent is the Back propagation algorithm. The main idea of this algorithm is to adjust weights from the neural by decreasing the total least square error of the network. To provide this algorithm, Gradient descent are also called as the method of steepest descent is used for calculating the weights [17,21].

Step 1 : Firstly, initialize with weights and offsets and then set small random values for all weights and node offsets.

Step 2 : Secondly, present the required input nodes and desired output nodes. The desired output is 1. From a training set, the input can be new on each trial or samples.

Step 3 : By using the sigmoid nonlinearity formulas calculate the actual outputs.

Step 4 : To get the desired output adapting the weights for each set.

Step 5 : Then repeat step 2 until the iteration stops.

In this algorithm it is better to choose the initial weights smaller. If we choose the initial weights larger then the network will be untrainable [21].

Recurrent Networks (Feed Backward Networks)

When one or more feedback loops are used in Neural networks are referred to as recurrent networks. From a feed forward neural network, it has one feedback loops when this network distinguishes. This system provides temporal and spatial behaviors, such as stable and unstable fixed points and limit cycles, and chaotic behaviors [21]. To model certain cognitive functions, these behaviors can be utilized. The functions are associative memory, self-organizing maps, unsupervised learning and temporal reasoning. A major role of recurrent networks is the feedback. There are two ways for feedback: one is local feedback that is used inside the network at the level of a single neuron and the other is global feedback which describes the whole network. To extend the standard back propagation algorithm to recurrent networks, Recurrent back propagation (RBP) is used. The dynamics of the network for computation is emphasized when the algorithm plays a nonrule-base continuous-time formalism. There is a limitation while understanding the recurrent networks with asymmetric

connections. For analyzing the dynamics of general nonlinear system, there will be a limited mathematical tools for reflection [21].

K-Nearest Neighbor Classifier

In pattern recognition, The k-nearest neighbor classifier (k-NN) is a supervised learning classifier, in a given dataset based on closest training set elements it classify test cases. The function is locally estimated and all computation is postponed until classification under the category of instance-based learning, or lazy learning which belongs to k-NN. An object being allotted to the class among its k nearest neighbors where the object will be classified with the help of many votes of its neighbors. The object is simply assigned to the respective class of its nearest neighbor when k=1. KNN classifier becomes popular because of their simplicity and good performance [24]. It will be expensive when a number of training set is given or dimension of training set is high in order to classify a given pattern to search the entire training set will be the major defect in KNN classifier [23]. K-nearest neighbors uses the local neighborhood to obtain prediction. A distance function can be calculated with the help of Euclidean distance, Hamming distance and Mahnattan distance. The distance between the test set and all the training sets will be computed, and a list of k nearest training set examples is maintained. The class to which maximum number of k-nearest training set belongs to, is assigned to test set [23] which is shown in Fig : 7. In continuous variables for calculating distance metric, Euclidean distance is used where as for discrete variables such as the overlap metric (or Hamming distance) is performed. If the distance metric is learned with specialized algorithms such as Large Margin Nearest Neighbor, then the classification accuracy of *k*-NN can be improved significantly [23,25].

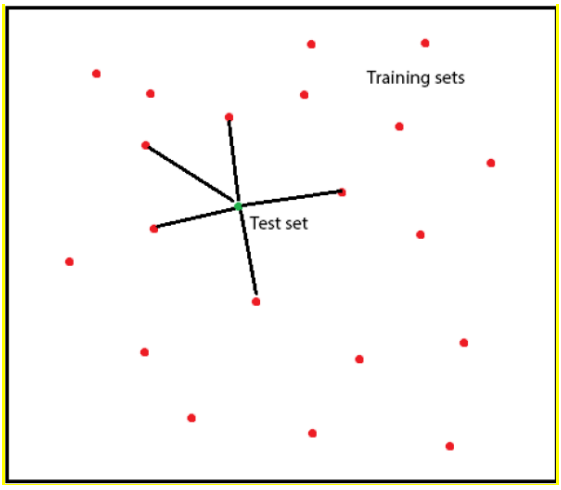

Figure 7: Diagrammatic representation of KNN

Algorithm:

- Let Training set, T = {xi}^n where i=1
- Let x be a pattern with unknown class label

- KNN = Ø
- For each t=T
  if |KNN| < K
  KNN = KNN ∪ T
  else
  find a pattern x' ∈ KNN such that
  dist(x, x') > dist(x, t)
- KNN = KNN – x'
- KNN = KNN ∪ {t}

When the class distribution is skewed "majority voting" classification occurs is the major drawback in this classification. The distance from the test point to each of its k nearest neighbors should take in account and there is a way to overcome this problem is to weight the classification. A weight proportional to the inverse of the distance from that point to the test point should be multiplied from the class of each of the k nearest points. There is other way to overcome this problem is data abstraction in data representation.

Observation

The proposed method on image classification inferred that the performance of different algorithm is to be analyzed. By calculating the performance metric for each classifiers, results will be known. By applying different classification technique in both training and testing, its accuracy can be performed so that we can easily find the maximum matching between the training and testing set. The accuracy of classification for the different set of classes of images is provided and examine which algorithm gives better performance while classifying the images.

No	Methods	Remarks
1	Support Vector Machines	This inferred that even the images are blur and noisy svm gives better result while comparing the algorithms with less computational time.
2	Neural Networks	This inferred that it can handle noisy or missing data when it works with large number of parameters and it also provide general solutions with good predictive result.
3	Bayesian Classifier	This classifiers inferred that it exhibits high accuracy and speed when applied to large databases and produce better result.
4	Fuzzy Logic Classifier	It inferred that already created simple rules and image classification will be equal even with less time consuming. Different conditions during image capture must be taken into account. Considering the level of classification accuracy, fuzzy logic can be satisfactory used for image classification.
5	k-Nearest Neighbor Classifier	This inferred that It is simple and intuitive that can be applied to the data from any distribution. If the number of samples is large enough then the accuracy of classification is good.

Conclusion

In the above, extracting and classifying of Color, shape and texture information for different images was presented. Generally, Gabor Filter methods are used for extracting the texture features, detecting edges for shape analysis and the use of different channels in the color image were experimentally tested and evaluated. By extracting the features in both training and testing images, feature vector is calculated. The results are shown with the help of Corel and Caltech databases by using classification techniques. The results also confirm that the proposed method achieved by matching the features vectors of both database and query images and also performance metrics are calculated. Finally, the applications of several combinations for feature extraction and also classification will be researched.

References

[1] C. Cortes and V. Vapnik, "Support vector network," Machine learning, vol. 20, pp. 1–25, 1995.

[2] E. Osuna, R. Freund, and F. Girosi, "Improved training algorithm for support vector machines", in IEEE NNSP'97, Amelia Island, FL, pp. 24-26, September 1997.

[3] S. B. Park, J. W. Lee and S. K. Kim, "Content based image classification using a neural network", Pattern Recognition Letters, Vol. 25, no. 3, pp. 287-300, 2004.

[4] Mao, K. Z.: "Feature Subset Selection for Support Vector Machines through Discriminative Function Pruning Analysis". IEEE Trans. Systems, Man and Cybernetics, vol. 34, no.1, pp. 60 -67, 2004.

[5] Haralick and Shanmugan, "Textural Features for Image Classification". IEEE Transactions on Systems, Man and Cybernatics, vol. SMC 3, no 6, pp. 610, November 1973.

[6] W. Niblack, R. Barber, W. Equitz, M. Flickner, D. Glasman, D. Petkovic, and P. Yanker, "The qbic project: Querying image by content using color, texture, and shape," SPIE, vol. 1908, pp. 173–187, February 1993.

[7] O. Chapelle, S.P. Haffner and V. Vapnik, Support vector machines for histogram-based image classification. IEEE Trans. on Neural Networks, vol. 10, no.5, pp. 1055–1064, May 1999.

[8] Hsu C.-W. and C.-J. Lin, "A comparison of methods for multiclass support vector machines," IEEE Trans. Neural Networks, vol. 13, pp. 415–425, Mar. 2002.

[9] Melgani F. and L. Bruzzone 2004. Classification of Hyperspectral Remote Sensing Images With Support Vector Machines. IEEE Transactions On Geoscience And Remote Sensing, Vol. 42, No. 8, August 2004.

[10] G.M. Foody, A. Mathur, A relative evaluation of multiclass image classification by support vector machines, Geoscience and Remote Sensing, IEEE Transactions, 2004, Vol. 42, No. 6, pp.1335-1343.

[11] Boiman, O., Shechtman, E. and Irani, M., In defense of Nearest- Neighbor based image classification, in IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp.1-8, 2008.

[12] O. Chapelle, P. Haffner, and V. Vapnik. Svms for histogram based image classification. IEEE Transactions on Neural Networks, 9, 1999.

[13] T. Randen, J.H. Husoy, "Filtering for Texture Classification: A Comparative Study," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 21, no. 4, pp. 291-310, 1999.

[14] K.I. Kim, K. Jung, S.H. Park, H.J. Kim, "Support Vector Machines for Texture Classification," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 11, pp. 1542-1550, 2002.

[15] R. Haralick, K. Shangmugam, L. Dinstein, "Textural Features for Image Classification," IEEE Transactions on Systems, Man, and Cybernetics, vol. 3, pp. 610 621, 1973.

[16] B. Sch¨olkopf, K. Sung, C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik, "Comparing support vector machines with gaussian kernels to radial basis function classifiers," A.I. Memo No. 1599, MIT, 1996.

[17] N. S. Manish, M. Bodruzzaman and M. J.Malkani, "Feature Extraction using Wavelet Transform for Neural Network based Image Classification", Proc. Int. Conf. on System Theory, Morgantwon, pp. 412-416, 1998.

[18] P. McGuire and G. M. T. D `Eleuterio, "Eigen pixels and a Neural-Network Approach to Image Classification", IEEE Trans. Neural Networks, Vol.12, No.3, pp. 625-635, 2001.

[19] Gardner MJ, Altman DG, " Calculating confidence intervals for proportions and their differences". In: Gardner MJ, Altman DG, eds. Statistics with confidence. London: BMJ Publishing Group, 1989:28-33.

[20] Bose, N. K.; & Liang P. (1996). Neural network fundamentals with graphs, algorithms and applications. McGraw-Hill, Inc.

[21] L.Fei-Fei and P. Perona, "A Bayesian hierarchical model for learning natural scene categories," IEEE Conf. Computer Vision and Pattern Recognition.,pp.524-531, March.2005.

[22] T.M. Cover and P.E. Hart, Nearest Neighbor Pattern Classification, IEEE Trans. Information Theory, vol. 13, no. 1, pp. 21-27, 1967.

[23] V. Suresh Babu and P. Viswanath. Weighted k-nearest leader classier for large data sets. In PReMI, pages 1724, 2007.

[24] P.E. Hart, The Condensed Nearest Neighbor Rule, IEEE Trans. Information Theory, vol. 14, no. 3, pp.515-516, 1968.

[25] N. Friedman, D. Geiger and M. Goldszmidt, "Bayesian Network Classifiers.," Machine Learning, vol. 29, pp. 131-163, 1997.