SVM approach for color image classification using shape and texture

Gayatri Vijay Patil , Pranali vivekanand More, Nandinee Dattatrey Pawar

Department of Computer Engineering, Late G.N.S.C.O.E, Nasik, India.

ABSTRACT

We propose a novel approach for content based plant image classification using Shape, Color and Texture features using Support Vector Machine (SVM) intended mainly for Agriculture, medical industry, botanical gardening and cosmetic industry. Farmers are suffering from the problem rising from various types of plant traits/diseases. Sometimes plant's doctors are also unable to recognize the disease that results in lack of identification of right type of disease and this leads to crop spoil if not taken care of at right time. The most significant part of research on plant disease to identify the disease based on CBIR (content based image retrieval) that is mainly concerned with the accurate detection of diseased plant. It has significant perspective in field of agriculture. Sample technique for identify plant disease. In this paper, diseases plant image classification is done on features extracted from histograms of color components and linear, non-linear SVM approach

Keywords-Support vector machine (SVM), Feature Extraction, Color histogram, image classification

I. INTRODUCTION

Plant diseases are turn into dilemma where it can cause of significant reduction of the quality and quantity of the agriculture products. Our research focuses on the detection of plants diseases based on color, shape detection and matching histogram technique. We need two very significance characteristic that is mainly concern with the accuracy of detection and speed to recognize the image diseases. Based on the color space, histogram, and shape detection techniques, we can able to find the disease of plant.

The content-based image retrieval (CBIR) systems have proven to be very useful in many fields to browse and search very huge image databases. Botanists are usually brought to use large collections of plants images. They need automatic tools to assist them in their work. This paper presents a plant/fruit/flower classification system which takes as input the image of a plant/fruit/flower and returns the most similar images from a database. The system is intended to be used as an e-commerce service where users can send the capture images of their house plants leaf/fruit/flower (which they often do not know by name and their diseases) to database to find their Latin names, care instructions and medicine. The problem involves identification of the matching plant, as well as retrieval of related varieties which may be also of interest to the user. A generalized diagram of a fruit image classification system is shown in Fig.1. The component which is of great importance is the feature extraction using support vector machine (SVM) algorithm. Feature extraction algorithm produces a feature vector, in which the components are numerical characterizations of the underlying biometrics.

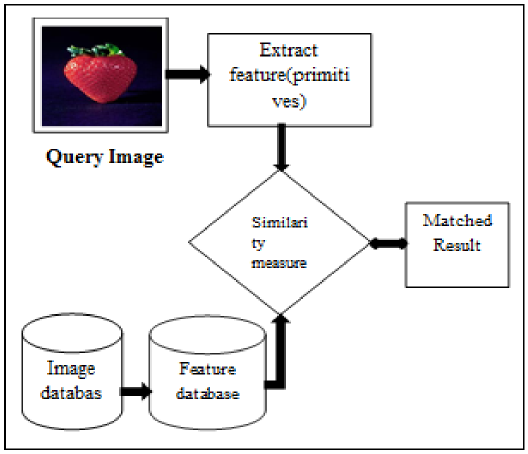

Fig.1 image classification system

The main focus of this paper work is devoted to finding suitable representation for images and classification generally requires comparison of images depending on the certain useful features.

II. PROPOSED SYSTEM

Support Vector Machines (SVMs)

SVM was first proposed by Vapnik and is gaining popularity in field of machine learning due to many attractive features and to show practical performance. It gives higher better performance in classification of image than other data classification algorithm. It is mainly used in real world problem like voice recognition, tone recognition, text categories, image classification, object detection, handwritten digital recognition, and data classification. Image classification is the process of grouping of similar types of image into a single unit i.e. called cluster of image. So the classification is a very exciting task to find exact result. To improve the result of classification, extract the related feature of image, because of this we also get good accuracy.

Support Vector Machines (SVMs) are supervised learning methods [2] used for image classification. It views the given image database as two sets of vectors in an ‘ n ' dimensional space and constructs a separating hyper plane that maximizes the margin between the images relevant to query and the images not relevant to the query. SVM is a kernel method and the kernel function used in SVM is very crucial in determining the performance.

There are some points in the feature space which are separated by some distance is called support vectors. It is the point between origin and that point and define the position of the separator. The distance from the decision surface tothe clos et data point determines the margin the classifier.

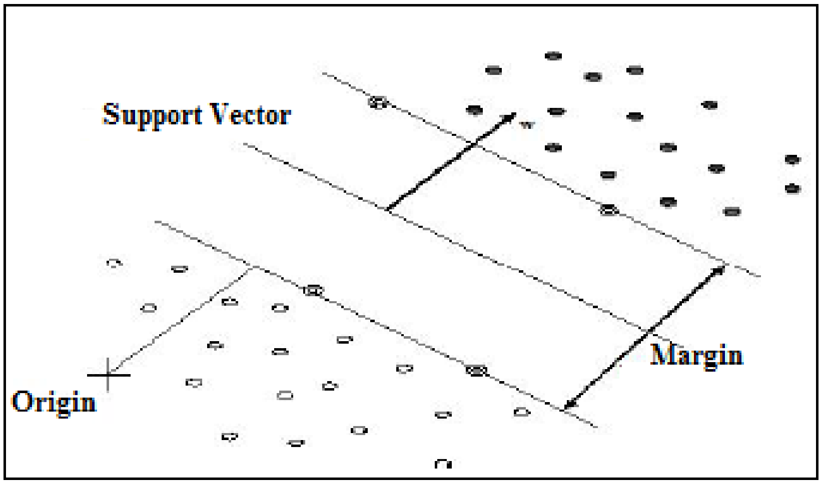

Fig.2Linear separating hyperplanes for two class separation

The main feature of SVM is to construct a hyper planes or a set of hyper planes with the help of support vectors in a higher dimension space [6]. These are mainly used for classification. It separate the space into two half space. A ‘good separation’ is achieved by hyper a plane that has the largest distance to the nearest data points. Here good separation means larger the separation between two hyper planes gives lower generalization error. That’s by it is called maximum margin classifier [5]. Two parallel hyper planes are constructed on each side of the hyper planes that separate the data. If geometric gap between hyper planes high than classification error is low [2], [5].

Image classification or categorization is a machine learning approach and can be treated as a step for speeding-up image retrieval in large databases and to improve retrieval accuracy. Similarly, in the absence of labeled data, unsupervised clustering is also found useful for increasing the retrieval speed as well as to improve retrieval accuracy. Image clustering inherently depends on a similarity measure, while image classification has been performed by different methods that neither require nor make use of similarity measures [2][5].

Feature extraction

It is very important step for image classification. In this, all the relevance or irrelevance features of image are extracted and on the basis of this classification of image performed. Basically feature extraction is a process of mapping image from image space to feature space. Feature space is a type of input space where similarity measures with the help of kernel function. In image, basically there are many features like colour, shape, text, size and dimension etc. which are mainly used for feature extraction but extracting those feature which are more relevance to our work in difficult task. The output given by this step is in the form of vector [5] [6].

General features are those which can be used for searching like colour, shape, texture and feature which are used for particular domain and have knowledge about them [5]. For example, we are searching for strawberries which belong to fruit category. These features are domain specific [5]. Some features are semantic which are very difficult to extract. Semantic features are those which have same meaningful information about image.

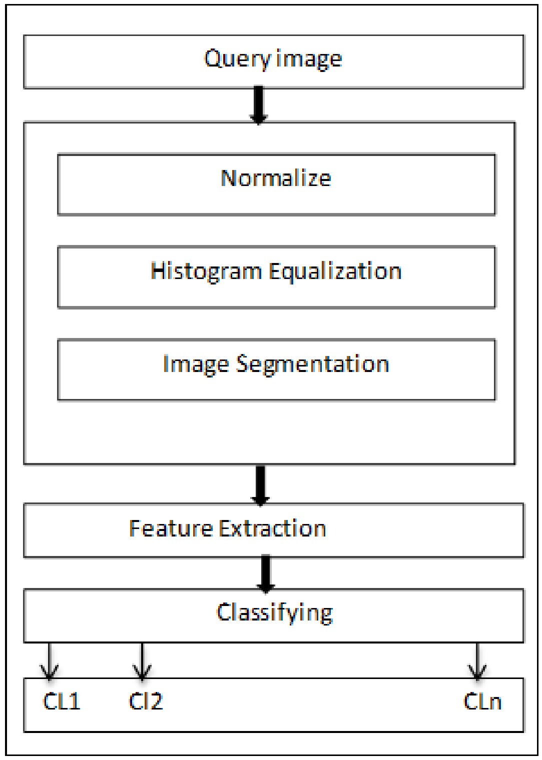

Fig.3Image Classification & Feature Extraction Process

In this category mean value, RGB value, Histogram value, Standard deviation and entropy are belong. These features are not easy to find [5]. So to analyze the set of feature with the help of input data is called feature extraction.

Image Classification

Image classification analyzes the numerical properties of various image features and organizes data into categories. Classification algorithms typically employ two phases of processing: *training* and *testing*. In the initial training phase, characteristic properties of typical image features are isolated and, based on these, a unique description of each classification category, *i.e. training class*, is created. In the subsequent testing phase, these feature-space partitions are used to classify image features.

The description of training classes is an extremely important component of the classification process. In supervised classification, *statistical* processes (*i.e.* based on an *a priori* knowledge of probability distribution functions) or *distribution-free* processes can be used to extract class descriptors. Unsupervised classification relies on *clustering* algorithms to automatically segment the training data into prototype classes. In either case, the motivating criteria for constructing training classes are that they are:

- *independent*, *i.e.* a change in the description of one training class should not change the value of another,
- *discriminatory*, *i.e.* different image features should have significantly different descriptions, and
- *reliable*, all image features within a training group should share the common definitive descriptions of that group.

Color Histograms

In image processing, a color histogram is a representation of the distribution of colors in an image. For digital images, it is basically the number of pixels that have colors in each of a fixed list of color ranges that span the image color space, the set of all possible colors. Color histogram technique is a very simple and low level method and in practice it has shown good results [2][3]especially for image indexing and retrieval tasks, where similar (not necessary identical) images are to be retrieved and easy feature extraction. This also ensures full translation and rotation invariance in the color images under classification task. The color histogram can be built for any kind of color space, although the term is more often used for three-dimensional spaces like RGB or HSV. For monochromatic images, the term intensity histogram is used.

Average RGB

Colors are commonly defined in three-dimensional color spaces. The color space models [7] can be differentiated as hardware-oriented and user-oriented. The hardware-oriented color spaces, including RGB and CMY are based on the three-color stimuli theory. The user-oriented color spaces, including HLS, HCV, HSV and HSB are based on the three human percepts of colors, i.e., hue, saturation, and brightness [7]. The RGB colorspace (see Figure 4) is defined as a unit cube with red, green, and blue axes; hence a color in an RGB color space is represented by a vector with three coordinates. When all three values are set to 0, the corresponding color is black. When all three values are set to 1, the corresponding color is white [1],[4]. The color histograms are defined as a set of bins where each bin denotes the probability of pixels in the image being of a particular color. A color histogram H for a given image is defined as a vector:

H={ H[*0*],H[*1*], ………H[*i*]…………H[*N*]}

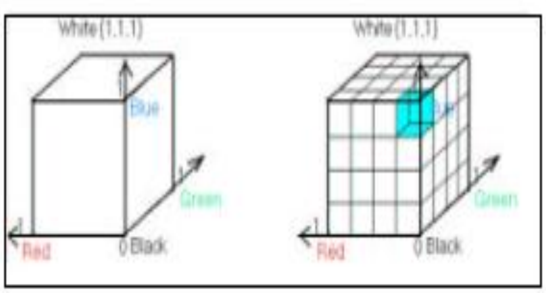

Fig.4 RGB color space

Where i represents a color in the color histogram and corresponds to a sub-cube in the RGB color space, $H[i]$ is the number of pixels in colors i in that image, and N is the number of bins in the color histogram, i.e., the number of colors in the adopted color model.

SVM ALGORITHM:-

SVM algorithm is introduce in 2002 by dust in boweil it introduce pattern reorganization svm consist of Two Types

- Linear SVM
- Non Linear

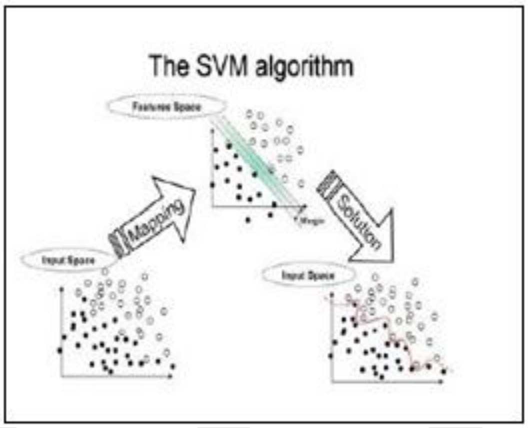

Fig.5 SVM algorithm

Step 1: Input Spaces

Initially take The Input Files Spaces and Mapping done to the using mathematical Formulae.

Step 2: Detection OF Contents

To detect the image related Content basic feature.

Step3: Margin

Margin Techniques used for Training the Dataset and testing of the dataset which include the maximum slab it consist of kernel function, polynomial mapping other kernel, RBF(Radial Basic Function), Signomoid Kernel(Specific Application)

Step 4: Solution

Give the proper solution for the images Contents Non linear input Solution data problem Solves that Ways.

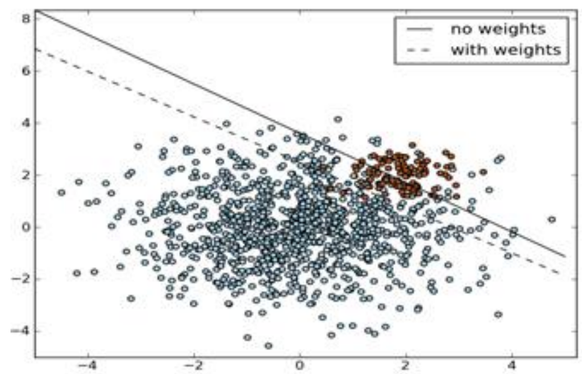

Fig.6 Non linear input Solution

III. CONCLUSION

The main focus of proposed work was to show the comparative study of neural network and support vector machine. Here we use SVM as a classifier for the classification of image and apply this classification process to all the features of image which are extract from feature extraction step. It is mainly used to find maximum margin hyper planes in a high dimensional feature space. Here we also explain all the conditions such as optimal separating hyper planes, linearly on-separable case and nonlinear support vector machines in which SVM work. A kernel based learning method used for the mapping purpose. So with the help of support vector machine we get much better performance than the other traditional method and get optimal result.

The histogram matching is based on the color feature and the edge detection technique. The color features extraction are applied on samples. The training process includes the training of these samples by using layers separation technique which separate the layers of RGB image into red, green, and blue layers and edge detection technique which detecting edges of the layered images. Once the histograms are generated then the testing image, immediately we applied the comparison technique based on the histogram.

The future work mainly concerns with the large database and advance feature of color extraction that contains a better result of detection. Another work concerns with research work in a particular field with advance features and technology.

ACKNOWLEDGEMENTS

We are very thankful to Prof. J. V. Sinde and all staff of Late G.N.S.COE, Nashik.

REFERENCES

[1] SudeepMarwaha, Subhash Chand, ArijitSaha, "Disease Diagnosis in crops using content based image retrieval" Division of Computer Application Indian Agricultural Statistics Research Institute New Delhi, India 2012.

[2] SaurabhAgrawal, Nishchal K Verma, PrateekTamarakar, PradipSircar, “Content Based Color Image Classification using SVM” Department of Electrical Engineering, Indian Institute of Technology Kanpur India 2011.

[3] Ahmed Naser Hussein, SyamsiahMashohor, M. IqbalSaripan, “A Texture-Based Approach for Content Based Image Retrieval System for Plant Leaves Images” Department of Computer and Communication Systems Engineering, Faculty of Engineering, University Putra Malaysia Serdang, Selangor, Malaysia 2011.

[4] SudhakarPutheti, Srinivasa Reddy Edara , “CBIR using Texels of RGB colourTextons” Computer Science and Engineering AcharyaNagarjuna University Guntur, A.P., India 2011.

[5] Sandeep Kumar, Zeeshan Khan, Anuragjain, “A Review of Content Based Image Classification using MachineLearning Approach”, International Journal of Advanced Computer Research, September-2012.

[6] Le Hoang Thai, Tran Son Hai and Nguyen ThanhThuy, “Image Classification using Support Vector Machine and Artificial Neural Network”, I.J. Information Technology and Computer Science, 2012.

[7] Avinash N Bhute, B. B. Meshram, “Integration of Color Feature Extraction Techniques for Intelligent Video Search Systems”, Dept. of Computer TechnologyVeermataJijabai Technological InstituteMatunga, Mumbai, India 2010