International Journal of Academic Research
ISSN: 2348-7666 Vol.2, Issue-1(4), January-March, 2015

An Improved Classification approach for Large Images

Marrapu Ramakrishna, M.Tech (Final Year), Department CSE, Baba Institute of Technology and Sciences (BITS), PM Palem. Madhurawada, Visakhapatnam, A.P

T Anand Kumar, Assistant Professor, Department CSE, Baba Institute of Technology and Sciences (BITS), PM Palem.Madhurawada, Visakhapatnam, A.P, India

Abstract: *Classification of image data is still an important research issue in the field of 1 image processing , Analyzes the unlabeled or sample data with the existing labeled or training samples ,Even this approach is traditional, but mechanism of analyzing the large scale images are very important. In this paper we are using an enhanced ID3 classifier by constructing the decision tree with enhanced features like mismatched Meta datasets and semantic comparison for classification of feature set of Images.*

Key words: Automatic, image, classification

I. Introduction

The term image classification refers to the labeling of images into one of a number of predefined classes. Although it is seemed not a very difficult task for humans, it has proved to be a difficult problem for machines. Therefore, image classification is also used for image-based CAPTCHA. To address image-based CAPTCHA problem and other problems like web searching, surveillance and sensor system, Designing and implementing automatic image classification algorithms has been an important research field for decades[1][3].

Many remote sensing systems record brightness values at different wavelengths that commonly include not only portions of the visible light spectrum, but also photo infrared and, in some cases, middle infrared bands.

The brightness values for each of these bands are typically stored in a separate grayscale image (raster). Each ground-resolution cell in an image therefore has a set of brightness values which in effect represent the "color" of that patch of the ground surface, if we extend our concept of color to include bands beyond the visible light range [2].

The Automatic Classification process uses the "colors", or spectral patterns, of raster cells in a multispectral image to automatically categorize all cells into a specified number of spectral classes. The relationship between spectral classes and different surface materials or land cover types may be known beforehand, or determined after classification by analysis of the spectral properties of each class[4]. The Automatic Classification process offers a variety of classification methods as well as tools to aid in the analysis of the classification results. The spectral pattern of a cell in a multispectral image can be quantified by plotting the raster value from each band on a separate coordinate axis to locate appoints in a hypothetical "spectral space". This spectral space has one dimension for each band in the image. Most classification methods use some measure of the distance between points in this spectral space to assess the similarity of spectral patterns. Cells that are close together in spectral space have similar spectral properties and have a high likelihood of imaging the same surface features [5].

Many remote sensing systems record brightness values at different wavelengths that commonly include not only portions of the visible light spectrum, but also photo infrared and, in some cases, middle infrared bands. The brightness values for each of these bands are typically stored in a separate grayscale image (raster). Each ground-resolution cell in an image therefore has a set of brightness values which in effect represent the “color” of that patch of the round surface, if we extend our concept of color to include bands beyond the visible light range [6].

The Automatic Classification process uses the “colors”, or spectral patterns, of raster cells in a multispectral image to automatically categorize all cells into a specified number of spectral classes. The relationship between spectral classes and different surface materials or land cover types may be known beforehand, or determined after classification by analysis of the spectral properties of each class. The Automatic Classification process offers a variety of classification methods as well as tools to aid in the analysis of the classification results.

The spectral pattern of a cell in a multispectral image can be quantified by plotting the raster value from each band on a separate coordinate axis to locate appoints in a hypothetical “spectral space”. This spectral space has one dimension for each band in the image. Most classification methods use some measure of the distance between points in this spectral space to assess the similarity of spectral patterns. Cells that are close together in spectral space have similar spectral properties and have a high likelihood of imaging the same surface features [7].

Our goal in this work is to develop a classification methodology for the automatic annotation of outdoor scene images. The training data is a set of images, each labeled with a list of one or more object (or concept) classes that it contains. There is no information on the locations of these entities in the image. For each class to be learned, a classifier is trained to detect instances of that class, regardless of size, orientation, or location in the image [8][9]. The solution that we propose is a generative/discriminative learning procedure that learns the object or concept classes that appear in an image from multiple segmentations of pre-annotated training images. It is significant in several respects:

1. It is able to work with any type of feature that can be extracted from an image by some automatic segmentation process and represented by a vector of attribute values. It can work with regions from a color or texture segmentation, groups of line segments, or small windows selected by an interest operator.

2. It can work with any number of different feature types simultaneously. As we will show, the formalism we developed for a single feature type generalizes easily to multiple feature types. Thus we can use several features types together for a more powerful recognition system.

3. Like the work of Dork and Schmidt and the more theoretical paper of Raina teal, our method consists of two phases: a generative phase followed by a discriminative phase. Our method is distinguished in the elegant framework we use for our discriminative phase. In

particular, although each segmentation process can produce a variable number of instances of its features, our methodology produces a fixed-length description of each image that summarizes the feature information in a novel way. This allows the discriminative phase to be implemented by standard classifiers such as neural nets or (linear kernel) support vector machines.

II. Related Work

Various classification approaches available in the traditional approaches of Image classification, they are may not be optimal with remote sensing images, these approaches like multi level classifier, distance classifier and likelihood classifier has various drawbacks like mismatched meta datasets in training and testing dataset in the process of classification[8].

K – Means Algorithm-Means algorithm is a well-known clustering algorithm popularly known as Hard C Means algorithm. This algorithm splits the given image into different clusters of pixels in the feature space, each of them defined by its center. Initially each pixel in the image is allocated to the nearest cluster. Then the new centers are computed with the new clusters. These steps are repeated until convergence. Basically we need to determine the number of clusters K first.

Then the centroid will be assumed for these clusters. We could assume random objects as the initial centroids or the first K objects in sequence could also serve as the initial centroids. The K means algorithm in a logical representation: Execute the below steps until convergence. Do the following while no object move group.

a) Determine the centroid coordinate (Random assignment).
b) Determine the distance of each Object pixel to the Centroids.
c) Group the object based on minimum distance with the Centroid.

The basic principle of the proposed algorithm is integrating the K-Means algorithm with LoG filter and Prewitt filter as follows.

3.1 *Algorithm: 1*

Step 1: Read the RGB Image available for classification. *Step 2:* Convert available image from RGB color space to L*a*b* color Space

Step 3: Classify the colors in 'a*b*' space using K- means clustering algorithm

Step 4: Label every pixel of the image using the results from K-means algorithm
Step 5: Create images that segment the image by color. *Step 6:* Segment the nuclei of the image into a separate image
Step 7: The Laplacian of Gaussian filter finds edges by looking for zero crossings.

III. PROPOSED WORK

We are proposing an empirical model of classification approach followed by clustering of feature set of large scale images, for clustering purpose we are using the incremental clustering for feature set of datasets and enhances the classification of ID3 classifier with mismatches attribute meta dataset comparison , semantics comparison while classifying the testing datasets with training datasets.

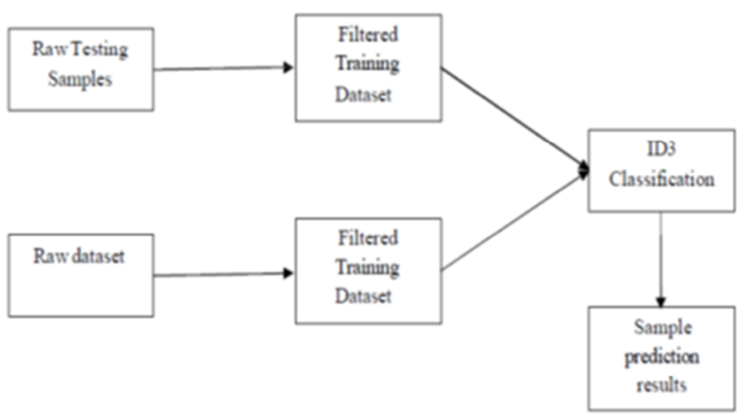


Fig 1. Image Data classification architecture

We gather the information of training datasets, which contains the previous synthetic data of images, properties of the image may vary, because it depends upon the objective. It is used for classify the testing sample with our proposed approach of decision tree construction.

Before classification of the testing samples with training datasets, we preprocess both testing and training samples for optimal performance instead of fail the learning approach and compare the semantically equal attribute sets.

We classify the testing sample of the image with training datasets, by computing the attribute information gain in terms of entropy, consider the maximum information gain attribute first and construct the decision tree to analyze the testing sample of data.

1) Establish Classification Attribute

2) Compute Classification Entropy.

3) For every attribute in R set, compute Information Gain using classification attribute.

4) Choose Attribute with the highest information gain to be the next Node in the tree (starting from the main root node).

the class name where as a non-leaf node is a decision node and the decision node is an attribute test with each branch (to another decision tree) being a possible value of the attribute and ID3 uses information gain to help it decide which attribute goes into a decision node and the advantage of learning a decision tree is that a program rather than a knowledge engineer that elicits knowledge from a final expert.

Gain measures how well a given attribute separates training examples into targeted classes. The only one with the highest information (information being the most useful for classification) is selected to define gain, we first borrow an idea from information theory called entropy and Entropy measures the amount of information in an attribute.

This is the formula for calculating homogeneity of a sample.

□◌1◎ $I2$

It helps to measure the information gain with respect to the attributes

5) Eliminate or remove Node Attribute, creating reduced table RS set.

6) Repeat steps 3 to 5 until all attributes have been used or

the same classification value remains for all rows in the reduced table.

ID3 builds a decision tree from a fixed set of examples and the resulting tree is used to classify future samples and the example has several attributes and belongs to a class (like yes or no) and the leaf nodes of the decision tree contain

In this module, we are enhancing the drawbacks in the previous traditional approach, those are attribute mismatching and semantic comparison of the datasets, it obvious improves the performance of the classification when training dataset column unknown or

$Gain(A)$ □□ $E(Current\ set)$ □□□

Miscellaneous Features:

$E(all\ child\ sets)$

when they are semantically equal.

Experimental Analysis:

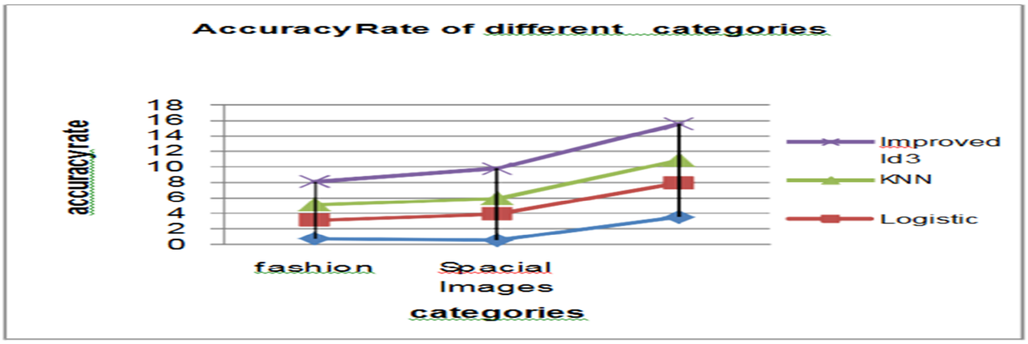
For experimental implementation we considered some sample datasets with the following features as angle ,ASM (angular second moment feature),Contrast and correlation and class labels as Grass land and water land, for the following synthetic dataset we construct a decision tree with respect to information gain in terms of entropy

Angle (in degrees)	ASM	Contrast	Correlation	Class
0	0.128	3.048	.8075	Grassland
0	.1016	2.153	.7254	Water land
45	.0080	4.011	0.6366	Grassland
45	.0771	3.057	.4768	Water land
90	.0077	4.014	.5987	Grassland
90	.0762.	3.113	.4646	Water land
135	.0064	4.7709	.4610	Grassland
135	.0741	3.129	.4650	Water land
160	.0087	3.945	.6259	Grassland
160	.0822	2.863	.5327	Water land

Different classifiers performance comparison:

We run different classifiers on different categories and compare their performances. For all classifiers, we can find flowers classification accuracy rate is always the highest while fashion classification accuracy rate is always the lowest. This implies the vector features for flowers are more distinguishable than the other two.

In the aspect of classifier performance on the same dataset, for our database, the logistic regression algorithm performance is not as good as the other two which guarantee us of meeting our first success metric.



Conclusion

Finally, we are concluding our research work with preprocessed classification approach with ID3 classification, initially testing and training samples can be preprocessed and compares only with available attribute set, instead of terminating classification of image data.

References

1. A tutorial of image classification by TNTMipspageno 0-36.
2. Image Classification Using SVM, KNN and Performance Comparison with Logistic Regression by Qi Gu, Zhifei Song.
3. Generative/Discriminative Learning Algorithm for Image Classification. Li, L. G. Shapiro, and J. Bilmes
4. Image Classification through integrated K- Means Algorithm by Balasubramanian Subbiah and Seldev Christopher. C 2014.
5. Flayed Sabine and W.H.Freeman, “Remote sensing, Principles and Interpretation”.
6. S.N. Sivanandam&Sumathi& S. N. Deepa,“Introduction to Neural Networks using Mat lab 6.0”.
7. K. Mumtaz and Dr. K. Duraiswamy, “A Novel Density based improved k-means Clustering Algorithm –Dbkmeans” International Journal on Computer Science and Engineering ISSN : 0975-3397 213; Vol. 02, No.02, 2010, 213-218
8. MadjidKhalilian, Norwati Mustapha, MD NasirSuliman, MD Ali Mamat “A Novel K-Means Based Clustering Algorithm for High Dimensional Data Sets” International Multi Conference of Engineers and Computer Scientists 2010 Vol I, IMECS 2010, March17-19, Hong Kong Baldacci,
9. R., Hadjiconstantinou, E., and Mingozzi, A. An exact algorithm for the capacitated vehicle routing problem based on a two commodity network flow formulation. Operations Research, 52(5):723–738,2004.