International Journal of Engineering Applied Sciences and Technology, **2016**
Vol. 1, Issue 4, ISSN No. 2455-2143, Pages 49-51
Published Online January-February 2016 in IJEAST (http://www.ijeast.com)

SCENE RECOGNITION WITH BAG OF WORDS

Laveti Lakshmideepak
Undergraduate student,
Department of CSE,
VIIT, Visakhapatnam,
A.P, INDIA

D.Aswani
Asst. Professor,
Department of CSE,
ANITS, Visakhapatnam.
A.P, INDIA

Dr. Challa Narasimham
Professor,
Department of CSE,
VIIT, Visakhapatnam,
A.P, INDIA

Abstract— **Scene recognition is one of the key features for the development of Robotics for better understanding of the environment. The aim of this paper scene recognition is an intelligence associating with the machine. This paper will analyze the activity of scene recognition starting with very simple methods i.e. tiny images and nearest neighbor classification and then we examine techniques that resemble the state of the art, bags of quantized local features and classifying techniques like linear classifiers learned by Support Vector Machines (SVM).**

Keywords— ***Image Representation, Scene Recognition, Robotics, Intelligence, Nearest Neighbor Classifier, SVM***

I. INTRODUCTION

Scene recognition is one of the toughest challenges of the artificial intelligence. Computer Vision (CV) is the science of training a computer how to identify a physical entity in its surroundings. Its task is to capture an image, understand it, reconstruct it and generates a meaningful and accurate description. Its ultimate aim is to imitate and improve on human visual perception. Computer vision is subset of artificial intelligence. Computer vision started with solving a problem as 'Summer Vision Project' in earlier days. Actually this task is not successful; if it's so then it will lead to a different study. Our main objective is not limited to work with the intensity of the pixels it has to understand the inner meaning of the system. Larry Roberts's Ph.D. dissertation in 1963 at the MIT, proposed the plan of extracting 3D geometrical information from related 2D views of blocks. Computer Vision can help robots in Localization, obstacle avoidance, Mapping, Object recognition (people and objects) and learning interaction with object.

II. BACKGROUND

In this paper we follow different techniques and methods to achieve the scene recognition. There are two methods like tiny image representation and nearest neighbour classifier. There are two techniques to techniques that resemble the state of the art, bags of quantized local features and classifying techniques like linear classifiers learned by support vector machines. There are three main modules to solve this problem. They are tiny image representation, feature extraction and trained classifiers. We general take image as input of .jpg format and we will produce a recognized output through our approach. We use bags of words model which is a popular technique for image classification inspired by models used in natural language processing.

III. PROPOSED SYSTEM

A. INPUT IMAGE-

Input image is the collection of pixels. We are giving the image as input of format .jpeg (joint photographer's expert group).

B. TINY IMAGE REPRESENTATION-

We have images of various sizes. We cannot apply the whole process to different images. So we are reducing the size of image to 16*16 so that we can process the set of images with less data. We can identify the pixels easily.

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 4, ISSN No. 2455-2143, Pages 49-51
Published Online January-February 2016 in IJEAST (http://www.ijeast.com)

C. BLOCK DIAGRAM-

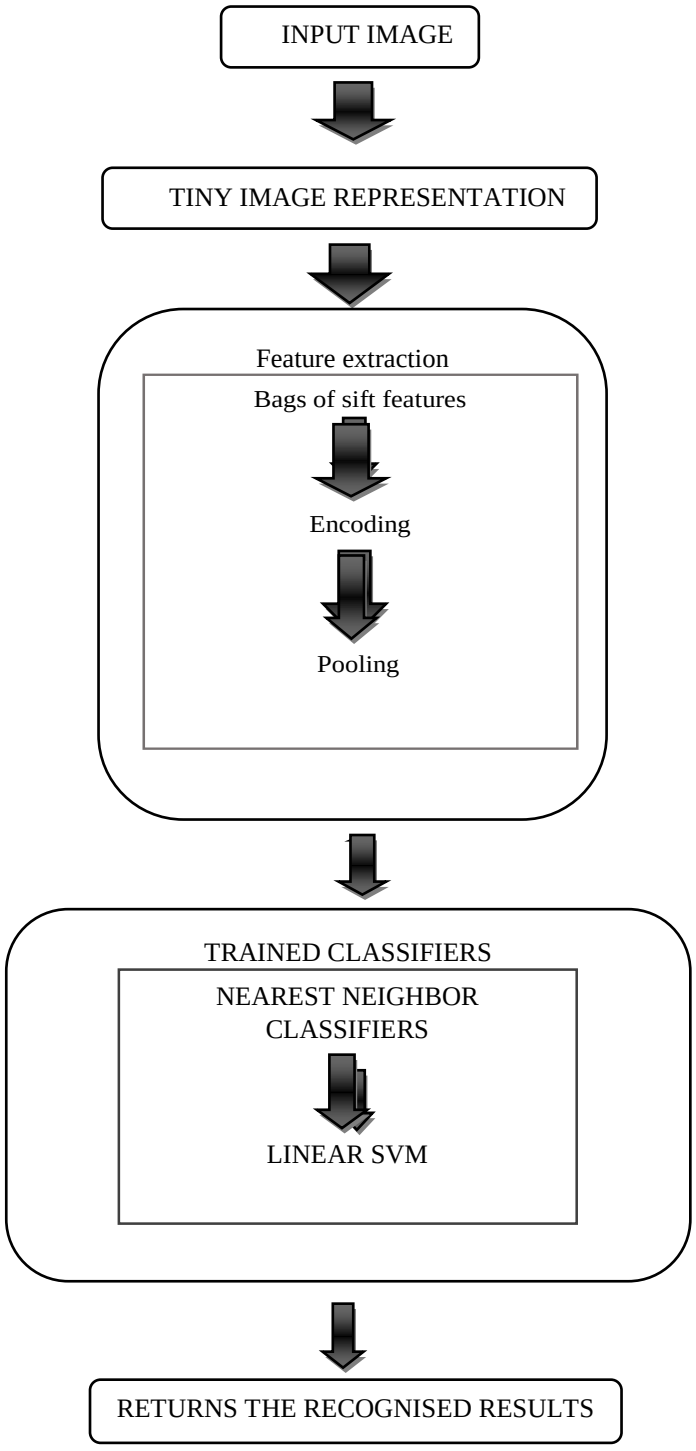


Fig. 1. Block Diagram

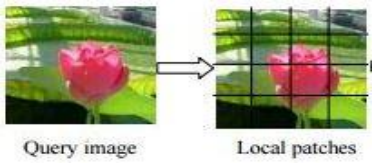


Fig 2: Tiny image representation.

D. FEATURE EXTRATION-

Feature extraction is one of the difficult processes to extract the correct features which can give best results for the given input image. Based on the features we extract we get that much accurate results. This module involves bags of sift features encoding and pooling.

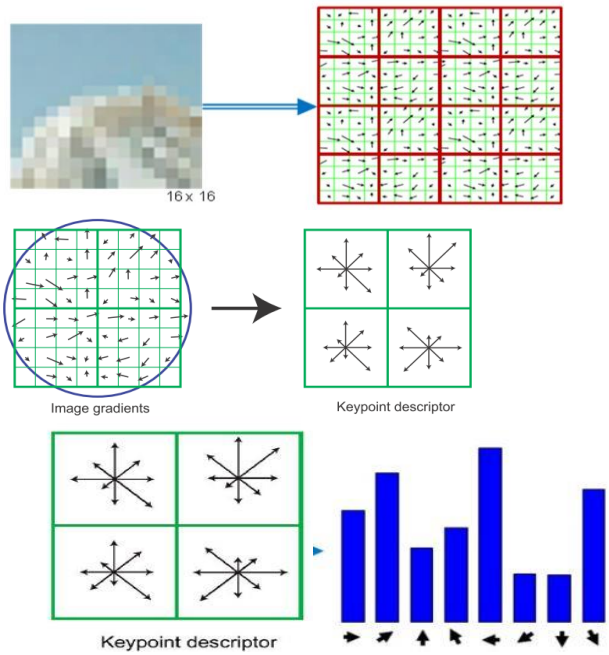


Fig 3: Feature Extraction.

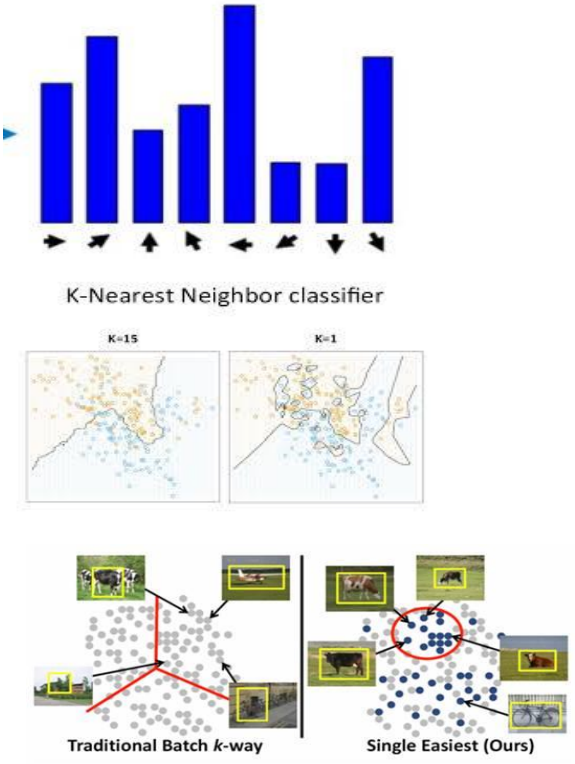
E. FEATURE EXTRATION-

Encoding is the quantization of the image patches that constitute the image or object to be classified. Basic encoding schemes work by first running K-means on the set of all (e.g. 3x3) pixel patches that you collect across multiple instances of your images/objects .This builds what is known a **dictionary** represented by the centroids obtained from the clustering. At the end of this process, you end up with K representative "visual words" (the centroid of each cluster after K means ends) of 3x3 patches. These "visual words" represent what is usually understood as your visual dictionary. Once you have these visual words, encoding is the process of assigning each 3x3 patch within your image/object the closest 3x3 word in the dictionary.

International Journal of Engineering Applied Sciences and Technology, **2016**
Vol. 1, Issue 4, ISSN No. 2455-2143, Pages 49-51
Published Online January-February 2016 in IJEAST (http://www.ijeast.com)

Pooling refers to the process of representing an image (or the object to be classified) as a "bag of words". The word bag here is meant to convey that once you have encoded each patch with a word (a number between 1 and K), you build a new representation (a bag) that discards the spatial relationship between the words that constitute your image or object. Pooling helps to achieve classification. It is a process of building a histogram of words (i.e. pooling ~ **"sampling" words** from the image to build a probability mass function of words).

F. TRAINED CLASSIFICATION-



The nearest neighbor classifier simply takes the description of a scene (Bags of SIFT) and calculates the "closest" Euclidean match to the training data. K-NN algorithm is a non-parametric method for classification.

G. LINEAR SVM(SUPPORT VECTOR MACHINES)-

Linear SVM is a one-vs.-all classification method that uses a learning algorithm to linearly fit the data as belonging to a category or not. The Linear SVM implemented is a simple one verses all implementation, which trains (in this case) 15 separate SVMs, one to recognize each type of scene. The results of the SVM trainers are then applied to each image, and the one which results in the highest response is considered to be the most likely scene match.

IV. CONCLUSION

Our output accuracy will be increased at each level of the module. At the end of the process a recognized output will be obtained. The results of this project will be published in my next paper. Scene recognition is done up to some level with this approach. Accuracy of the outputs has to improve for the better results of scene recognition. There is one of the good approaches for the scene recognition mechanism.

V. REFERENCE

[1] Seymour Papert,"The Summer Vision Project", *Artificial Intelligence Group, Vision Memo,July 7,1966.*

[2] Svetlana Lazebnik[1], Cordelia Schmid[2], Jean Ponce[1,3], "Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories," *Citied by 5298..*

[3] Jianxiong Xiao, James Hays, Krista A.Ehinger, Aude Oliva, Antonio Torralba, "Sun Database: Large scale recognition from abbey to zoo." .

[4] Bolei Zhou[1], Agata Lapedriza[1,3], Jianxiong Xiao[2], Antonio Torralba[1], and Aude Oliva[1], " A Learning Deep Features for Scene Recognition using Places Database", *Cited by 192.*

[5] Megha Pandey and Svetlana Lazebnik, "Scene Recognition and Weakly Supervised Object Localization with Deformable Part-Based Models", *In ICCV, 2011*..