Binary scene classification

B S N V Chaitanya (viswachaitanya.b16@iiits:in)
B.Tech,ECE,Indian Institute Of Information Technology Chittoor, Sri City, A.P., India

Abstract— In this paper,we present a procedure to classify real world scenes into two categories:Natural and Man-made.Given an image as an input we wish to classify as Natural or Manmade.Given a set of images of scenes which contain different objects,for example sea,mountains,buildings,we discover these objects in every image and use the same to perform the classification. The image objects are produced by an image segmentation algorithm using multiclass SVM(support vector machine)classifier integrated with histogram intersection kernel, which we use for our classification.SIFT(Scale Invariant Feature Transform) is a feature descriptor that describes a object in an image in terms of number of interest points.As SIFT feature descriptor is invariant to scaling and translation we use this more commonly.In this paper,we propose a method to design an efficient image classifier by combining the SIFT feature descriptor with intersection kernel SVM.Our experimental results show that the proposed method has good accuracy.

Keywords: SIFT,Support Vector Machine(SVM),Histogram intersection kernel Feature Descriptor

I. Introduction

Our project mainly focuses on classifying a given image into its respective category (Manmade or Natural).We essentially need to figure out features such that Image = f(features). The most important characteristic of feature that we require is a very strong co-relation between the feature value and the class of the image.This makes learning an easier, faster and a less error prone job. Thus, feature identification is among the most important task. Once, the features are identified a classifier can be constructed using state of art algorithms like SVM, k-means or any other suitable method. With help of this classifier, identification of an input image needs to be done. Steps involved in doing this are:

1.Feature Extraction
2.Quantization
3.Classification

STEP 1:Feature Extraction:

Features are extracted from scale invariant image regions.The SIFT algorithm is widely used for object recognition and detection which is invariant to illumination changes and affine or 3D projection.The SIFT key-descriptors that produce a compact feature descriptor describes an image with its key-points.So,the features are extracted using SIFT.

STEP 2:Quantization:Now,using k-means clustering descriptors are grouped.As we have 2 categories,using k-means algorithm 2 clusters are formed.From the clusters obtained from k-means clustering,each cluster is assigned with each visual word and frequency histograms are built.Then,the k-d tree is built on the cluster centers.The SIFT descriptors are vectors of

128 elements, i.e. points in 128-dimensional space. So you can try to cluster them, like any other points. You extract SIFT descriptors from a large number of images, similar to those you wish classify using bag-of-features. Then you run k-means clustering on this large set of SIFT descriptors to partition it into 200 clusters, i.e. to assign each descriptor to a cluster. k-means will give you 200 cluster centers, which you can use to assign any other SIFT descriptor to a particular cluster.

STEP 3:Classification: When an image is given as an input,it is classified into its respective category by SVM.Then you take each SIFT descriptor in your image, and decide which of the 200 clusters it belongs to, by finding the center of the cluster closest to it. Then you simply count how many features from each cluster you have. Thus, for any image with any number of SIFT features you have a histogram of 200 bins. That is your feature vector which you give to the SVM.

II. Related Work

A. Yangzihao Wang and Yuduo Wu, University of California, Davis proposed ”Scene classification based on neural networks and deep convolution”

They have proposed a novel approach for scene classification based on deep convolutional neural networks. They tried to fill in the semantic gap between the large deep convolutional neural network features from the massive dataset like ImageNet and the high-level context in the scene categories. Their method, which works by extracting spatial pyramid features from region proposals of images, has shown that deep convolutional neural network is capable of achieving promising results on highly challenging, large-scale dataset which contains both scenes that can be well characterized by global spatial properties and the scenes that can be well characterized by detailed objects they contains. It is notable and significant that we achieved these results by using a combination of classical computer vision approaches and deep convolutional neural networks

B. Chern Hong Lim and Chee Seng Chan proposed A Fuzzy Qualitative Approach for Scene Classification”

They show the implementation of FQS and the usage of it in natural scenes classification. The experiments show positive results in term of crisp classification and fuzzy classification results. However, there are more research to be done to fine tune the proposed framework.Their approach train a model directly from the training data in fuzzy qualitative quantity space and our results are defined in a ranking system.

C. Limin Wang, Sheng Guo, Weilin Huang,Yuanjun Xiong, and Yu Qiao proposed ”Scene Classification with Multi-Resolution CNNs”

They have studied the problem of scene recognition on large-scale datasets such as the Places, Places2, and LSUN. Large-scale scene recognition suffers from two major problems: visual inconsistence (large intra-class variation) and label ambiguity (small inter-class variation). They developed powerful multi-resolution knowledge guided disambiguation framework that effectively tackle these two crucial issues. We introduced multi-resolution CNNs which are able to capture visual information from different scales. Furthermore, they proposed two knowledge guided disambiguation approaches to exploit extra knowledge, which guide CNNs training toward a better optimization, with improved generalization ability

III. DATASET

We used the 15-class scene category dataset for testing and training. From this, we made a new dataset containing two categories (manmade and natural). Each category contains 200 images for training. Each image size is 300X250 on an average. Manmade set contains scenes like: Kitchen, Tall Building, Bedroom, Office, Suburban. Natural set contains scenes like: Forests, Countryside, Mountains, Coast. Input is an image and it is classified as natural or manmade. Accuracy is calculated using confusion matrix.

(i) Man-made scenes

(ii)Natural scenes

IV. VARIOUS SCENE CLASSIFICATION TECHNIQUES

A. GIST:Image Feature Descriptor

Given an input image, a GIST descriptor is computed by:
1.Convolve the image with 32 Gabor filters at 4 scales, 8 orientations, producing 32 feature maps of the same size of the input image.
2.Divide each feature map into 16 regions (by a 4x4 grid), and then average the feature values within each region.
3.Concatenate the 16 averaged values of all 32 feature maps, resulting in a 16x32=512 GIST descriptor.
Intuitively, GIST summarizes the gradient information (scales and orientations) for different parts of an image, which provides a rough description (the gist) of the scene.

B. SIFT:Image Feature Descriptor

For feature matching in object detection,SIFT is used.SIFT is a localized image patch descriptor. A typical image has a few thousand SIFT descriptors, each of 128 dimensions. SIFT was designed for scale and affine invariance in wide baseline image matching tasks, which were part of stereo vision. It was subsequently utilized for image classification and performed well.SIFT descriptors are scale invariant.SIFT can robustly identify objects even among clutter and under partial occlusion, because the SIFT feature descriptor is invariant to uniform scaling, orientation, illumination changes, and partially invariant to affine distortion.

C. K-NEAREST NEIGHBOUR ALGORITHM:Classifier

The k-NN algorithm is among the simplest of all machine learning algorithms. In pattern recognition, the k-nearest neighbors algorithm (k-NN) is a non-parametric method used for classification and regression.In both cases, the input con- sists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression.In k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbour.

D. Support Vector Machine SVM:Image Classifier

Support Vector Machines are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (super-vised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. In two dimentional space this hyperplane is a line dividing a plane in two parts where in each class lay in either side.

V. IMPLEMENTATION

A. Resizing the image:

First, the input image is converted into grayscale image and resized to 256X256.

B. Finding the pyramid of SIFT descriptors:

Using the dense sift impementation in VLFeat:vl_dsift we find sift descriptors of the image.vl_sift bundles a feature detector and a feature descriptor.A number of frames(attributed regions) from an image are extracted by the detector in a way which is consistent with some variations of the illumination,viewpoint. The descriptor associates to the regions a signature which identifies their appearance compactly and robustly.A dense set of SIFT features are extracted from an image using vl_dsift.We create the pyramid of sift descriptors by splitting the image into multiple sublevels.Then vl_dsift is made to run on the smaller portion of the images in smaller steps and appended to the images feature vector.

C. Creating Vocabulary:

We have to generate a vocabulary using the descriptor we have,to classify the given image.We chose a random set of images in our data set and constructed a vocabulary using the pyramid of sifts descriptor mentioned above.And then these descriptors were grouped together into 200 clusters using k-means clustering with VLFeat: vl kmeans(clustering algorithm). Its purpose is to partition a set of vectors into k groups that cluster around common mean vector. It is a lengthy process,as large number of clusters are clustered everytime.So,the vocabulary that has been generated for the first time when training is saved as a .mat file.

D.Creating Histogram:

Once the vocabulary was defined, we define an im-age's pyramid of sifts with a histogram using VLFeat: vl_kdtreequery function.This is a branch-and-bound technique that maintains an estimate of the smallest distance from the query point to any of the data points down all of the open paths. vl kdtreequery function supports two important_ operations: approximate nearest-neighbor search and k-nearest neighbor search.We first run the pyramid of sift descriptors function defined initially on the image. This gives us a large vector of features defining an images sifts at multiple levels steps. We then run vl kdtreequeryfunction to compare these features to our vocabularies features to give us the distance between the features we found and its closest vocabulary fea-ture. Finally, we make a histogram of these closest vocabulary features. These histograms are used to compare the training and the test images. So as to reduce the running time of the code, the histograms that are created are also saved as a .mat file.

E. Support Vector Machine:

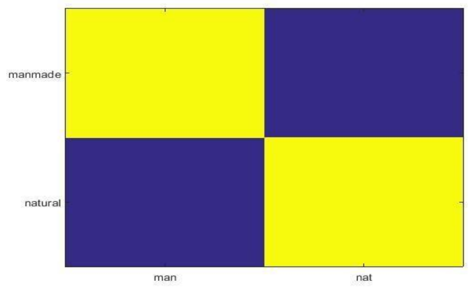
Now,as the features are defined for the images,we can clas-sify the test images by comparing with the training set.We per-form a one vs many strategies against our test images,for each category that we are trying to classify. We call vl_svmtrain against training images' features to generate weight and offset vectors. We then multiply our weights against our test image features and add the offsets to get our one vs many result. We place this result into the confusion matrix as we go through each category and take the max to figure out which category the image is closest to our test image.

F. Testing:

1.First the same set of images are used for training and testing and the output is observed.
2.Next,a new set of testing images are used and output is observed.
3.At last,a single image is given as input and classified as manmade or natural.

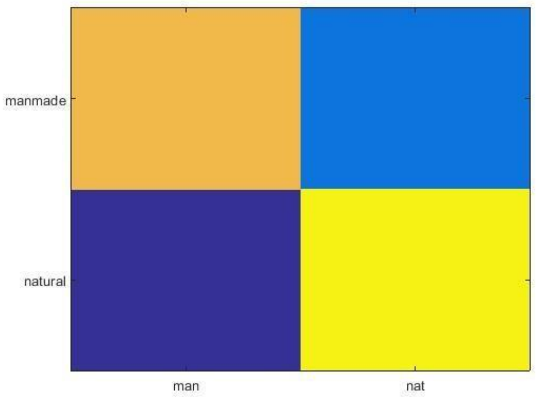
VI. Results:

1.When the testing is done with same set of images that are used for training using SIFT and GIST the accuracy we got was 99% and 100% respectively.

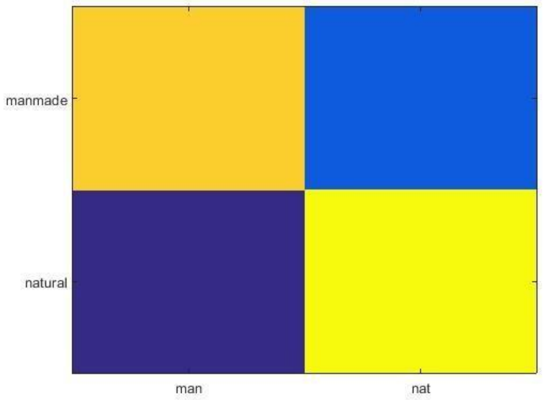


CONFUSION MATRIX USING SIFT AND GIST FOR TRAINING IMAGES

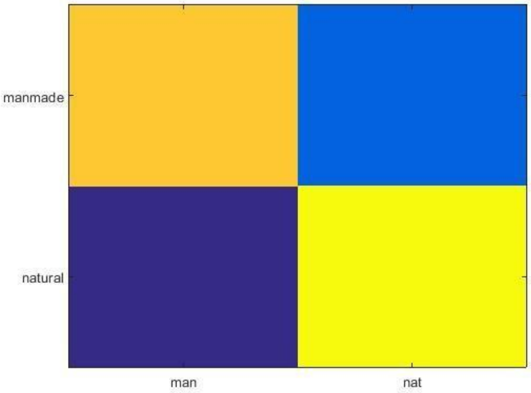
2.When a new set of images is used for testing,using SIFT we got an accuracy of 90%,using GIST we got 93% and using GIST+SIFT,we got approximately 95.6%.



CONFUSION MATRIX USING SIFT FOR TESTING IMAGES



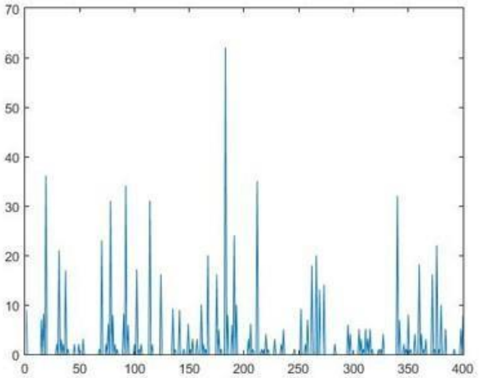
CONFUSION MATRIX USING GIST FOR TESTING IMAGES



CONFUSION MATRIX USING SIFT+GIST FOR TESTING IMAGES

3.When a single image is taken as an input the classification is done at three levels of the tree.

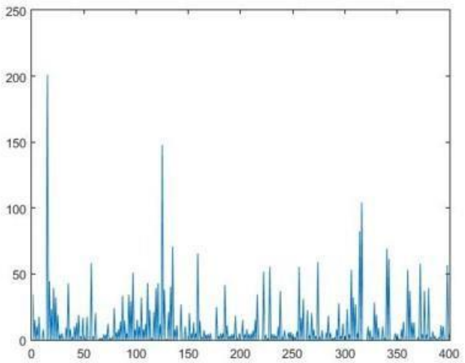
(i)At level L = 1 the image is taken as a whole.The accuracy we got was 80%.



(i)HISTOGRAM AT LEVEL-1

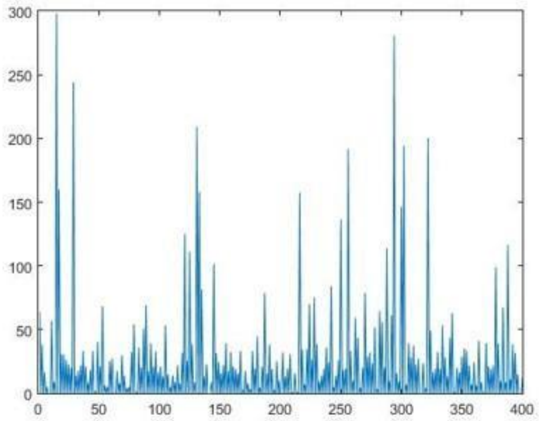
At level L = 2 the image was divided into smaller parts.The accuracy we got was 96%.

For strong features, performance drops from L = 2 to L = 3 because the highest level of L = 3 is too finely subdivided. As it is finely divided individual frames yield too few matches,that is it is difficult to match the features.So at level L = 2 we get strong feature so the accuracy is more.



(ii)HISTOGRAM AT LEVEL-2

At level L = 3 the image was further divided into smaller parts and the accuracy we got was 94.1%.



(iii)HISTOGRAM AT LEVEL-3

VII. SUMMARY AND CONCLUSION

Tasks such as scene classification pose a tough problem especially due to large variations in everyday images.It is rather tough to capture this tremendous amount of variation. Selection of appropriate features pose a tough challenge.Our proposed model has been designed by using SIFT with the Support vector machine(SVM) for the binary classification in adaptive manner.The problistic classification with the binary class support vector machine has been utilised for the robustness in the classification.

FUTURE SCOPE:

The scene classification can be extended for multi-class classification.This can be implemented at a high level like using neural networks and feature extractions using SIFT and GIST together.

References

[1] Bosch, Anna, Andrew Zisserman, and Xavier Muoz. ”Scene classifica-tion via pLSA.”, Computer VisionECCV 2006 (2006): 517-530..

[2] Zhou, Bolei, et al. ”Learning deep features for scene recognition using places database.”, Advances in neural information processing systems. 2014.

[3] Dutt, BVV Sri Raj, Pulkit Agrawal, and Sushoban Nayak. ”Scene Classification in images.”

[4] Gokalp, Demir, and Selim Aksoy. ”Scene classification using bag-of-regions representations.” ,Computer Vision and Pattern Recognition, 2007. CVPR’07. IEEE Conference on. IEEE, 2007.

[5] Hare, Jonathon S., Sina Samangooei, and Paul H. Lewis. ”Efficient clustering and quantisation of SIFT features: exploiting characteristics of the SIFT descriptor and interest region detectors under image inversion.” , Proceedings of the 1st ACM International Conference on Multimedia Retrieval. ACM, 2011.

[6] Hassaballah, M., Aly Amin Abdelmgeid, and Hammam A. Alshazly. ”Image Features Detection, Description and Matching.” , Image Feature Detectors and Descriptors. Springer International Publishing, 2016. 11-45.

[7] Hu, Junlin, and Ping Guo. ”Combined Descriptors in Spatial Pyramid Domain for Image Classification.”, arXiv preprint arXiv:1210.0386 (2012).

[8] Xiao, Jianxiong, et al. ”Sun database: Large-scale scene recognition from abbey to zoo.” , Computer vision and pattern recognition (CVPR), 2010 IEEE conference on. IEEE, 2010.

[9] Tomaev, Nenad, and Dunja Mladeni. ”MODIFIED K-MEANS ALGO-
RITHM FOR FINDING SIFT CLUSTERS IN AN IMAGE.”

[10] Li, Li-Jia, et al. ”Objects as Attributes for Scene Classification.” , ECCV Workshops (1). 2010.