Implementation and Analyzing SURF Feature Detection & Extraction on WANG Images Using Custom Bag of Features Model

Roohi Ali and Manish Maheshwari

September 13, 2021

Implementation and Analyzing SURF Feature Detection & Extraction on WANG Images Using Custom Bag of Features Model

Roohi Ali*
Department of Computer Science and Applications,
Makhanlal Chaturvedi University of Journalism and Communication,
Bhopal (462001). M. P, India.
roohi.ali2006@gmail.com

Manish Maheshwari
Department of Computer Science and Applications,
Makhanlal Chaturvedi University of Journalism and Communication,
Bhopal (462001). M. P, India.
manishbhom@yahoo.com

Abstract: A novel technique of image classification using BOVW model also known as Bag of Visual Words is very popular for retrieval of images using features instead of text vocabulary. The entire process first involves feature detection of images by selecting key points or forming a Grid over images, the choice made in order to speed up the process of detection. Then comes the stage of feature extraction for which SURF, a binary feature descriptor is employed. K-means clustering is then applied in order to quantize and make the bag of visual words. Every image, expressed as a histogram of visual words is fed to a supervised learning model, SVM for training. SVM is then tested for classification of images into respective classes. Matlab is used for implementation using bag class with Extractor fuction over 1000 image dataset WANG with 10 different categories.

Keywords: BOVW, K-means clustering, SURF, Extractor Function, SVM.

1.1. Introduction

"An image is a set of signals sensed by the human eye and processed by the visual cortex in the brain creating a vivid experience of a scene that is instantly associated with concepts and objects previously perceived and recorded in one's memory."

One of the prominent applications of image processing is Image feature extraction that laid the base for many image processing algorithms prominently image classification. Images are classified between the predefined classes within the big datasets of images using compact vector representations of local neighborhood property. There are many techniques available to categorize the images. All they need some feature an extraction technique that helps in categorizing them on the basis of the extracted features. A diverse range of extractors and detectors are available vary upon kinds of detection over interest points, repeatability, complexities over time and space. Features such as Blob, Corner, and Grid etc are available along with certain descriptors. The Feature Detection and Extraction tab under The Computer Vision Toolbox TM delivers Corner detection features methods such as Harris–Stephens algorithm, Features from Accelerated Segment Test (FAST)[1], Methods by Shi & Tomasi, Oriented FAST and rotated BRIEF (ORB)[3][4].Blob detection features such as Maximally Stable External Region Tracking(MSER),Speeded Up Robust Features (SURF)[2] and KAZE. In Addition to this certain detectors are also available like FREAK [6], BRISK [5], ORB [4], HOG, SURF and KAZE descriptors. A good mix of a detector along with the corresponding descriptors will be used according to the need of the application.

In this paper classification of images is performed to decide between one of the ten categories present in the WANG Dataset. Despite of the various techniques available in this genre, the methodology used here is unique as it uses multiple feature extraction style technique and implementation. Detection Methods like SURF and GRID are put to use along with descriptors deployed in the Bag of Feature model accelerate the efficiency and fastness of the algorithm.

K-means clustering provides a base to form the dictionary of the visual words known as the bag of visual words, which is the most important framework used in the entire process. Support vector machine a supervised learning technique is deployed as a classifier for the classification purpose. Supervised learning model, SVM is an important classifier and used in many linear and non-linear classification problems. This problem, being a linear classification problem, SVM has been used to train the positive and negative images as car and non-car images respectively and hence perform the classification of test images.

1.2. Detection and extraction of local features

Various patterns and different structures that comprise an image are nothing but the local features of an image. It may be the small patches, edges of an image etc. Basically significance of a feature is the difference it possesses from its corresponding surroundings i.e. local neighborhood like texture, intensity or color of one patch will be different from other. It laid the foundation of most of the image processing algorithms like image tracking, object detection and classification etc. Vector of numeric values (i.e. nothing but a raw pixel value) which describes a piece of image is a descriptor. Key point is again a driving force for a descriptor; it is actually the interest point (intensity of certain patch or corner) of an image. Descriptors along with the key points will define the local features.

A local feature is a good mix of gradient- based and intensity variation approaches includes edges, blobs and regions.**Fig.1.1** shows different techniques for detection and extraction of features according to interest points.

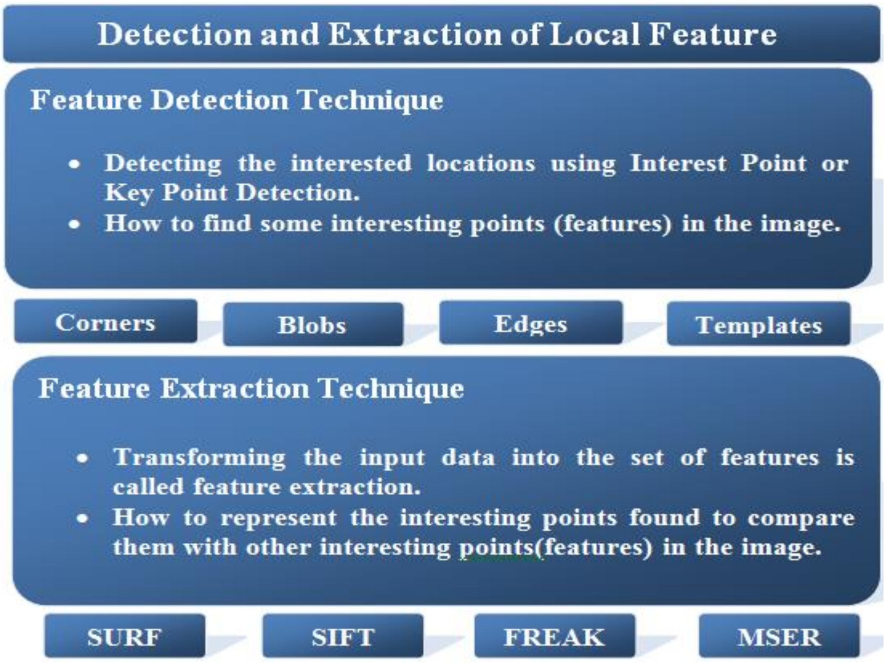


Fig.1.1. Depicting various Detection and Extraction techniques.

One of the main advantages of using the local features is that it represents contents of image precisely without the use of any segmenting technique for the tasks of classification, extraction and Detection.

Repeatable detections	When given two images of the same scene, most features that the detector finds in both images are the

	same. The features are robust to changes in viewing conditions and noise.
Distinctive	The neighborhood around the feature center varies enough to allow for a reliable comparison between the features.
Localizable	The feature has a unique location assigned to it. Changes in viewing conditions do not affect its location.

Table 1.1. Different approaches for local features extractions.

1.2.1. Feature Detection

While processing an image the initial step is to perform Feature Detection. It will detect the low level entities by the selection of unique interest/key points (IP/KP). These selected distinct key points ensure the competency of the Procedure. IP/KP are the features like a Corner, edge or Blob identified and detected by a detector from an image that seem to be useful for further processing. Feature Detector can be selected according to the Region to be abstracted such as Edge, Corner, and Blob etc.

1.2.2. Feature Extraction

Feature extraction, is a form of reduced dimensionality, the extraction of particular set of features from full size input to form a feature vector. Extracted features contain only relevant information about the image and describe the input data accurately.

1.2.3. Feature Extraction using SURF

SURF is inspired by Scale-Invariant Feature Transform (SIFT).SURF was first introduced by Herbert Bay et al. at the 2006, European Conference on Computer Vision. In Computer Vision, Speeded up Robust Features (SURF) is a technique used for object recognition, classification, and registration. SURF is inspired by Scale-Invariant Feature Transform (SIFT) [8].To detect points of interest, SURF uses an integer approximation of Hessian blob detector. Its feature vector is based on the Haar Wavelet response around the interested features [8] [9] [10].

1.3. Bag of visual words

The dictionary of visual words which can be used to define each image in terms of the frequency of each word present in an image is known as the bag of visual words.

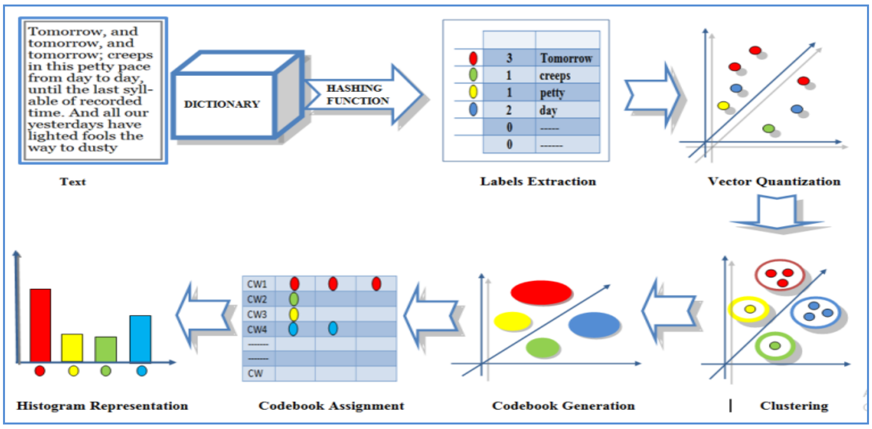


Fig.1.2. Bag of Visual Words Model

Fig.1.2. above demonstrates the bag of visual words model showing the text images as histogram of visual words. Now, this gives a fixed length vector of descriptors irrespective of the number of key points detected. This is one of the advantages of using a bag of visual words model. The fixed length feature vector thus obtained is fed to Support Vector Machine or any other supervised learning model for training.

1.4. Supervised Vector Machine Classifier

In machine learning, support vector machines (SVMs) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis [7][14]. SVM can be defined as a discriminative classifier, formed of an optimal hyper plane. The hyper plane separates the various classes of examples which help in classification. SVM runs an algorithm which helps in finding the hyper plane which is optimal based on the training data that is fed to it. The optimal hyper plane is chosen such that the distance of the hyper plane from the nearest data point on either side is maximum. However, SVM can also be used for performing non-linear classification.SVM is thus an important model used in machine learning and finds its application in object classification.

1.5. Dataset, Implementation and Results

1.5.1. Dataset

Wang database with 10 different categories as listed in Table.1 is used. Each group consists of 100 images in JPEG format, from Wang database, downloaded from the website http://wang.ist.psu.edu/iwang/test1.tar. All these images in the database are natural images. Each image is of size 384*256 or 256*384 pixels. All the images are in the RGB color space [12] [13].

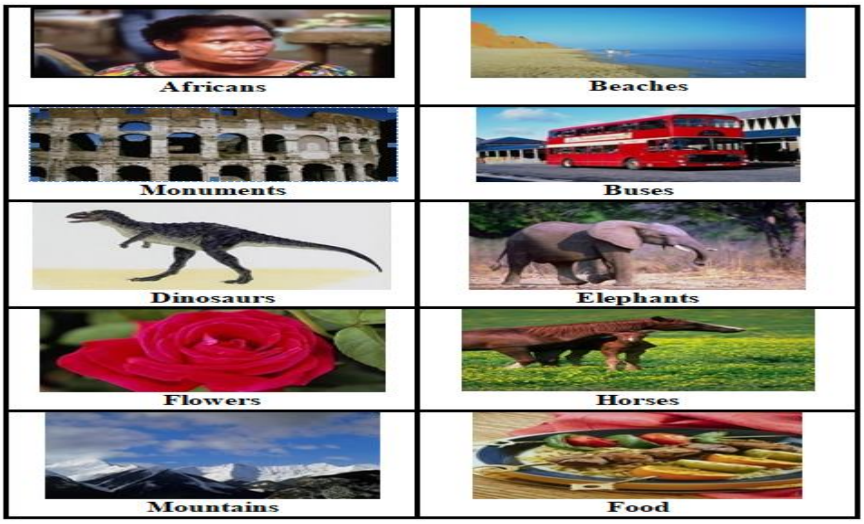


Fig.1.3. WANG Dataset with 10 image categories

1.5.2. BOVW Bag of Feature model

As Shown in Fig.1.4. Below shows the stepwise visual representation of how the process of Bag of feature works. How to create a histogram or clusters of visual words that represent an image and then these histogram/clusters are used to train classifiers.

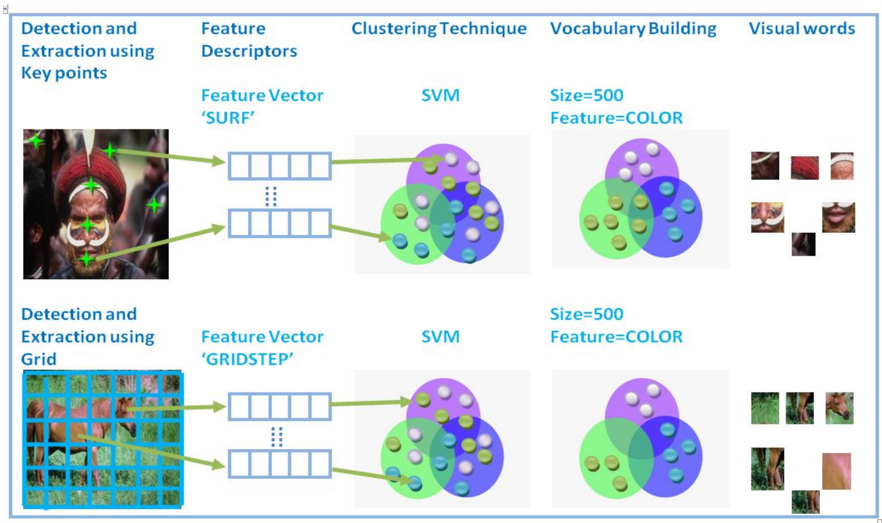


Fig.1.4. BoVW model for representation of images

In the **Fig.1.4.** shown above two techniques are depicted on two different images of African and Horses. The first one is using Point selection dense SURF for feature extraction and the next is using Grid step method. These are two different strategies for feature extraction only while all the other steps are same like vector formation, quantizing and training phase are all same.

1.5.3. Algorithm BOVW for Bag of Feature model

Following are the steps involved in extracting BoVW features model implemented on WANG images:

ALGORITHM		
Step 1	Preprocessing the images	RGB to Grayscale
Step 2	Images are divided into test and training sets.	TEST (30%) TRAINING (70%)
Step 3	Feature Detection and Extraction	SURF DetectSURFFeatures.
Step 4	Training Classifier	Bag of features metric MultiscaleSURFPoints.Metric
Step 5	Generate feature Histograms for each image using Clustering	SVM

Prediction method is used to classify the new images.

1.5.4. Implementation

The Feature Detection and Extraction tab under The Computer Vision Toolbox TM of MATLAB TM delivers methods for creating bag of visual features using bagOfFeatures class [11].

There are two ways for implementation:

Without Extractor

```
bag = bagOfFeatures (..., Name, Value)
```

Using Custom Extractor function

```
bag=bagOfFeatures (imds,'CustomExtractor', extractorFcn)
```

It will return a bag of features that uses a custom feature extractor function to generate the vocabulary features. Table below shows all the arguments of bag class along with their descriptions.

CustomExtractor	**A function handle to a custom feature extraction function**
VocabularySize	Number of visual words held by the bag
StrongestFeatures	Fraction of strongest features to use from each label.
PointSelection	Method used to define point locations for feature extraction
GridStep	Step in X and Y directions defining the grid spacing
BlockWidth	Patch sizes from which SURF descriptor is extracted
Upright	Whether or not to extract upright SURF descriptors

Table 1.2. Arguments (name-value) used with bag of visual features using bagOfFeatures class

While implementing there are two files one is an ExtractorMain file named **'Bag1withExtractorsMain.m'** which is the main file and the other one is **BagOneExtractor2SURF.m** which is a SURF extractor fuction file. By executing the **'Bag1withExtractorsMain.m'** on the various categories of WANG images, with vocabulary size of 500.The SURF features is detected and extracted numerically, after training and testing Results are displayed in Result Section.

extractor = @BagOneExtractor2SURF;
bag = bagOfFeatures
(trainingSets,'CustomExtractor', extractor,'VocabularySize', 500);

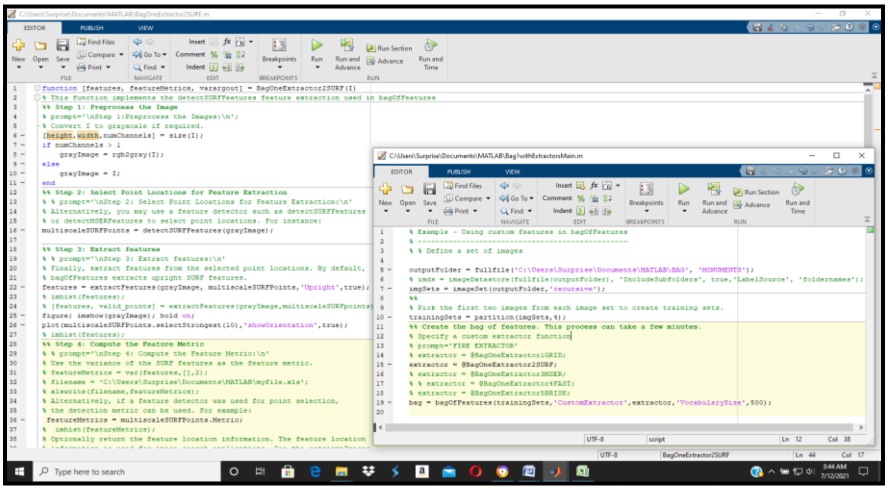


Fig.1.5. shows Editor /Implementation of Extractor and main file for SURF in Matlab

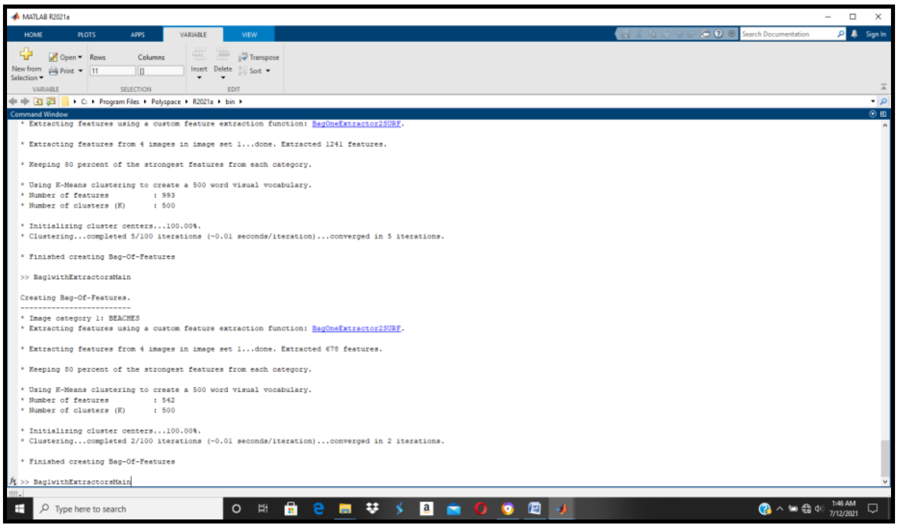


Fig.1.6. shows command line view of results in Matlab.

Meanwhile the implementation could be done with or without the use of Extractor function. The results show some of the wide differences in the number of features extracted. The extracted features and strongest features are more promising with the Extractor function. Therefore, it will be continued in implementation part and corresponding results are displayed in the next section.

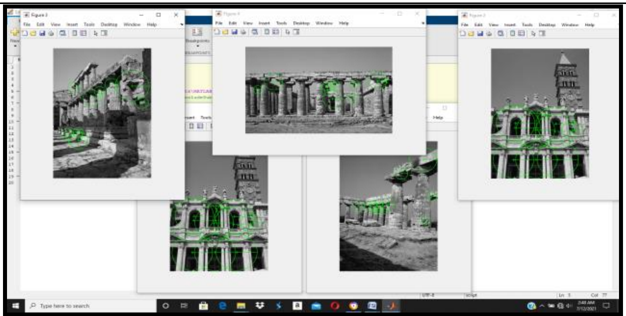
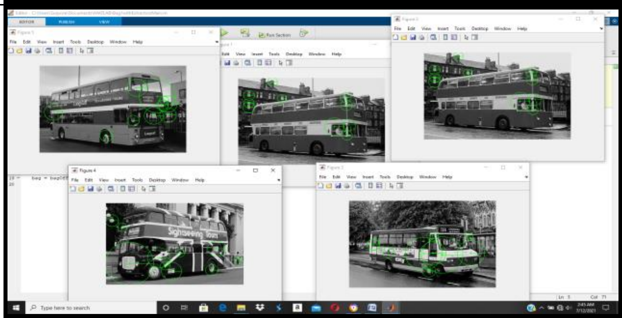
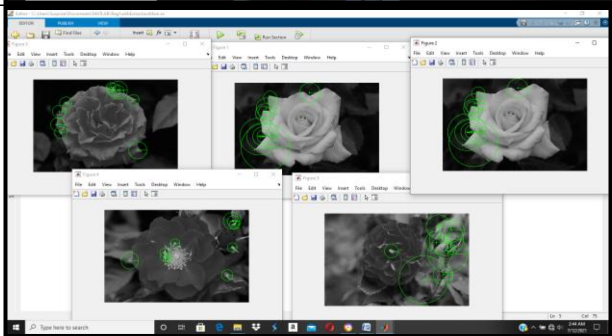
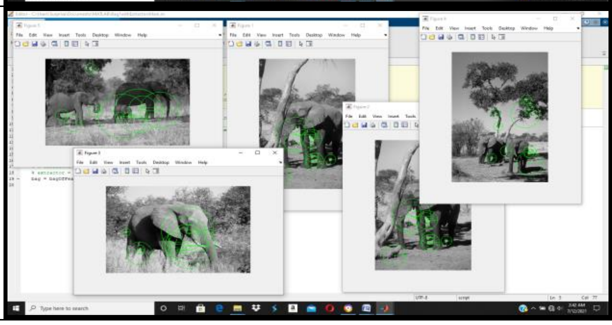
	WITHOUT EXTRACTOR Total images=1000 WANG DATASET Testing set=30% Training set=70%						
			Categories(No. of images per category after preprocessing)	No. of strongest features from each category	Number of features	Number of clusters (K)	Clustering completed
1	Selecting feature point locations	Detector method.	102(*31)	546	55692	500	20/100 iterations (~0.64 seconds/iteration)
	Extracting Features	SURF					
	GridStep	[8* 8]					
	BlockWidth	[32 64 96 128].					
	918 images Extracted features.	202674					
	least number of strongest features	Image category 6 (546)					
	WITH EXTRACTOR FUNCTION Testing set=30% Training set=70%						
			Categories(No. of images per category after preprocessing)	Number of strongest features from each category	Number of features	Number of clusters (K)	Clustering completed
2	Selecting feature point locations	BagOneExtractor2SURF	102(*31)	11059	1128018	500	24/100 iterations (~15.26 seconds/iteration)
	Extracting Features	SURF					
	GridStep	[8* 8]					
	BlockWidth	[16 32 48 64].					
	918 images Extracted features.	4139784					
	least number of strongest features	Image category 4 (11059)					

Table 1.3. Results of WANG images when Bag of features Implemented without Extractor and with Extractor Function.

1.5.5. Results

The graphical representation of detected SURF features in the images of all the 10 categories when implemented using feature extractor function are presented in the **Fig.1.7.** below. The features are marked in green circles taking radial distance from targeted objects. It can be inference from the results that 'flower' category images have least number of feature detected.

Category		

Monuments			
Bus			
Flowers			
Elephants			

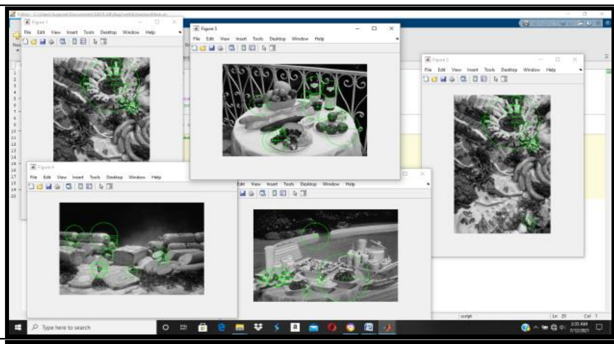
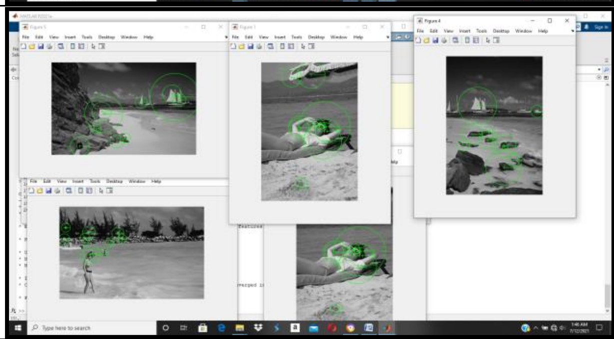
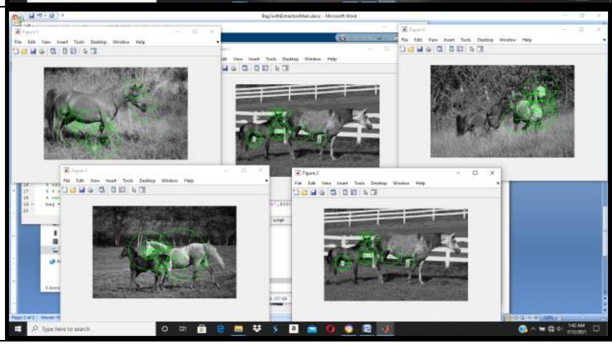
Food		
Beaches		
Horses		

Fig.1.7. Display the detected features in green circles in all the 10 categories of images.

Images from all the 10 categories are trained and tested for detecting SURF features. It uses grayscale images therefore in preprocessing stage all the images are converted from RGB to grayscale.

1.5.6. Analysis

Table 1.5. Depicts the analytical view of Extracted features using SURF detector. The size of clusters (K) is fixed to 500 for all Categories. Number of Iterations and time taken per second for iterations is also specified.

Keeping 80 percent of the strongest features from each category. Extracting features using a custom feature extraction function: BagOneExtractor2SURF.						
S.NO	Category	Number of Feature Extracted	Number of Cluster(K)	Number of features Using K-Means clustering to create a 500 word visual vocabulary.	Clustering completed / 100 iterations	seconds/iteration
1	AFRICAN	1216	500	973	10	~0.01
2	MONUMENT	1869	500	1492	22	~0.02
3	BUS	2053	500	1642	16	~0.02
4	FLOWER	306	245	245	1	~0.00
5	ELEPHANT	1282	500	1026	6	~0.04
6	FOOD	1726	500	1381	11	~0.02
7	BEACH	678	500	542	2	~0.01
8	HOURSE	1121	500	897	10	~0.02
9	MOUNTAIN	1241	500	993	5	~0.01
10	DIANASOUR	734	500	587	1	~0.01

Table 1.5. Analysis of extracted feature using SURF

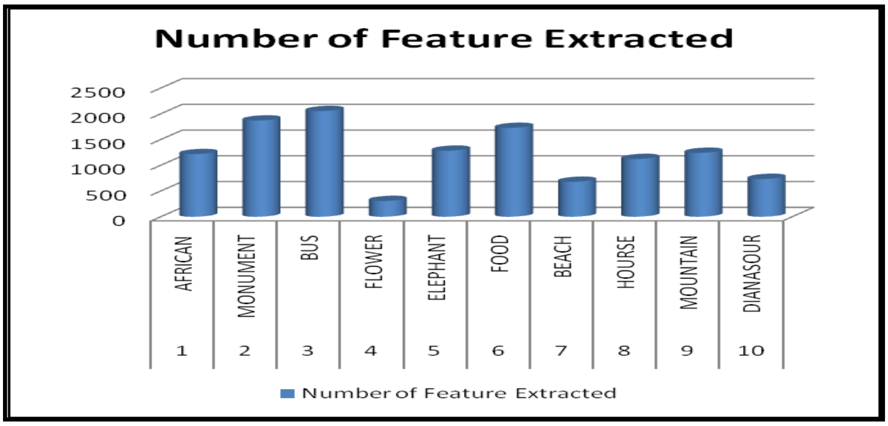


Fig.1.8. Chart showing Number of Features Extracted from images per category

It is cleared from the above figure **Fig 1.8.** that FLOWER category performance is quite weak,as it has lowest number of features detected and extracted.while BUS has maximum number of features extracted as the local feature vectors extracted are based on color features.followed by MONUMENT and FOOD.

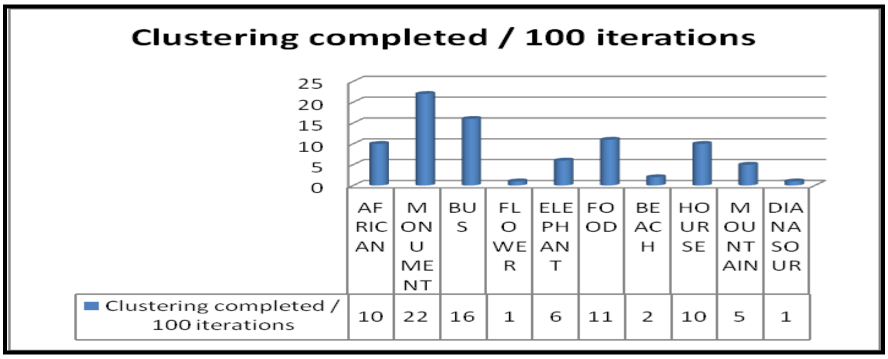


Fig.1.9. Chart showing Number of cluster formed from Features Extracted of images per category

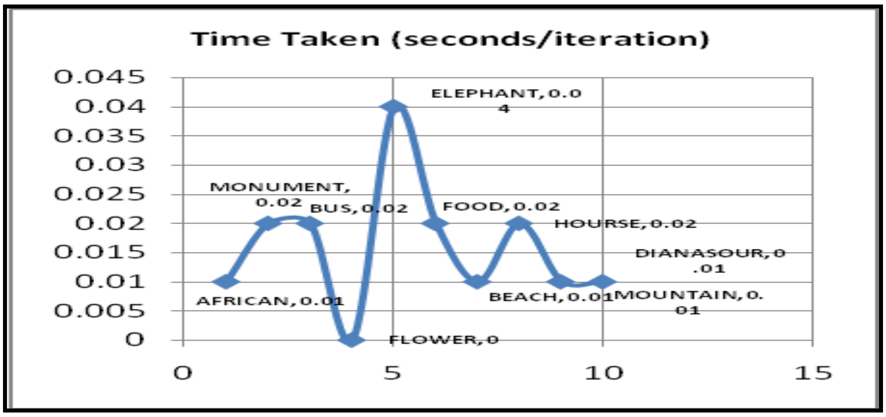


Fig.1.10. chart showing Time Taken in second for completing iteration of Features Extraction of images per category

It is also evident from **Fig 1.9.** that iterations for clustering is directly proportional to the number of features.But it is not the case with time taken for completing iterations as shown in **Fig 1.10.**.

1.6. Conclusions and Future Work

In this paper, we have discussed a new approach Bag of Visual Words for Image retrieval based on Bag of Word model that is successful in text mining. There are many positive points that drive researcher's attention like fast retrieval and bulk amount of features extracted and used. Therefore the retrieval performance goes quite well. In future it could be clubbed with different detectors other than SURF and with/without Custom Extractor function. There is a wide approach available to club it with traditional approaches like color Moments, Histogram or any new color, texture approaches for better classification results. It will be our next promising strategy to combine these Bag feature vectors with Generative Adversarial Networks along with CNN.

References:

[1] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, “SURF: Speeded Up Robust Features,” Computer Vision and Image Understanding, vol. 110,no. 3, pp. 346–359, June 2008.

[2] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, “BRIEF: Binary Robust Independent Elementary Features,” in European Conference on Computer Vision, vol. 6314, September 2010, pp. 778–792.

[3] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “ORB: An Efficient Alternative to SIFT or SURF,” in IEEE International Conference on Computer Vision, November 2011, pp. 2564–2571.

[4] M. C. Stephan Leutenegger and R. Siegwart, “BRISK: Binary Robust Invariant Scalable Keypoints,” in IEEE International Conference on Computer Vision, November 2011, pp. 2548 –2555.

[5] Alahi, A., Ortiz, R., Vandergheynst, P.: FREAK: fast retina keypoint. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 510– 517 (2012)

[6] Wohrer, A.: Model and large-scale simulator of a biological retina with contrast gain control. Ph.D. thesis, University of Nice Sophia-Antipolis (2008)

[7] K. Alsabti, S. Ranka, and V. Singh, An Efficient k-means Clustering Algorithm, Proc. First Workshop High Performance Data Mining, Mar. 1998.

[8] H. Chandel, S.Vatta,”A Vision based Vehicle Detection System”, Communications on Applied Electronics (CAE) – ISSN: 2394-4714, Volume 2 – No.6, August 2015.

[9] J. Farooq, “Object Detection and Identification using SURF and BoW Model”, International conference on Computing, Electronic and Electrical Engineering (ICE Cube), April 2016, DOI: 10.1109/ICECUBE.2016.7495245.

[10] S.Kripa, M.Menon, “Single-Viewpoint Panorama Construction With Wide-Baseline Images using SIFT And SURF Features”, IOSR Journal of Electronics and Communication Engineering, The National Symposium on Antenna Signal Processing & Interdisciplinary Research in 63 | Page Electronics 2017 (ASPIRE-2017).

[11] https://ww2.mathworks.cn/help/vision/ug/image-classification-with-bag-of-visual-words.html

[12]Manish Maheshwari, et al. Content based Color Image Clustering. In: International Journal of Computer Applications 2012; vol. 57, No.15, p. 38-43.

[13]M. Maheshwari. Image Clustering Using Color and Texture. In: Proceeding of 2009 First International Conference on Computational Intelligence Communication Systems and Networks, 2009.

[14] https://en.wikipedia.org/wiki/Support-vector_machine
[15]http://www.emt.tugraz.at/~pinz/data/GRAZ_02/.
[16] http://ai.stanford.edu/~jkrause/cars/car_dataset.html.
[17] V. Vapnik. The Nature of Statistical Learning Theory. NY: Springer-Verlag. 1995.