IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 5, Ver. II (Sep.- Oct. 2017), PP 67-74
www.iosrjournals.org

A Training Independent Visual Codebook using SURF for Large Scale Image Retrieval

*Vipin Gopalakrishnan Nair[1], Nadir N. Charniya[2]
[1]*(ME Student, Electronics and Telecommunication Department, VES Institute of Technology, Mumbai University, India)*
[2]*(Professor, Electronics and Telecommunication Department, VES Institute of Technology, Mumbai University, India)*
Corresponding Author: Vipin Gopalakrishnan Nair

Abstract: *The number of digital images created and uploaded on the internet is approximately over a billion each day. With the advent of modern image capturing devices namely digital cameras and smartphones, there is a huge surge in the quantity of images generated. Retrieving any of these images at a later stage for reference and analysis is cumbersome. Hence we need computer vision algorithms which can sift through the vast image database and return the results with considerable accuracy and speed. In this paper, we detect the local features in an image and describe its neighborhood in terms of SURF descriptor. The feature vectors are converted into code words by indexing and a codebook is generated. The visual codebook is training independent as it doesn't update or modify its contents by itself based on new image features which are added later on. We tested our algorithm on a self made dataset of 2400 images out of which we trained only 240 images and the remaining 2160 images were used for testing. The retrieval method generated results with an average accuracy of approx. 99%. The algorithm can also be applied to larger databases containing thousands of images. This form of image identification method can be used in automatic annotation in search engines. The retrieval method can also be of use in various zones namely medical diagnosis to determine the disease corresponding to a patient's symptoms or in the field of crime prevention or fake product detection, and also intellectual property to name a few.*

Keywords: *Local Image Features, SURF, Visual codebook*

Date of Submission: 11-09-2017 Date of acceptance: 23-09-2017

I. Introduction

The current advancement in technology in the field of imaging devices have led to an explosive increase in the number of images captured every moment. The vast number of images poses a problem for retrieving them at a later instant. In such a scenario, the image retrieval systems gain prominence. Text Based Image Retrieval [1], [2] is one of the most traditional methods of image search where the objects and scenes in an image are annotated manually. This method suffers from the problem of semantic gap [3], [4] as the annotations cannot completely gauge the user intentions resulting in a large number of irrelevant results. Content Based Image Retrieval (CBIR) [5], [6] rely on the content or features within an image rather than depending on metadata like keywords or tags. It classifies the digital image data based on their visual features namely color, texture, shape, intensity etc. The key steps in CBIR involve feature extraction and feature matching with the query image. Feature extraction detects the interest points in all the images in the database and the neighborhood around each interest point is described by a Feature Descriptor [8]. Speeded up Robust Features (SURF) [6],[7], Scale Invariant Feature Transform (SIFT) [9], Gradient Location and Orientation Histogram (GLOH) [10] are a few feature descriptors which make the neighborhood region immune to scale or rotation variance along with clutter and occlusion. For feature matching, similarity between features of query image and images in the database are measured. Most of the recent approaches follow the Bag of Words (BoW) model [11] to index large amount of images in a database. Feature Detectors [6],[7],[9] identify the interest points within an image and the region surrounding each point is described by a Feature Descriptor. Bag of Words model groups the feature vectors computed by the descriptors into clusters by the k-means clustering method [12]. This generates visual words which is the instrument of comparison for matching features with the query image. BoW relies heavily on the feature detector and descriptors to detect the interest points and regions in the image. The most common descriptor used for image retrieval purposes, SIFT creates a scale space representation of the image followed by the Difference of Gaussian (DoG) [13] to detect all the key points. The neighborhood around key points is converted into a vector by assigning them a magnitude and orientation. This allows SIFT to detect objects robustly as it is invariant to scale, rotation and changes in illumination. SURF can outperform SIFT in terms of speed and accuracy. SURF uses integral images [7] instead of actual ones to perform convolutions and this reduces computation time to a great extent. Discrete Wavelet Transform (DWT) [14] is used to retrieve images based on the description of objects with an image. The features are extracted by wavelet decomposition of the wavelets computed for an image. Measuring the similarity between features extracted is a crucial step in CBIR. To speed up the similarity measurement process, algorithms like Distance Metric Learning (DML) [15] and Artificial Neural Network (ANN) are used and the images closest to the query image is retrieved.

II. Related Work

Content Based Image Search based on local features is the most widely adopted method for image retrieval. The key components in the method are detecting and identifying the local features, feature description of the neighborhood, translating the feature vectors into a codebook by any clustering means and indexing the image features, and finally measuring the similarity between features to identify a potential match. In this section we have a look into the related work in each area.

2.1 Local Feature Representation

Local features are the peculiar patterns or typical structures found in an image such that it differs from its immediate neighbors in terms of color, shape, intensity, texture etc. Detection of local features form the building blocks of content based search. So the features should be highly distinctive and should be able to repeat it over various changes. The distinctive pattern in the image may be a point, corner, edge or a small image patch. In [16] Raman Maini *et al* studied and analyzed various edge detection techniques to conclude that Canny's Edge Detection algorithm outperforms all other operators. However it is computationally more complex than other operators. In [17] Tingjin *et al* proposes a fast and effective method based on Harris Corner Detector and azimuth consensus to detect corner or junction points. Harris detector filters out the weak pixels in the flat region to speed up the process of corner detection.

2.2 Feature Description

After the interest points have been detected in an image, the neighborhood of each point is described in terms of a vector which assigns the region a magnitude and orientation. This vector has to remain invariant to scale changes or rotation, illumination effects or change in view point. In [18], Nouman Ali *et al* integrates the visual words of both the descriptors SIFT and SURF to increase the robustness of both features in image retrieval. This method incorporates the positives of both descriptors as SIFT is more robust to scale and rotation changes while SURF takes care of change in illumination. In [19], Xiaojing *et al* proposed an algorithm based on GLOH descriptor to reduce the complex calculations in high dimensional feature vectors. An image is divided into sub regions and each region is given a different weight based on its distinctiveness and then the features from each region are clustered.

2.3 Codebook Generation

The amount of features that are extracted from a single image is so vast that we need represent them in a compact manner for operations. The high dimensional features can be quantized to a cluster of visual words which is then used to represent the image. Most clustering algorithms present for large size codebook generation employ the k-means clustering which is sensitive to the selection of the initial cluster centers. In [20] , W. Zhou *et al* generated a codebook by binarising the SIFT features terming them as BSIFT. In [21] Celebi *et al* presented various initialization methods to increase the efficiency of k-means and analyzed their computational efficiency. As an alternative to k-means, [22] Y. Benezeth *et al* proposed Markov Cluster Algorithm (MCL) which does not require any initialization and the number of classes can be chosen arbitrarily. It is demonstrated that MCL is highly competitive to k-means with any number of clusters selected randomly.

2.4 Indexing

To increase the speed and performance in finding suitable matches corresponding to a query, an index of all relevant features are stored. In [23] Amato *et al* represented every object in a database according to the order of a number of reference objects whose distance from the object is calculated. The orders of these reference objects are then compared to measure the degree of similarity between two objects. Z. Zhou *et al* [24] proposed a global context verification to filter out irrelevant matches during a search scheme. The overlapping region based global context descriptor encodes the global context information in SIFT and allows an efficient verification.

III. Methodology

The proposed algorithm involves two steps namely codebook generation and on-line querying.

3.1 Codebook Generation

1) Select any (M by N) 2-D Image from the dataset.
2) Convert the corresponding image to Grayscale
3) Detect SURF features over the image and return the SURF object points which contain information about SURF features detected in the 2-D gray scale input image. Select 3 octaves and 3 scales per octave (but can vary depending on the image size). As number of octaves increases, larger blobs can be detected as higher octaves employ large filters and subsample image data.
4) Retain only the valid SURF Feature Points on the image.
5) Calculation of SURF descriptors:
   Derive the feature vectors and their locations for the pixels surrounding each interest points. These Feature vectors are the SURF Descriptors.
6) Storage of Feature Vectors as cells of an object:
   Codebook is an index based approach of storing image features in terms of SIFT Features. These Features include scaling, orientation, rotation, shifting, stretching and overall power in the transform domain (1, 1).
7) Repeat step 1 to 5 for all images in the dataset and store the Feature Vectors and Valid SURF Points into a codebook which is a 'X-by-2' (X is total number of images in the dataset) array.

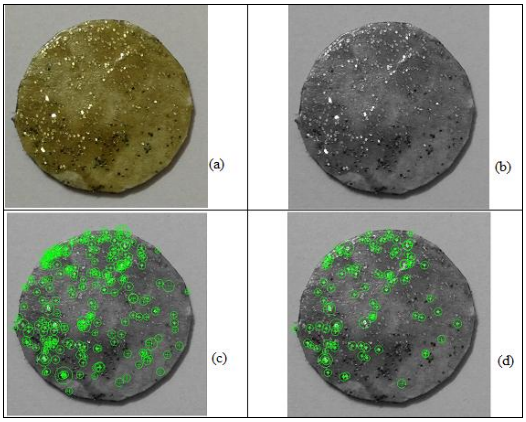

Fig.1. (a) A random Dataset Image. (b) Grayscale Image. (c) SURF Points detected on the image (250) and (d) Valid SURF points on the image

3.2 Querying

1) Accept Query image from user: Prompt a Query Image from the user, store the path of the file and retrieve the image from the corresponding location.
2) Convert the corresponding image into Grayscale.
3) Detect the SURF feature points in the query image.
4) Retain only the valid SURF Feature Points.
5) Extract SURF descriptors for all the interest points:

Derive the feature vectors and their locations for the pixels surrounding each interest points. These Feature vectors are the SURF Descriptors which contain scale, sign of Laplacian, orientation, location, metric, and count (number of index points) of the image. Store the descriptors as an M-by-N matrix of M feature vectors and length N. Find M valid points that contains index of all the keypoints corresponding to each descriptor.

6) Comparison of database images with current image:
   Load the codebook and for each of the 240 images, match its features and valid points with those of the query image. Generate a query table SSD and pair the index of the query with the database and vice versa.

7) Thresholding the number of matched points and maxima class definition:
   Set the threshold for number of matched points at 30 (arbitrary) and the matches below 30 should be ignored. Determine the class which has maximum number of matched points and that is the class of the query result.

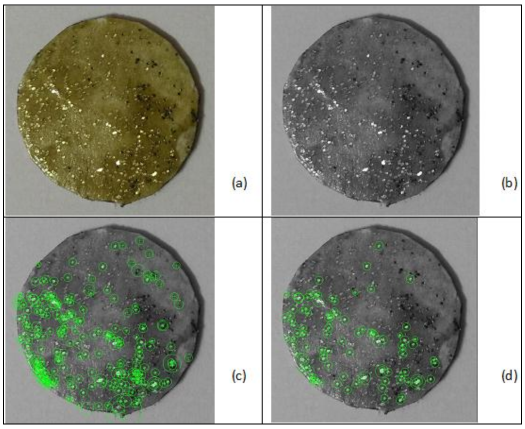

Fig.2. (a) A query image given as input by the user. (b) Grayscale Image. (c) SURF points detected on the query image shown by green circles. (d) The reduced number of valid SURF points shown on the image.

IV. Experiments

We evaluate our algorithm on our self composed dataset. It is generated randomly which is distinct texture seal data using glitter paint. Each pattern is random, unique and difficult to imitate. We classify the parent images into sets of 20 and assign a unique parent image to each set. Further within each set, the parent image is subject to modifications like rotation, change in scale, illumination and viewpoint. Each set contains a total of 120 modified images of the parent image. Thus the dataset has a total member of 2400 images classified into 20 classes. Out of 2400 images in the dataset, we train only 240 images for the codebook generation and test the algorithm on the remaining 2160 images. The training and testing images are independent of each other. In this experiment, we try to identify and match the correct class for each of the 2160 query image.
We encode the algorithm on MATLAB R2016. The program has been run and the results have been recorded on a system with the following specifications.

TABLE 1 SYSTEM SPECIFICATIONS

Processor	Intel® Pentium® Dual Core Processor T23710 @ 1.73GHz
Installed Memory (RAM)	2 GB
System Type	64-bit operating system, x-64 based processor

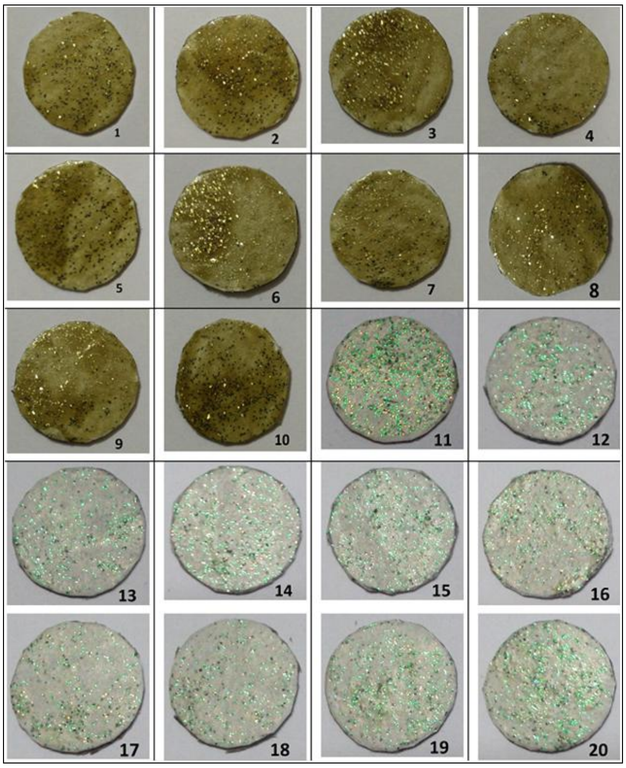

Fig.3. Dataset Images (1-20) corresponding to member of each class created.

We give an image from class 9 as a query. The results are observed as follows.

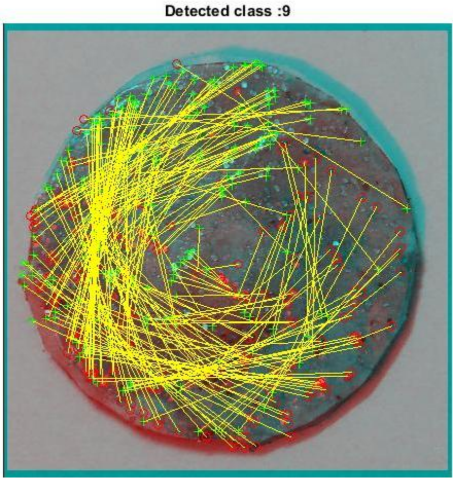

Fig.4. The query image superimposed onto the matched dataset image, and the identical SURF points being matched on both images. The nature of matched lines indicates that the query image is a rotated version of the dataset image.

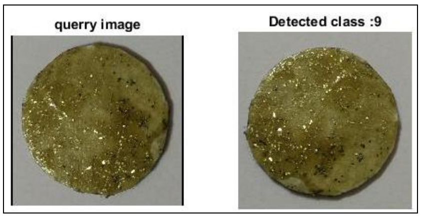

Fig.5. The final output, indicating that the query image belongs to class 9.

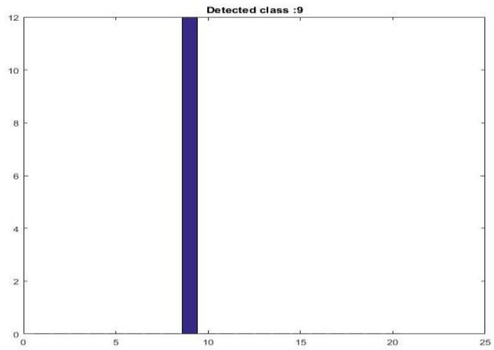

Fig.6. Histogram showing the maximum matches belonging to class 9.

4.1 Accuracy for Image Search Results

After generating the codebook and indexing the images, we test our algorithm with random query images from the different classes. A total iteration of 32 images per class is performed, taking the total number of iterations to 640. The algorithm showed exemplary accurate results for each class with the average accuracy being approximately 99%.

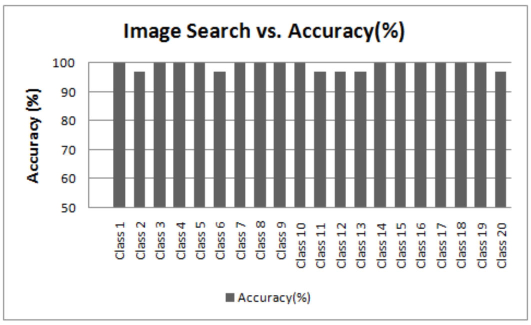

Fig.7. The graph representing the performance of SURF for normal image search.

4.2 Rotations vs. Accuracy

Out of the 2160 images for testing, we rotate 32 images from each query class by varying angles of orientation and the resultant images were fed as query. We performed 32 iterations per class, 640 iterations in all. The average accuracy for the image rotated images was 98.90%.

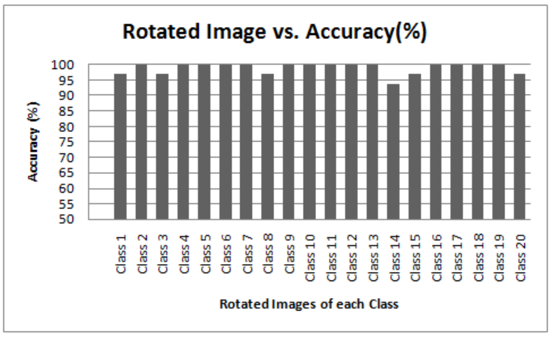

Fig.8. Rotated Image Class vs. Accuracy. Accuracy ranged from 94 – 100%

4.3 Performance Analysis: Intensity Variations vs. Accuracy

The intensity of the images in the query data were varied between 0 and 256 corresponding to the darkest and brightest levels of illumination. As illumination increases gradually from 0 to 50, accuracy of detected images start increasing and it becomes fairly constant for intensity in the range of 75 to 225. Hence we could achieve considerable success with dark and very bright images too.

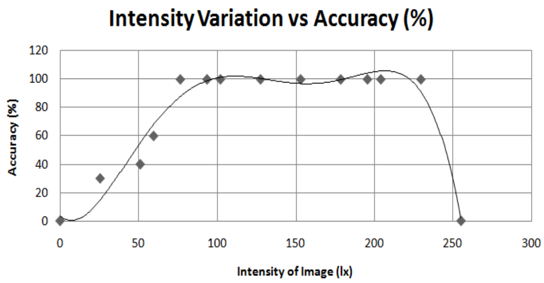

Fig.9. Intensity of the image varies between 0-256. Accuracy remains fairly constant in the mid range of intensity.

4.4 Image Blur vs. Accuracy

The query images are blurred with blur varying from 0% to 100%. Our algorithm gives considerable accuracy when the blur is approximately 30%, post which it cannot detect the class to which the query belongs.

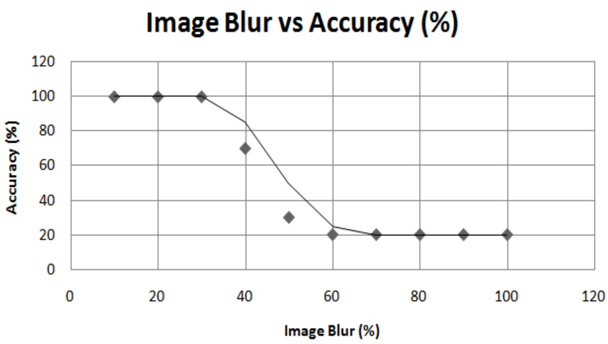

Fig.10. Image blur vs. Accuracy

4.5 Image Crop vs. Accuracy

The images are cropped (both centre cropped and corner cropped) and the graph is plotted with accuracy as a function of the amount of image retained. Our method detects the images even at an approximate image crop of 80%.

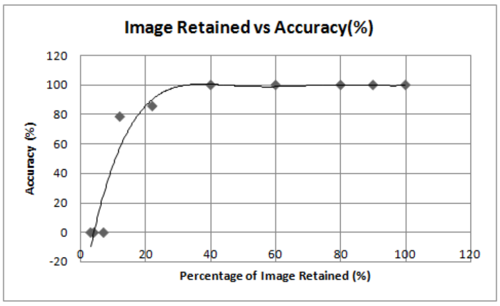

Fig. 11. Accuracy as a function of Image cropped.

V. Discussions

We compare the performance of our approach with each of the feature description algorithms using ORB, FREAK, BRISK, and BSIFT.

TABLE 2 PERFORMANCE COMPARISON OF ALL APPROACHES

Method	ORB	FREAK	BRISK	BSIFT	SURF (our approach)
Average Time Cost per Query(sec)	0.11	0.58	0.51	0.30	1.187
Time cost to index Features (sec)	5.89	10.80	10.24	18.86	1.283
Mean Average Precision (mAP)	0.28	0.10	0.18	0.54	0.81

VI. Conclusion

Large scale image search using SURF competes well with other methods in terms of mean average precision as it is highly accurate and discriminates well. However the time required to search is slightly on the higher side. This method can be quite useful for applications where accuracy is of paramount importance.

References

[1] W. Li, L. Duan, D. Xu and I. W. H. Tsang, "Text-based image retrieval using progressive multi-instance learning" in 2011 International Conference on Computer Vision, pp. 2049-2055, Nov. 2011.

[2] Jose Costa Pereira and Nuno Vasconcelos, “Cross-modal domain adaptation for text based regularization of image semantics in image retrieval systems” in Computer Vision and Image Understanding, vol.124, pp. 123-135, Jul. 2014.

[3] Rong Zhao and W. I. Grosky, "Narrowing the semantic gap - improved text-based web document retrieval using visual features," in IEEE Transactions on Multimedia, vol. 4, no. 2, pp. 189-200, Jun 2002.

[4] Xirong Li, Tiberio Uricchio, Lamberto Ballan, Marco Bertini, and Cees Snoek, “ Socializing the Semantic Gap: A Comparative Survey on Image Tag Assignment, Refinement, and Retrieval” in ACM Computing Surveys (CSUR), vol. 49, Jul. 2016.

[5] Karegowda Asha Gowda and Bharathi PT, “Enhancing CBIR Performance Using Evolutionary Algorithm-Assisted Significant Feature Selection: A Filter Approach” in International Journal of Applied Research on Information Technology and Computing, vol.7, pp. 53-59, May.2016.

[6] C. Benavides Alvarez, J. Villegas Cortez, G. Roman Alonso and C. Aviles Cruz, "Face Classification by Local Texture Analisys through CBIR and SURF Points" in IEEE Latin America Transactions, vol. 14, no. 5, pp. 2418-2424, May. 2016.

[7] Herbert Bay, Andrea Ess, Tinne Tuytelaars and Luc Van Gool, “Speeded-Up Robust Features (SURF)”, in Computer Vision and Image Understanding, vol.110, no. 3, pp.-346-359, Jun. 2008.

[8] S. H. Cheon, I. K. Eom and Y. H. Moon, "Fast descriptor extraction method for a SURF-based interest point," in Electronics Letters, vol. 52, no. 4, pp. 274-275, Feb. 2016.

[9] D. G. Lowe, "Object recognition from local scale-invariant features," in Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, vol.2 , pp. 1150-1157, Sept. 1999.

[10] K. Mikolajczyk, B. Leibe and B. Schiele, "Local features for object class recognition," in Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, Beijing, vol. 2 , pp. 1792-1799, 2005.

[11] M. M. Farhangi, M. Soryani and M. Fathy, "Informative visual words construction to improve bag of words image representation," in IET Image Processing, vol. 8, no. 5, pp. 310-318, May. 2014.

[12] C. Boutsidis and M. Magdon-Ismail, "Deterministic Feature Selection for K-Means Clustering," in IEEE Transactions on Information Theory, vol. 59, no. 9, pp. 6099-6110, Sept. 2013.

[13] M. Basu, "Gaussian-based edge-detection methods-a survey," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol.32, no. 3, pp. 252-260, Aug. 2002.

[14] E. Gupta and R. S. Kushwah, "Combination of global and local features using DWT with SVM for CBIR," in 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), Noida, pp. 1-6, 2015.

[15] S. C. H. Hoi, Wei Liu, M. R. Lyu and Wei-Ying Ma, "Learning Distance Metrics with Contextual Constraints for Image Retrieval," in 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), pp. 2072-2078, 2006.

[16] Raman Maini, Himanshu Aggarwal, “Study and Comparison of Various Image Edge Detection Techniques”, in International Journal of Image Processing (IJIP), vol. 3, no. 1, pp. 1-11, Feb. 2009.

[17] Tingjin, Luo, Zhang Jun, and Lian Lin, "Speed up junction detector based on azimuth consensus by Harris corner" In Optical Review, vol.21, no.2, pp. 135-142, 2014.

[18] Nouman Ali, Khalid Bashir Bajwa, Robert Sablatnig, Savvas A. Chatzichristofis, Zeshan Iqbal, Muhammad Rashid, and Hafiz Adnan Habib. "A novel image retrieval based on visual words integration of sift and surf" in PloS one, vol.11, no. 6, pp. e0157428, 2016.

[19] Xiaojing, Gao, Luo Xiaoling, Li Yufeng, Liu Xia, Pan Xin, and Zhang Kaixue, "Face Recognition Based GLOH Descriptor and Integration of Local Features" in Proceedings of International Conference on Internet Multimedia Computing and Service, ACM, p. 71. Jul. 2014.
[20] W. Zhou, H. Li, R. Hong, Y. Lu and Q. Tian, "BSIFT: Toward Data-Independent Codebook for Large Scale Image Search," in IEEE Transactions on Image Processing, vol. 24, no. 3, pp. 967-979, Mar. 2015.
[21] Celebi, M. Emre, Hassan A. Kingravi, and Patricio A. Vela. "A comparative study of efficient initialization methods for the k-means clustering algorithm" in Expert systems with applications, vol.40, no. 1, pp.200-210, 2013.
[22] Y. Benezeth, A. Bertaux and A. Manceau, "Bag-of-word based brand recognition using Markov clustering algorithm for codebook generation" in 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, pp. 3315-3318, 2015.
[23] Amato, Giuseppe, Claudio Gennaro, and Pasquale Savino, "MI-File: using inverted files for scalable approximate similarity search" in Multimedia tools and applications, vol.71, no. 3, pp.1333-1362, 2014.
[24] Z. Zhou, Y. Wang, Q. M. J. Wu, C. N. Yang and X. Sun, "Effective and Efficient Global Context Verification for Image Copy Detection," in IEEE Transactions on Information Forensics and Security, vol. 12, no. 1, pp. 48-63, Jan. 2017.

IOSR Journal of Computer Engineering (IOSR-JCE) is UGC approved Journal with Sl. No. 5019, Journal no. 49102.

Vipin Gopalakrishnan Nair. “A Training Independent Visual Codebook using SURF for Large Scale Image Retrieval.” IOSR Journal of Computer Engineering (IOSR-JCE) , vol. 19, no. 5, 2017, pp. 67–74.