HANDWRITTEN NANDINAGARI IMAGE RETRIEVAL SYSTEM BASED ON MACHINE LEARNING APPROACH USING BAG OF VISUAL WORDS

Prathima Guruprasad[1], Dr. Jharna Majumdar[2]
[1]Dept. of Computer Science and Engineering, NMIT, Université of Mysore
PB No. 6429, Govindapura, Gollahalli, Yelahanka, Bangalore, India
[2]Dept. of CSE(PG) and Center for Robotics Research, NMIT, Université of Mysore
PB No. 6429, Govindapura, Gollahalli, Yelahanka, Bangalore, India

ABSTRACT

Human handwriting comes with different styles, highly variable and inconsistent which makes recognition of rare handwritten Nandinagari scripts extremely challenging. This paper proposes a Bag of Visual Words (BoVW) based approach to retrieve similar Handwritten Nandinagari character images from visual vocabulary. This technique is useful in the retrieval of images from a large database, efficiently and accurately also helps in achieving required scalability. The character images are represented as histogram of visual words present in the image. Visual words are quantized representation of local regions, and for this work, SIFT descriptors at key interest points are used as feature vectors. For the efficient handwritten Nandinagari character retrieval system, the optimal number of clusters to be chosen is 32 with the maximum mean Average Precision (mAP) as 0.904.

Keywords: Character Image retrieval, Bag of Visual words, SIFT Features, Code book generation, Clustering, Indexing and retrieval, Visual Vocabulary.

I. INTRODUCTION

The Indian regional languages are cursive in nature so, reliable and efficient Recognition system is still not available.Many of these languages come in different forms subjected to changes region to region and time to time. They have rich heritage which are in manuscript Palm Leaf forms and also in handwritten forms. Recognition is even more challenging in case of rare ancient handwritten Nandinagari manuscripts which is earlier form of Devanagari scripts. In recent years, large-scale image indexing and retrieval show sufficient scope in both corporate sector as well as in research community. The focus of this work is to use invariant descriptors to represent such rare manuscripts. Retrieval of similar images from a large image database is a challenging problem.

This is because the memory requirement of the images, time of processing, indexing and retrieval time will grow as database size grows. Also effective and accurate retrieval when the images are with various sizes, orientations and occlusions is a matter of concern. In the proposed model, the Bag of Visual Words (BoVW) concept is used for effective handwritten Nandinagari character image retrieval as the technique is widely used for text based system. This is the way in which human mind interpret the similar objects by generating the visual words. In BoVW, the Nandinagari character image is represented as discrete visual words, the quantized representation of large feature descriptors extracted from the images. During recognition, similar images are retrieved by effectively computing the histogram of visual word frequencies with the closest histogram matching in very less time. In order to process the handwritten Nandinagari script which is having complex 52 characters which include 15 vowels and 37 consonants, the following steps are used. First the Nandinagari character dataset is prepared by converting them to gray scale images in different sizes and orientations.

Then detect the interest points and generate the invariant feature descriptors around the key points using SIFT (Scale Invariant Feature Transform) technique. This is the pre-processing step for BOVW approach which is based on data visualization concept. Next is code book generation by quantizing local SIFT descriptors into visual words. This is also called as vocabulary generation. The last step is indexing and searching similar images from the visual vocabulary. This approach is highly scalable. To measure the recognition accuracy mean Average Precision method is used.

II. REVIEW OF LITERATURE

Large scale image retrieval systems is using the BOVW model in which each image can be viewed as a sparse vector of visual words to text-retrieval system[2]. BOVW model is designed for the local descriptors of images to describe regions around the key points which is detected in the images [3]. Instead of global features, an image can have a set of patches around key interest points. And local descriptor of 128-dimension SIFT features is used to be a good way to represent the characteristics of these patches [4][5]. But when such local descriptors for each image is extracted, the total number of such features is very huge. And comparing and searching similar matches for each local descriptor in the query image becomes too cumbersome. Therefore, BOVW is proposed as a way to solve this problem by compact representation of these descriptors into visual words, which decreases the size of the descriptors [6][8]. This is helpful in scalable indexing and fast search on the vector space.

III. PROPOSED METHODOLOGY

The training phase in this proposed model as shown in Figure 1 of recognizing the hand written Nandinagari scripts using data visualization has following steps. The visual vocabulary is used for matching query images in the recognition phase.

Algorithm Steps

Step 1: Preparation of dataset: Standard dataset for handwritten Nandinagari characters is not available .As input, Handwritten Nandinagari characters of different orientations and styles are prepared. A total of 1049 handwritten Nandinagari character images are taken for result analysis with minimal or without Pre-processing by converting them to Greyscale.

Step 2: Feature Extraction: This is done using SIFT method. First identify the key interest points which usually lie on high-contrast regions of the image, such as corners which remain invariant even change in scale, rotation and illumination. Then 128 feature descriptors extracted for each indentified key points on the images. The steps are as shown in Figure 2.

Step 3: Representation of visual words: The large set of SIFT local feature descriptors are represented as visual words using BOVW approach. This technique groups the similar descriptors using k means clustering [7] approach and codebook of k X 128 dimension is generated by k cluster centroids. This is based on number of clusters formed and called as k visual words. Each key point descriptor of an image is assigned to the closest cluster centroid. The number of assignments is represented as histogram of frequencies of visual words and generated by aggregating of local SIFT descriptors.

Step 4: Then indexing and retrieval of similar characters from the codebook by querying is done using nearest neighbour approach in the recognition phase. Top N similar images retrieved from the database.

IV. EXPERIMENTAL RESULTS

As input, Handwritten Nandinagari characters of different sizes 256 X 256, 384 X 384,512 X 512 and 640 X 640 and 5 Rotations are 0o, 45o, 90o, 135o and 180oare considered. This forms a 1049 characters in the database.

For SIFT feature extraction, a set of images (.jpg or .png files) contained in a folder writes the extracted features into text files. For each image, this first key points of each image is identified and 128 feature descriptors are generated for each of these key points. In this way a total of 1049 input images are fed to the feature extraction module and feature files are written in a directory which represents different types and styles of images. Average number of interest points for these images are ranging from 63 to 268 key points. And the average execution time is ranging from 45 milli seconds to 218 milli seconds and the feature descriptor size is ranging from 35 MB to 166 MB. This feature analysis statistics is as shown in Figure 3.

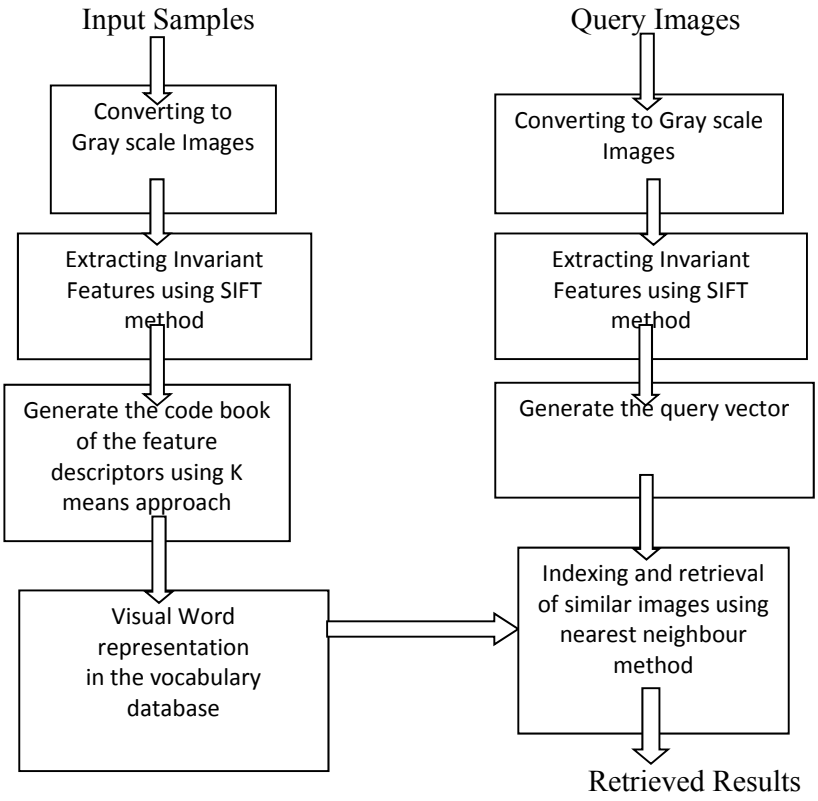

Figure 1: Proposed model

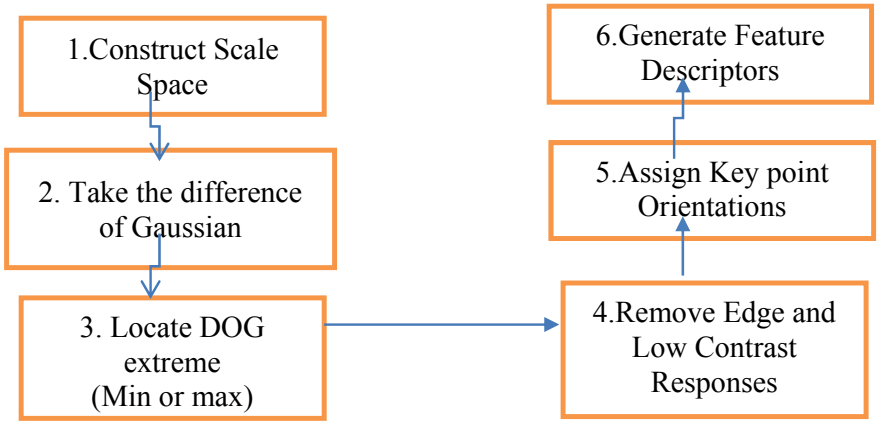

Figure 2: Steps in SIFT Method

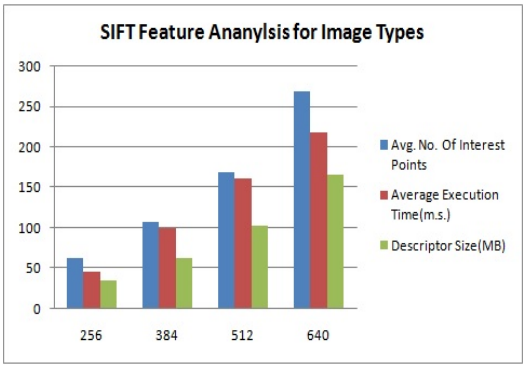

Figure 3: Average no. of interest points, Average execution time and Feature Descriptor size for Character images

In SIFT codebook generation phase K means clustering approach is used and code book is generated only once. The code book file stores only cluster centres aFnd the number of cluster centres depends on number of clusters formed based on the value of K. To analyze the performance the codebooks with 8, 16, 24, 32 and 52 cluster centroid are generated. These are called visual words.

Table I. Code book size and Clustering time different number of clusters

Sl. No.	No. of Clusters	Index Size(KB)	Indexing Time(m.s.)
1	8	8,176	95250
2	16	16,895	88513
3	24	24,648	93430
4	32	33,251	91417
5	52	53,761	85466

These internally get mapped to all key points present in that particular cluster. The table 1 shows the clustering time and code book size for different values of K.

Then character images are represented by frequencies of visual words i.e. as histogram of visual words present in the image. Then for all images in the database, index is generated and BDB (Berkeley Database) library for indexing purpose is used.

This is based on persistent storage of the linear indexing approach and assigns the index for the images stored in the image folder. The index size varies from 8176 KB to 53761 KB for different values of K ranging from 8 to 52 clusters as shown in table 2.

Table 2 *:* Index size and Indexing time different number of clusters

Sl. No.	No. of Clusters	Index Size(KB)	Indexing Time(m.s.)
1	8	8,176	95250
2	16	16,895	88513
3	24	24,648	93430
4	32	33,251	91417
5	52	53,761	85466

The retrieval of similar images is obtained by comparing the query vector against quantized BoW feature representation. Thus the answer using the index of the images folder and the query image vector with nearest neighbour method is computed.

Here top N similar images are retrieved from visual vocabulary. From the image set a representative of 10 query samples as shown in table 3are taken for analyzing the performance of this proposed framework.

Table 3: Query Image samples

Query Name	Query Image	Size and Orientations
A		512X512, 0 Deg.
AE		256X256,45 Deg.
AE		256X256, 0 Deg.
RE		640X640, 0 Deg.
KA		512X512, 0 Deg.
CA		384X384, 0 Deg.
BA		384X384, 0 Deg.
PA		384X384, 0 Degr.
YA		512X512, 0 Deg.
LA		512X512, 0 Deg.

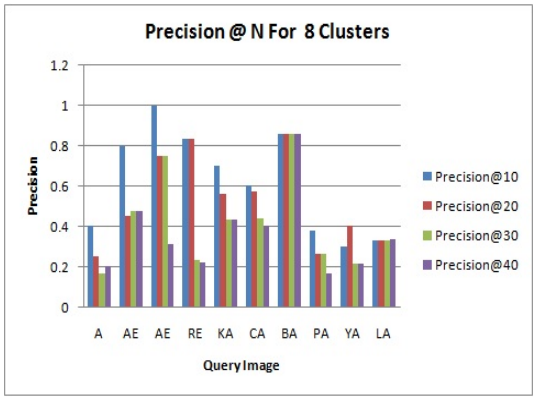

Figure 4: Precision at 10,20,30 and 40 for 8 clusters

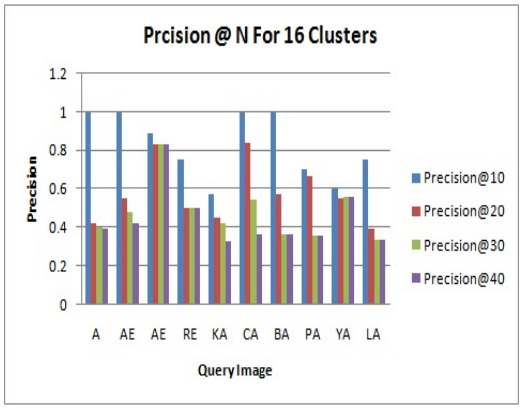

Figure 5: Precision at 10,20,30 and 40 for 16 clusters

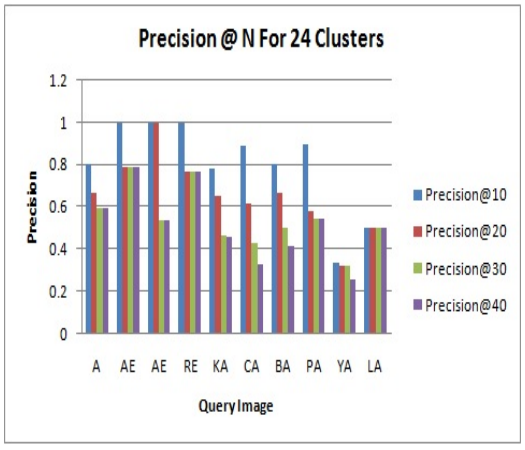

Figure 6: Precision at 10,20,30 and 40 for 24 clusters

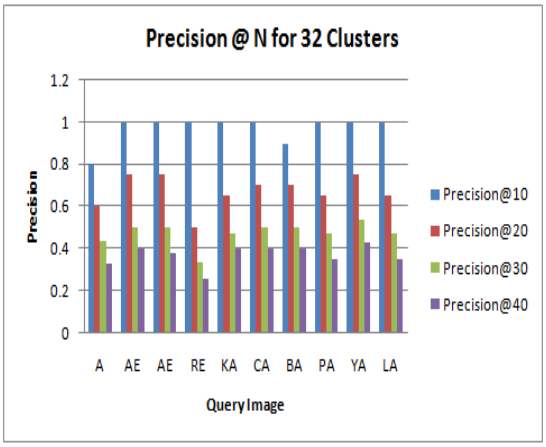

Figure 7: Precision at 10,20,30 and 40 for 32 clusters

The precision at N shows how accurate top N retrieved results. The precision at N shows how accurate top N retrieved results. The retrieval performance is measured from visual vocabulary by precision at 10, precision at 20, precision at 30, precision at 40 at different cluster numbers ranging from 8,16, 24, 32 and 52 clusters as shown in Figure 4 to 8. The Average Precision (AP) is shown in Figure. 9.

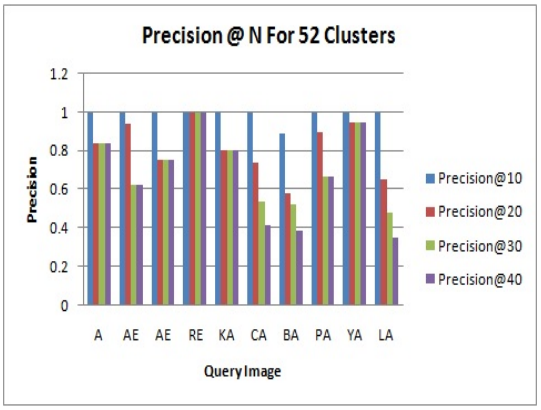

Figure 8: Precision at 10,20,30 and 40 for 52 clusters

Then Mean Average Precision (mAP) and retrieval time for all query images is computed as shown in Table 4. From below table, we conclude that mAP is maximum for a specific number of clusters where the recognition accuracy is also more. Thus for the expert system on handwritten Nandinagari character recognition the optimal number of clusters to be chosen is 32 with the maximum mAP as 0.904 .

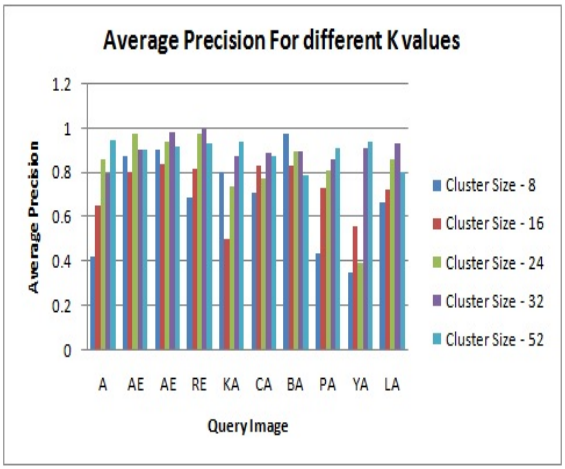

Figure 9: Average Precision for of query Images for different cluster sizes

Table 4: Mean Average Precision for different number of clusters

Sl. No.	No. of Clusters	mAP	% - MAP
1	8	0.681938231	68.19382306
2	16	0.727342196	72.73421962
3	24	0.822027367	82.2027367
4	32	0.904111858	90.41118576
5	52	0.895355485	89.53554851

V. DISCUSSION AND CONCLUSION

The ancient handwritten Nandinagari manuscripts contain valuable information on various discipline but there are only a few scholars who can read and interpret this fluently. The invariant feature descriptors are extracted using SIFT method. The codebook is generated using K means approach. Here each image descriptor is mapped to cluster centres and represented as visual words. Performance of this proposed framework is analyzed by taking different number of clusters and finding mean average precision in each case. This is to select the appropriate number of clusters for generating SIFT code book for handwritten Nandinagari characters. As a future work, extending this research to different feature extraction and classification methods to minimize error rates.

VI. REFERENCES

[1] P. Visalakshi, "Nandinagari Script", DLA Publication, First Edition, 2003.

[2] Jialu Liu, "Image Retrieval based on Bag-of-Words model", arXiv Preprint arXiv: 1304.5168, 2013.

[3] D. G. Lowe, “Distinctive Image Features From Scale-Invariant Key Points”, IJCV, vol. 60, no. 2, pp. 91–110, 2004.

[4] Mortensen, E.N., Deng, H., Shapiro, L., "A SIFT descriptor with global context", In Computer Vision and Pattern Recognition , 20-25 IEEE, Vol. 1, 184-190, June 2005.

[5] Ives Rey-Otero, Jean-Michel Morel and Mauricio Delbarcio, “An analysis of Scale-space sampling in SIFT”, ICIP, 2014.

[6] Ravi Shekhar, C.V. Jawahar, “Word Image Retrieval using Bag of Visual Words”, IEEE, DOI 10.1109/DAS.2012.96, 2012.

[7] Akanksha Gaur, Sunita Yadav," Handwritten Hindi Character Recognition using KMeans Clustering and SVM",4th International Symposium on Emerging Trends and Technologies in Libraries and Information Services, 2015.

[8] Ayman Al-Dmour, Mohammed Abuhelaleh, "Arabic Handwritten Word CategoryClassification using Bag of Features ", Vol.89. No.2, JATIT, 2016.

[9] Yin Zhang, Rong Jin, Zhi-Hua Zhou, " Understanding Bag-of-Words Model: A Statistical Framework", International Journal of Machine Learning and Cybernetics, Dec 2010.

[10] Stephen O’hara , Bruce A. Draper, " Introduction to the Bag of Features Paradigm for Image Classification and Retrieval",arXiv:1101.3354v1, Jan 2011.

[11] Rakesh Mehta, Ondrej Chum, Jir Matas, “Towards Visual Words to Words", 13th International Conference on Document Analysis and Recognition (ICDAR), IEEE, 2015.

[12] Antonio Foncubierta Rodr´ıguez, Alba Garc´ıa Seco de Herrera, Henning M¨uller, “Meaningful bags of words for medical image classification and retrieval", Health Monitoring and Personalized Feedback Using Multimedia Data, Springer 2015.

[13] Patrice Bellot, Marc El-Bèze, " A Clustering Method for Information Retrieval",ICSC 1999.

[14] Marc T. Law, Nicolas Thome and Matthieu Cord, "Bag-of-Words Image Representation:Key Ideas and Further Insight", Advances in Computer Vision and Pattern Recognition, Springer 2014.