Inscription Image Retrieval Using Bag of Visual Words

N Jayanthi and S Indu
1. Assistant Professor, ECE Department, Delhi Technological University, Delhi-110042, India
2. Professor, ECE Department, Delhi Technological University, Delhi-110042, India

Abstract: This paper presents a technique for efficient and veracious retrieval of ancient inscriptions and manuscripts from a large database of images by using the Bag of Visual Words (BoVW) technique. The proposed method can be used to recognize inscription images across the world. SURF (speeded up robust features) is used as an image feature extractor. A visual vocabulary is created by representing the image as a histogram of visual words which helps in the retrieval process. Usage of SURF ensures scalability, faster processing better results with darkened and blurred images. We demonstrate the method on a combination of 300 inscriptions images comprising of several inscription around the world.

Keywords: BoVW, Inscriptions, SURF.

1. INTRODUCTION

Since ancient times inscriptions have been used as a medium of storing and passing the information from one generation to the next. These inscriptions give us an insight into their worlds and carry answers to many of our questions. For ages we have tried to decipher their meanings and determine their origins but with the increasing amount of data and constant digitization of the world it is only natural that this treasure of knowledge is archived as well.

Once the inscription images are archived we move to the next step which is the retrieval process. With a database having innumerable images it is necessary to ensure that – i) we have the fastest possible retrieval solution. ii) The retrieval technique is language independent considering the inscriptions from various cultures are inscribed in separate languages. iii) It should be able to perceive images that have undergone damages such as blurring and darkening over time (more resilient to noise).

Different methods have been suggested for efficient retrieval, word spotting is popular for recovering the relevant images without recognition. Various word spotting techniques have been used, such as the Dynamic Time Warping (DTW) algorithm [3], Hidden Markov Model (HMM) [2] etc. DTW compares a series of feature vector for image retrieval, though successful in many cases it has a drawback when it comes to large databases, having a computation time of 1 second for comparing two word images hence it loses its practicality. In [5] Zagoris et al. define a segmentation based word spotting technique using document specific local features that was tested on two historical datasets. In [1] Rajakumar and Bharathi suggest a contour let transform to recognize Tamil characters from stone inscriptions. Sankar and Mamantha [4]
have suggested automatic word annotation for document retrieval where the digital image is assigned with a metadata by the system automatically.

The Bag of Visual Words has come across as one of the most efficient techniques for document image retrieval [6], [7], [8], [9], [10]. Augereau et al. [6] talk of combining visual and textural features for document classification and retrieval using BoW and BoVW. Shekhar and Jawahar [13] use the BoVW for image retrieval using SIFT as the feature vector. Aldavert et al. [11] work on the George Washington data set for keyword spotting using BoVW. In [13] it has been observed that SIFT descriptor is not very robust for all possible image degradations and hence it calls out for the need of a more robust descriptor. Nabeel et al. [12] have summarized the performance of SIFT and SURF on several datasets showing SURF to be outperforming SIFT in case of blurred images and darkened images. SIFT though known to work well with images from different datasets is replaced by SURF here which has proved to be more efficient when it comes to the highly blurred inscription images.

With the inscriptions dating back to medieval times the effect of blurring tends to become very prominent in the inscribed texts or pictorial representations. The biggest problem that then arises is how to retrieve them accurately in the shortest time span possible. To overcome this issue we have come up with an image retrieval procedure that employs the bag of visual words (BoVW) technique using the SURF (Speeded Up Robust Feature) descriptor. SURF has been known to perform exceptionally well with the shortest execution time and hence we have replaced the originally used SIFT (Scale Invariant Feature Transform) with the former technique. SURF works well with blurred and darkened images. The BoVWs technique ensures that the correct output is obtained with its help. The major advantage of our methodology is that its works well with images that have undergone degradation of different forms like darkening or blurring over time and its language independence. Inscriptions from all over the world belonging to separate timelines in any language can be retrieved using this technique making it highly robust.

2. Methodology

2.1 Bag of Visual Words:

Fig 1: Proposed Image Retrieval system
The bag of visual words technique has evolved from the bag of words technique where we move from document classification in BOWs to image classification in BOVWs. The Bag of Words technique is used for document classification where the documents are represented based on the frequency of word occurrence. The document is considered as a stream of words where punctuation marks and blank spaces are ignored. A vocabulary is created and its size reduced by removing words that occur too often (as they might not hold much significance) or words that occur too less say once or twice. After filtering the vocabulary to the desired size we proceed further by indexing the documents. The desired document is retrieved from the indexed vocabulary using their histogram (frequency of word occurrence) by comparing the query image with the indexed files. Similar to the BoWs we use the Bag of Visual Words (BoVWs) technique for image classification. Here instead of text words we use visual words to build our vocabulary. An image of an inscription can be represented as a set of non distinctive discrete visual features, where the visual features corresponds to the vocabulary, and the image is defined as a histogram of visual word occurrence.

We set off by defining the interest points (either corners or blobs) and then determining the features of those interest points. We have used SURF (speeded up robust features) as our key feature detector. SURF uses an integer approximation of the determinant of Hessian blob detector to find its interest points. After extracting the features we segment those using K mean clustering. This is done to get distinct clusters and then eventually to derive a code book where each cluster is represented by its centroid.

2.2 K-mean Clustering:

The k mean clustering is an unsupervised learning algorithm which forms k clusters form M observations. It is a cyclic process where the cluster centroids keep on rearranging their location until they reach a point beyond which they remain in a fixed positions and the clustering process ends. First k centroids are defined and each point of the data set is grouped into clusters based on the nearest centroids. Once this is done we recalculate k new centroids as barycentre of previous clusters and the points are regrouped according to the new centroids. This process is repeated to get k fixed centroids. And these are then used in the codebook generation process. Once the clustering is done the extracted features are given labels of the closest centroid and thus the
SURF features are quantized. Figure 2. Depicts how from a database of inscription images each image is taken and its SURF extracted followed by generation of a codebook and histogram computation of the quantized SURF features.

2.3 SURF (speeded up robust features):

SURF was first introduced by Herbert Bay et al. [12] as a feature detector and descriptor that is both scale and rotation invariant. The SURF algorithm makes use of the Hessian matrix which acts by defining interest points. The Hessian matrix uses the Difference of Gaussian (DoG) which is a basic laplacian based descriptor. The SURF works in four steps, first the Hessian matrix is used to determine interest points, then the major interest point in scale space are found, next feature direction is taken into account as it should be rotation invariant and finally the feature vectors are generated. Haar transforms are used to assess the direction of the features.

2.4 Inverted File Indexing:

Inverted file indexing is introduced as it has proven to enhance the systems efficiency by several orders. The images can be presented as a catalogue of characteristics or attributes. The indexed file then behaves as a sorted list of these attributes and each attribute has a link to the image comprising it. Hence whenever an attribute is recalled all the images from the catalogue comprising of it will be retrieved. The retrieval process as depicted in figure 1 starts with inputting the query image. From our entire collection of images if we want to retrieve an inscription image we need to input a query image which will be used to compare and extract the correct output image. The initial processing on the query image is similar to those on the inscription image from the database. First the SURF features of the query are extracted. Once this is done we cluster them using the k means algorithm. From the vocabulary the closest cluster to the derived cluster of the query image is located and assigned to it. This is represented in the codebook and from the indexed files all the images that are assigned to the given cluster are brought forward of which the best match is retrieved.

3. RESULT AND DISCUSSIONS

In this section we show the result of the inscription retrieval process. Inscriptions written in several languages are taken to form the data set, where the dataset and test images are divided into three categories, namely English, Tamil and other languages. The other language includes inscription images belonging to different periods having different symbols and pictorial representations from all over the world. The results are calculated by finding the precision p, which can be defined as

\[ P = \frac{\text{true positive}}{\text{true positive} + \text{false positive}} \]  

Where true positive is means a hit and false positive a miss. Some of the query images taken from different languages have been shown in the figure 3.

3.1 Performance Statistics:

Table 1 depicts the precision performance of our algorithm where we have created a database of 72 images in English language, 56 in Tamil and 212 in other languages images in English, Tamil and others respectively. The precisions are as shown. Figure 3. depicts the output taken on several query images. Inscriptions in Tamil, English, symbolic inscriptions have been shown as an example of our methods robustness. Figure 4 shows the algorithm’s scale invariant property where the system is able to detect an image from a query image that has been rotated by
90 degrees. Figure 5, 6 and 7 show how the blurring and darkening to a large extent does not affect our retrieval process.

Table I Precision Performance

| LANGUAGE | NO. OF IMAGES | QUERY IMAGES | PRECISION |
|----------|---------------|--------------|-----------|
| English  | 72            | 20           | 0.781     |
| Tamil    | 56            | 15           | 0.752     |
| Others   | 212           | 42           | 0.734     |

Fig 3. Output for query inputs of Inscriptions
Fig 4. Scale invariant property

Fig 5. (a) and (b) depict the results when the query image is blurred and the extent of blurring acceptable to give the correct output.

Fig 6 (a) and (b) depict different degrees of blurring in the image to be retrieved

![Query image](image1.png) ![Original image](image2.png) ![Darkened image](image3.png)

| [1] | [2] |
|-----|-----|
| Query image | Original image | Darkened image |

Fig 7. Retrieval of a darkened image

4. Conclusion

We have proposed an inscription image retrieval system that is based on the Bag of Visual Words technique using SURF as the feature descriptor which has proved to be highly efficient and gives us the advantage of being scale and rotation invariant. It is also a language independent technique making it the best suited for retrieving inscription images considering the ancient inscriptions to be written in varied styles and forms. We have worked with three sets of inscriptions, i) English, ii) Tamil iii) and inscriptions in several other language. The
algorithm has proved to work well in all three cases. Usage of SURF has enhanced the scope of the algorithm and enabled us to work with varied forms of image degradation.

ACKNOWLEDGMENT

We would like to thank Arunima Sidhu for the data collection of the inscription images.

REFERENCES

[1] S. Rajakumar, Dr. V. Subbiah Bharathi, “Century Identification and Recognition of Ancient Tamil Character Recognition,” India International Journal of Computer Applications (0975 – 8887) Volume 26– No.4, July 2011.
[2] Rodríguez-Serrano, José A., and Florent Perronnin. "Handwritten word-spotting using hidden Markov models and universal vocabularies." Pattern Recognition 42(9) 2106-2116, (2009).
[3] Brown, M., and L. Rabiner, "Dynamic time warping for isolated word recognition based on ordered graph searching techniques." Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP’82. Vol. 7. IEEE, 1982.
[4] K. pramod Sankar, R. Mannmatha, C.V Jawahar, “Large scale document image retrieval by automatic word annotation” International Journal on Document Analysis and Recognition archive springer-verlag Berlin, Heidelberg volume 17 Issue 1, March 2014.
[5] Konstantinos Zagoris Ioannis Pratikakis Basilis Gatos,” Segmentation-based Historical Handwritten Word Spotting using Document-Specific Local Features “14th International Conference on Frontiers in Handwriting Recognition, 2014.
[6] Olivier Augereau, Nicholas Journet, Anne Vialard, Jean-Philippe Domenger,” Improving classification of an industrial document image database by combining visual and textual features”.
[7] Rusinol, Marcal, et al. "Browsing heterogeneous document collections by a segmentation-free word spotting method." International Conference on Document Analysis and Recognition. IEEE, 2011.
[8] Ataer, Esra, and Pinar Duygulu. "Matching ottoman words: an image retrieval approach to historical document indexing." Proceedings of the 6th ACM international conference on Image and video retrieval. ACM, 2007.
[9] Sankar, P., Jawahar, C., Mannmatha, R, “Nearest neighbor based collection OCR” In: Proceedings of the IAPR Workshop on Document Analysis Systems, pp. 207 - 214 (2010).
[10] Rothacker, Leonard, Marcal Rusinol, and Gernot A. Fink. "Bag-of-features HMMs for segmentation-free word spotting in handwritten documents." 2013 12th International Conference on Document Analysis and Recognition. IEEE, 2013.
[11] Aldavert, David, et al. "A study of Bag-of-Visual-Words representations for handwritten keyword spotting," International Journal on Document Analysis and Recognition (IJDAR) 18.3 223-234, (2015).
[12] Bay, Herbert, et al. "Speeded-up robust features (SURF)." Computer vision and image understanding 110(3): 346-359, 2008.
[13] Ravi Shekhar and C.V. Jawahar “Word Image Retrieval using Bag of Visual Words” 10th IAPR International Workshop on Document Analysis Systems, 2012.
N. Jayanthi received her M.Tech. degree in 2007 from Department of Electronics and Communication Engineering, National Institute of Technology, Tiruchirappalli, Tamilnadu. She had been working as a Assistant Professor for the past 8 years in the Department of Electronics and Communication Engineering, Delhi Technological University, Delhi. Currently she is working towards her PhD degree in Image Processing and Applications. Her areas of interest are Image Processing, Machine Learning and Communication System.

Dr. S. Indu received her PhD. degree in 2012 from Department of Electronics and Communication Engineering, Delhi University, Delhi. She is working as Professor having 18 years of teaching experience and presently holds the position of Head of the Department in the Department of Electronics and Communication Engineering, Delhi Technological University, Delhi. Her area of interest is Image Processing, Computer Vision and Wireless Sensor Networks and in these areas she has published several research papers. She also holds position of Branch Counsellor IEEE DTU Branch and worked on several DST sponsored projects.