BAG OF VISUAL WORDS FOR WORD SPOTTING IN HANDWRITTEN DOCUMENTS BASED ON CURVATURE FEATURES

Thontadari C and Prabhakar C.J

Department of Computer Science, Kuvempu University, Shimogga, India

ABSTRACT

In this paper, we present a segmentation-based word spotting method for handwritten documents using Bag of Visual Words (BoVW) framework based on curvature features. The BoVW based word spotting methods extract SIFT or SURF features at each keypoint using fixed sized window. The drawbacks of these techniques are that they are memory intensive; the window size cannot be adapted to the length of the query and requires alignment between the keypoint sets. In order to overcome the drawbacks of SIFT or SURF local features based existing methods, we proposed to extract curvature feature at each keypoint of word image in BoVW framework. The curvature feature is scalar value describes the geometrical shape of the strokes and requires less memory space to store. The proposed method is evaluated using mean Average Precision metric through experimentation conducted on popular datasets such as GW, IAM and Bentham datasets. The yielded performances confirmed that our method outperforms existing word spotting techniques.

KEYWORDS

Corner Keypoints, Curvature Feature, word image segmentation, Bag of Visual Words, Codebook, Similarity Measure, Feature Extraction, Word Spotting.

1. INTRODUCTION

Retrieving information from huge collection of historical and modern documents is useful for interpreting and understanding documents in various domains. Document digitization provides an inspiring alternative to preserve valuable historic and modern manuscripts. However, digitization solitary cannot be obliging until these collections of manuscripts can be indexed and made searchable. Spotting particular regions of interest in a digital document is easy owing to the possibility to search for words in huge sets of document images. The procedure of manual or semi-automatic transcription of the whole text of handwritten documents for searching any particular word is a tiresome and expensive job and automation is desirable in order to reduce costs. In earlier research work, character recognition is used widely in order to search the required word in a document. Available OCR engines designed for different languages yield excellent recognition results on scanned images of good quality printed documents. However, the performance of OCR engines significantly degraded when applied to handwritten documents due to faded ink, stained paper, and other adverse factors of handwritten documents. Another factor is that traditional optical character recognition (OCR) techniques that usually recognize words character by character. The performance of the available OCR engine is highly dependent on the burdensome process of learning. Moreover, the writing and font style variability, linguistics and script dependencies are the impediments of such systems. Recently, research has been emphasized on word spotting in order to overcome the drawbacks of OCR engines on handwritten documents. Word spotting is a moderately new alternative for information retrieval in digitized document images and as possible as to retrieve all the document images, paragraphs,

DOI:10.5121/ijcsit.2017.9406

and lines that contain words similar to a query word by matching the image of a given query word image with document images.

In the handwritten word spotting literature, we can classify two different families of word spotting methods depending on the representation of the handwritten words [1]. The first category of techniques called as sequential word representations [2] describes handwritten words as a time series by using a sliding window in the writing direction. The second category of techniques called as holistic word representations [3] extracts a single feature vector with fixed size that represents the word as a whole. The drawback of sequential word representations is that size of the word's descriptors will depend on the width of the word which leads two different words and cannot be directly compared by means of a distance between points. Holistic word representations are giving promising results in recent years. The main advantage is that these techniques extract fixed sized feature vectors and thus, two handwritten word images can be compared using statistical pattern recognition technique or any standard distances.

In Holistic word representations, a good description of the word images is a key issue. The researchers have developed word spotting techniques using different feature representations. The shape descriptors are widely used in word spotting and shape descriptors can be classified into statistical and structural. The statistical descriptor represents the image as an n-dimensional feature vector, whereas, the structurally based techniques represent the image as a set of geometric and topological primitives and relationships among them. Statistical descriptors are the most frequent and they can be classified as global and local features. Global features are computed from the image as a whole, for example, widths, height, aspect ratio, the number of pixels. In contrast, local features are those which refer independently to different regions of the image or primitives extracted from it. For example, position/number of holes, valleys, dots or crosses, gradient orientation based SIFT and SURF features. These features have been proved useful due to invariance to scale and rotation as well as the robustness across the considerable range of distortion, noise, and change in intensity. Hence, Scale Invariant Feature Transform (SIFT) [4] and Speeded Up Robust Feature (SURF) [5] features are extensively used in different computer vision applications such as image retrieval [6] and image classification [7] and sign board detection [8]. SIFT and SURF local feature descriptors have recently achieved a great success in the document image analysis domain. Hence, the researchers are adapting these local features for development of many applications in document analysis such as word image retrieval [9-10], logo retrieval [11], page retrieval [12].

Many authors have proposed handwritten word spotting techniques based on the matching of keypoints extracted using SIFT or SURF. The techniques have been used to directly estimate similarities between word images, or by searching the query model image within complete pages in segmentation free circumstances. However, the key points matching framework presents the disadvantage that such methods are memory intensive and requires alignment between the keypoint sets. In order to avoid matching all the keypoints among them, the Bag of Visual Words (BoVW) technique has been used for word spotting in handwritten documents. The BoVW based word spotting methods yield holistic and fixed-length image representation while keeping the discriminative power of local descriptor. Rusinol, et al. [9] have proposed a segmentation free word spotting technique using BoVW model based on SIFT descriptors. The query by example model [13] where the local patches expressed by a bag of visual words model powered by SIFT descriptors. Rothacker, et al. [14] have proposed to combine the SIFT descriptors based on BoVW representation with Hidden Markov Models in a patch-based segmentation-free framework in handwritten documents. The drawback of SIFT-based word spotting in BoVW technique is that they are memory intensive; window size cannot be adapted to the length of the query, relatively slow to compute and match. The performances of these methods are dependent on length of the query with respect to the fixed size of the window.

In order to overcome the drawbacks of SIFT based word spotting using BoVW, we proposed segmentation-based word spotting technique using a Bag of Visual Words powered by curvature features which describe spatial information of the local keypoint. The main contribution of our approach is that detection and extraction of curvature feature from word image. The literature survey on theories of vision reveals that curvature is important in shape perception and measures of curvature plays an important role in shape analysis algorithms. High curvature points are the best place to break the lines where a maximal amount of information can be extracted which are necessary for successful shape recognition. This is based on the information that the corners are points of high curvature. Asada, et al. [15] proposed an approach for representing the significant changes in curvature along the bounding of planar shape and this representation called as curvature primal sketch. The scale space approach [16] to the description of planar shapes using the shape boundary. The curvature along the contour was computed and the scale space image of the curvature function was used as a hierarchical shape descriptor that is invariant to translation, scale, and rotation. Recognition of handwritten numerals based on gradient and curvature features of the gray scale character proposed in [17]. Recently, in [18] proposed an approach for offline Malayalam recognition using gradient and curvature feature. An advantage of using curvature feature is that it reduces the dimension of the feature descriptor when compared to other local features. Hence, in this paper, for the purpose of shape description of handwritten word, curvature features at corner keypoints are used. Then, we constructed a BoVW framework based on curvature features. After construction of Bag of Visual Words of the training set and query image, Nearest Neighbor Search (NNS) similarity measure algorithm is used to retrieve word images similar to a query image. The remainder of this paper is organized as follows: in section 2, we present a brief review on existing word spotting techniques. This is followed by proposed handwritten word spotting. Section 4 provides experimental results of proposed approach and finally, the conclusion is given.

2. RELATED WORK

Word spotting was initially proposed by Jones, et al. [19] in the field of speech processing, while later this concept was agreed by several researchers in the field of printed and handwritten documents for the purpose of spotting and indexing. This approach enables to localize a user preferred word in a document without any syntactic restriction and without an explicit text recognition or training phase. The concept of word spotting [20] has been introduced as an alternative to OCR based results. The word spotting methods have followed a well-defined process. Initially, layout analysis is carried out to segment the words from document images. Then, the segmented word images are symbolized as sequences of features such as geometric features [21], profile based features [22-23], local gradient features [24]. Finally, similarity measure methods, such as XOR comparison, Euclidean distance, Scott and Longuet Higgins distance, Hausdorff distance of connected components, sum of Euclidean distances of corresponding key points. More, recently Dynamic Time Warping and Hidden Markov Models are used to compare the feature of query word image and set of word images presented in the dataset. Finally, retrieved word images are ranked according to this similarity.

The state of the art word spotting methods can be categorized into two groups, segmentation based approach, and segmentation free approach. In segmentation based approach, the sequences of operations are applied to the document images. First, the document is pre-processed and based on text layout analysis; the document is segmented into word images. Then, analyzing and extracting feature descriptor from the segmented word image. Based on these feature descriptors, a distance measure is used to measure the similarity between the query word image and the segmented word image. Rath et al. [22] proposed an approach which involves grouping similar word images into clusters of words by using both K-means and agglomerative clustering techniques. They constructed an index that links words to the locations of occurrence which helps to spot the words easily. The authors in [25] have proposed word spotting technique based on

gradient based binary features can offer higher accuracy and much faster speed than DTW matching of profile features. Rodriguez, et al. [24] proposes a method that relaxes the segmentation problem by requiring only segmentation at the text line level.. Unsupervised writer adaptation for unlabelled data has been successfully used for word spotting method proposed in [27] based on statistically adaption of initial universal codebook to each document. Khurshid et al. [29] represent a word image as a sequence of sub-patterns, each sub-pattern as a sequence of feature vectors, and calculate a segmentation-driven edit (SDE) distance between words. The authors in [30] proposed a model-based similarity between vector sequences of handwritten word images with semi-continuous Gaussian mixture HMMs. Coherent learning segmentation based Arabic handwritten word spotting system [31] in which can adapt to the nature of Arabic handwriting and the system recognizes Pieces of Arabic Words (PAWs). Based on inkball character models, Howe [28] proposed a word spotting method using synthetic models composed of individual characters.

On the other hand, Gatos, et al. [32] presented a segmentation free approach for word spotting. In this method, words are not individually segmented from a document, instead significant regions are detected, document image feature descriptors are allotted to detected region and matching procedure is performed only on the detected regions of interest. A segmentation free word spotting [33] method based on zones of interest. Initially, they detect the zones of interest and compute the gradient angles, and construct a graph of zones of interest. Finally, matching is performed by searching through the tree. A segmentation free approach [34] based on flexible ink ball model, allowing for Gaussian random-walk deformation of the ink trace. Thus, the query word model is warped in order to match the candidate regions in the document. The segmentation free method for word spotting in handwritten documents [35] based on heat kernel signatures (HKS) descriptors are extracted from a local patch cantered at each keypoint. A statistical script independent line based word spotting method [36] for offline handwritten documents based on hidden Markov models by comparing a filler models and background models for the representation of background and non-keyword text. An unsupervised Exemplar SVM framework for segmentation free word spotting method[37] using a grid of HOG descriptors for documents representation. Then, sliding window is used to locate the document regions that are most similar to the query. A segmentation free query by string word spotting method [38] based on a Pyramidal Histogram of Characters (PHOC) are learned using linear SVMs along with the PHOC labels of the corresponding strings. Line-level keyword spotting method [39] on the basis of frame-level word posterior probabilities of a full-fledged handwritten text recognizer based on hidden Markov models and N-gram language models.

3. OUR APPROACH

3.1. Bag of Visual Words (BoVW) Framework

The Bag of Visual Words [40] is an extension of Bag of Words [41] to the container of digitized document images. The BoVW framework consists of three main steps: in the first step, a certain number of image local interest points are extracted from the image. These keypoints are significant image points having rich of information content. In the second step, feature descriptors are extracted from these keypoints, and these feature descriptors are clustered. Each cluster corresponds to a visual word that is a description of the features shared by the descriptors belongs to that cluster. The cluster set can be interpreted as a visual word vocabulary. Finally, each word image is represented by a vector, which contains occurrences of each visual word that appears in the image. Based on these feature vectors, a similarity measure is used to measure the likeness between given query image and the set of images in the dataset.

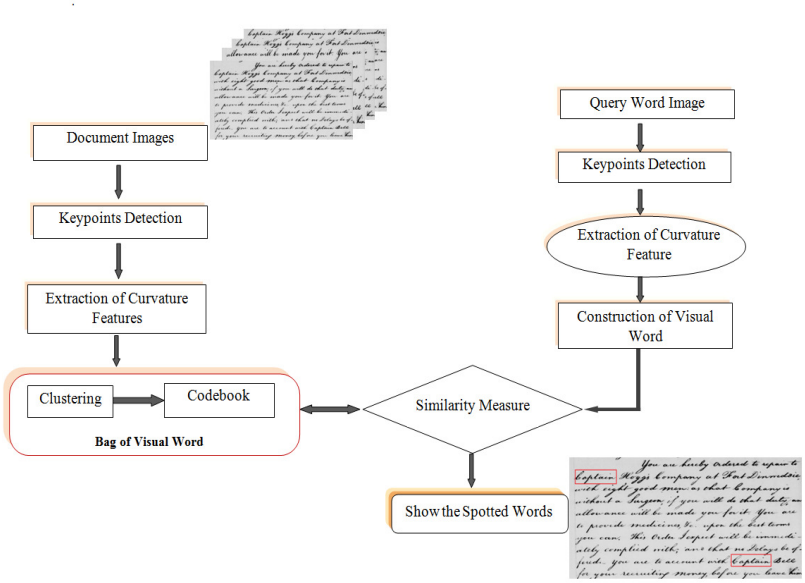


Figure 1. Block diagram of the our approach for word spotting

To the best of the author's knowledge, there is no approach using curvature feature in the BoVW framework for word spotting in handwritten document images. The proposed method based on the BoVW framework for word spotting is illustrated in Figure 1. The proposed method composed of four stages: (i) keypoints detection (ii) extraction of curvature features (iii) codebook generation and (iv) word retrieval. In the first stage, the document image is segmented into text lines and each text lines are segmented into primitive segments i.e. words using directional local profile technique [42]. The corner keypoints are extracted from word image using Harris-Laplace Corner detector. Then, curvature features are extracted at each corner keypoint. The codebook is used to quantize the visual words by clustering the curvature features using K-Means algorithm. Finally, each word image is represented by a vector that contains the frequency of visual words appeared in the image. In the word retrieval phase, for a given query image, we construct the visual word vector. Then, Nearest Neighbor Search is used to match the visual word vector of the query word image and the visual word vectors presented in the codebook. Finally, based on the ranking list, retrieved word images are presented.

3.2. Keypoints Detection

Before extracting the keypoints from word images, the document must be segmented into words. We segment document images into lines and then into words using robust algorithm [42]. The sample results of segmented word images are shown in Figure 2.

Figure. 2. Segmented word images (a) from GW database (b) from IAM database (c) from Bentham database

Once the document image is segmented into words, next step is the detection of corner points (keypoints). In this work, we detect the corner points of the word images using Harris-Laplace corner detector [43].

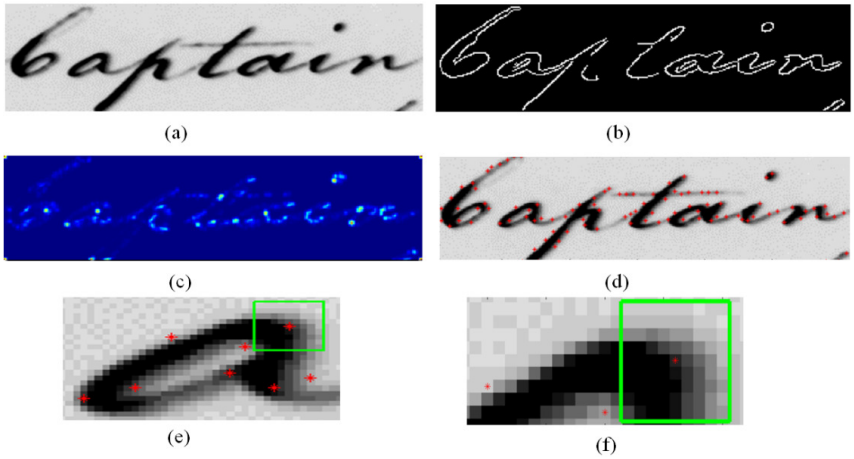


Figure. 3. The intermediate results for keypoints detection: (a) segmented word image (b) canny edge map (c) Detected corner keypoints on canny edge map image using Harris-Laplace operator (d) detected corner keypoints on original word image (e) keypoint at the global level (e) sample keypoint at the local level

Harris-Laplace corner detector is an accepted interest point detector due to its strong invariance to scale, rotation, and image noise. It is believed that most of the information on a contour is concentrated at its corner points. These corner keypoints, which have high curvature on the contour, play very important role in shape analysis of handwritten word images. The Figure 3 (d-f) shows the result for detection of corner points in the word image.

3.3. Curvature Features

We extract curvature features from each detected corner keypoints of word image. The curvature is a measure of how geometric object surface deviates from being a flat plane, or a curve is turning as it is traversed. The curvature features are the salient points along the contour of word image. At a particular point (M) along the curve, a tangent line can be drawn; this line making an angle θ with the positive x-axis (Figure 4). The curvature at point is defined as the magnitude of the rate of change of θ with respect to the measure of length on the curve. Mathematically, the curvature of a point M in the curve C is defined as follows:

$$K(M) = \lim_{\Delta S \to 0} \left| \frac{\Delta \theta(M)}{\Delta S} \right| \qquad (1)$$

where, $\theta(M)$ is the tangential angle of the point M and S is the arc length.

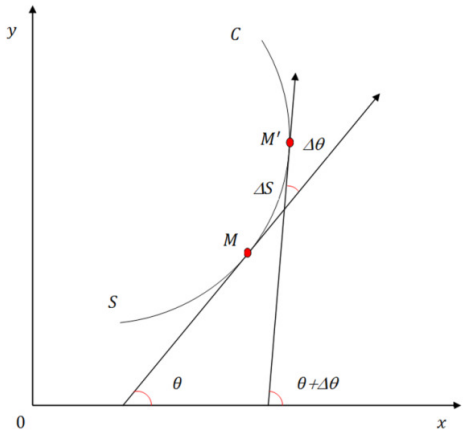


Figure.4. Curvature of the curve C

3.4. Codebook Generation

The single codebook is the house of all possible visual words that correspond to spotting the word image in handwritten document images. Generally, the codebook must hold the following constraints, the codebook should be small, that guarantees to minimum computational rate through minimum dimensionality, redundant visual words minimization, and provide high discrimination performance.

In order to generate the codebook for given training samples, we extract curvature features at each detected corner keypoint and it is followed by clustering of curvature features using K-means algorithm. The curvature feature computed from detected corner keypoint is allotted to a cluster through minimum distance from the centre of the corresponding cluster. The clusters are treated as visual words. The number of clusters characterizes the size of the codebook. Finally, word image is formally represented as visual word vector with the help of frequency of occurrences of the visual words in the codebook. For example, consider a codebook with size 7 visual words and a word image that contains 16 local curvature features, which are assigned as follows: 3 local curvature features for the first visual word, 5 local curvature features for the second, 2 local curvature features for the third, 3 local curvature features for the fourth visual word, 2 local curvature features for the fifth visual word and 1 local curvature feature for the seventh visual word. Then, the visual word vector which represents the Bag of Visual Words of the word image is [3, 5, 2, 3, 2, 0, 1]. The dimension of this visual word vector is equal to the number of visual words in the codebook.

3.5. Word Retrieval

In order to retrieve the word images similar to query word, we compute the similarity between visual word vector of the query word image and visual word vector of word images present in the dataset. Lowe [49] proposed to obtain matches through a Nearest Neighbour Search (NNS) approach, by using the Euclidean distance between a training samples and query image descriptor by considering appropriate threshold T.

$$NNS = \sqrt{\sum_{i=1}^{n} \left(d_j(i) - q(i)\right)^2} \quad < T \qquad (2)$$

where, n is the dimension of visual word vector, d_j and q are visual word vectors of j^{th} training sample and query word image respectively.

4. EXPERIMENTAL RESULTS

In this Section, we present the experimental results of the proposed word spotting method in comparison with the state of the art word spotting methods. The proposed method is evaluated on three handwritten datasets of different nature, such as George Washington (GW) dataset [23], IAM English dataset [47] and Bentham dataset [44]. Figure 5 shows sample document images of GW, IAM, and Bentham dataset.

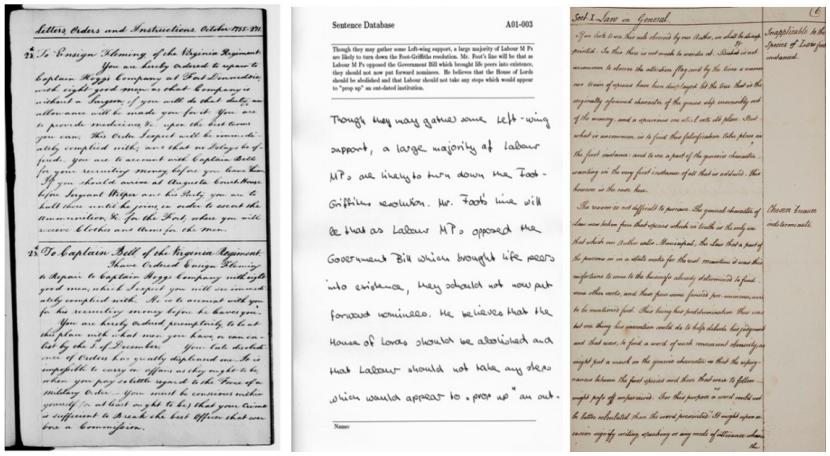


Figure. 5. Samples handwritten document images from (a) GW dataset (b) IAM dataset (c) Bentham dataset

In order to evaluate the performance of our approach, we used mean Average Precision (mAP) metric. Precision gives the percentage of true positives as compared to the total number of word images retrieved by the approach. The mAP provides a single value measure of precision for all the query word images. The Precision for each query word is computed as follows:

$$\Pr ecision(P) = \frac{TP}{TP + FP} \tag{3}$$

where, True Positive (TP) is the number of relevant word instances correctly retrieved, False Positive (FP) is the number of word images retrieved as other than query image.

We employed five-fold cross-validation method to validate our approach. The dataset is partitioned into five complementary subsets i.e. four subsets are used for training and remaining one subset is used as validation (test) set. The cross-validation process is repeated five times, with each subset used exactly once for validation. To estimate the overall evaluation results of our approach, we compute the average of the five validation results.

4.1. Selection of Codebook Size

The size of the codebook is pre-computed and selection of the optimal size of a codebook is one of the important factors in achieving highest accuracy. Predicting the desirable clusters and optimal codebook size is non-straightforward and it is dataset-dependent. Hence, in order to find an optimal size, we conducted the experiments using GW, IAM and Bentham datasets by varying

codebook size. For each dataset, we employed five-fold cross validation process. To estimate the overall evaluation results of our approach, we compute the average of the five validation results. The mean Average Precision obtained on three datasets using our approach for a different codebook size is presented in Table 1 and its pictorial representation is shown in the Figure 6.

Table 1. Performance evaluation of our approach for varying codebook size using three datasets.

Codebook size	(mAP %)		
	For GW dataset	For IAM dataset	For Bentham dataset
36	63.43	42.06	40.17
40	65.78	43.89	44.91
45	68.43	48.72	51.36
60	87.65	63.12	71.59
72	**96.72**	79.64	80.46
90	93.42	81.12	84.72
120	90.01	**94.46**	88.21
180	87.57	91.13	90.34
360	81.32	89.24	**91.84**
720	70.13	71.64	86.42

From the Table 1 and Figure 6, it is observed that, when codebook size is large the accuracy of our approach decreases. It is concluded that for GW dataset, the performance of our approach is significantly better when a size of the codebook is set to 72. Similarly, for IAM dataset and Bentham dataset, the performance of our approach yields good accuracy when a size of the codebook is set to 120 and 360 respectively. Hence, in all the experiments, the codebook size is 72, 120 and 360 for GW, IAM and Bentham dataset respectively.

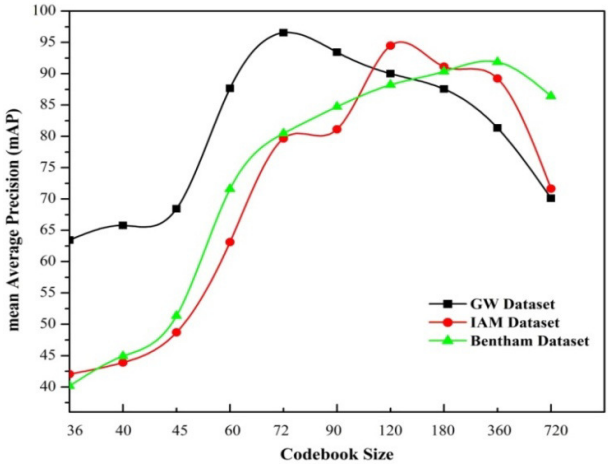


Figure 6. The performance comparison of our approach for varying codebook size on three different datasets based on mAP.

4.2. Experiments on GW Dataset

The GW dataset contains 20 handwritten pages with a total of available 4860 words written by several Washington's secretaries. From this dataset, we have taken correctly segmented 1680 word instances of 42 different classes of words from all the 20 pages of scanned handwritten documents. Each class of words may occur at least two times per document. While performing

five-fold cross-validation, we partitioned the dataset into five disjoint subsets and each subset consists of 336 word images (from each class of words we have considered 8 word instances).

In the training stage, 1344 word instances are used and remaining 336 word images are used for testing. Hence, for every experiment, we have 336 word images as test samples. The cross-validation process is repeated five times, with each subset used exactly once for validation. Table 2 shows the qualitative results of our approach for the GW dataset. The first column shows query words and corresponding retrieved word instances of a given query word are shown in subsequent columns. To estimate the overall quantitative evaluation results of our approach, we have taken the mean Average Precision. It is noticed that our approach achieves the highest accuracy with mAP 96.72. The result confirms that the robustness of our approach for different kinds of words with a small variation in writing style.

Table 2. Query word images (first column) and corresponding retrieved word instances from GW dataset.

We compared the performance of our approach with other existing word spotting methods using GW dataset. The work of Rath, et al. [22] is considered as a baseline of our experiment. A comparison with other word spotting methods, such as Bag of Visual Words based methods generated by dense SIFT feature descriptor to detect the query keyword proposed in two different papers by [9] and [14]. The fourth method used for comparison is unsupervised word spotting [45] using a grid of HOG descriptors by sliding window framework. We compared the performance of proposed method with our previous work for word spotting based on Co-HOG feature descriptor extracted in scale space representation [46]. The results of existing methods are extracted from their papers where results are reported as shown in Table 3. For our approach, fivefold cross-validation method is used to validate based on mean Average Precision. In Table 3, the size of dataset, feature descriptors used and comparison results of our approach with existing methods are shown. It is observed that, compared to existing methods, our approach yields the highest accuracy of 96.72.

Among existing methods, [9] and [14] are based on BoVW concept and [22], [45] and [46] are non-BoVW methods. Hence we extended these non-BoVW methods to BoVW framework and evaluation results are obtained. Table 3 shows the performance comparison of our approach with results of extended existing methods to BoVW framework. It is concluded that our approach yields highest retrieval result compared to BoVW extended existing methods.

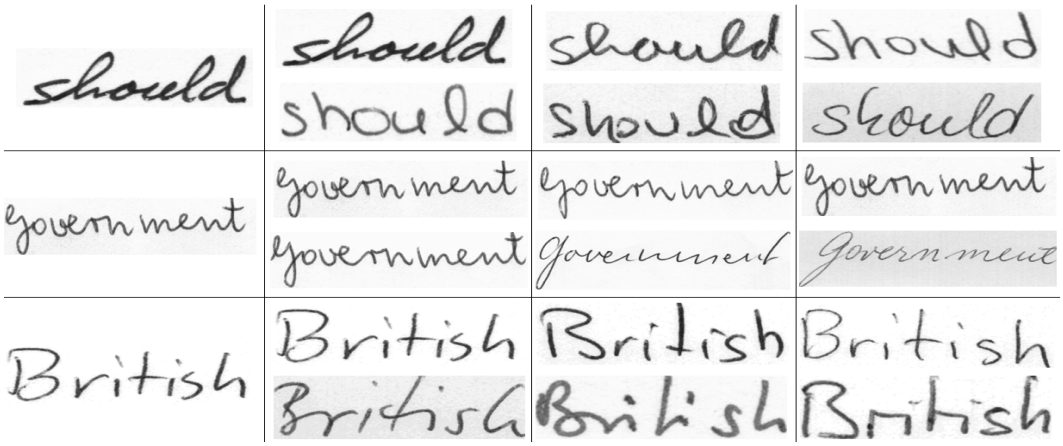
Table 3. The performance comparison of our approach with existing methods for GW dataset

Methods	Features	Experimental setup	mAP (%)	
			Originally reported	BoVW Framework
[22]	Projection profile, and Background/Ink transitions	10 pages, 2381 queries	40.90	77.20
[9]	SIFT	20 pages, 4860 queries	30.40	30.40
[45]	HOG	20 pages, 4856 queries	54.40	54.40
[14]	SIFT	20 pages, 4860 queries	61.10	61.10
[46]	Co-HOG	20 pages, 1680 queries	98.76	80.90
Our approach	Curvature	20 pages, 1680 queries	--	96.72

4.3. Experiments on IAM Dataset

A Modern English handwritten IAM dataset consists of 1539 pages text from the Lancaster-Oslo /Bergen corpus [26]. The dataset has been written by 657 writers. From the IAM dataset, we have taken 2070 segmented non-stop word images of 46 different classes i.e., each class of word image has 45 instances of different writing styles. For five-fold cross-validation process, each subset consists of 414 word images (from each class of words we have taken 9 word images). In the training stage, 1656 word images are used and remaining 414 word images are used for testing. Table 4 shows the experimental results of our approach for the IAM dataset. It is observed that our approach achieves the highest accuracy with mAP is 94.46.

Table 4. Query word image (first column) and corresponding retrieved word instances from IAM dataset.



We compared the performance of our approach with two existing word spotting methods: the first method is SIFT descriptors based method [37]. SIFT features are densely extracted at different patch size and then aggregated into Fisher vector representation of the word image and second method is Co-HOG feature descriptor based word spotting proposed in our previous work [46]. The results of existing methods are extracted from their papers where results are reported as shown in Table 5. For our approach, fivefold cross-validation method is used to validate based on mean Average Precision. Table 5 shows the comparison results of our approach with existing methods and extended existing methods to a BoVW framework. It is observed that our approach

yields the highest accuracy of 94.46 compared to non-BoVW methods as well as extended existing methods to a BoVW framework.

Table 5. The performance comparison of our approach with existing word spotting methods for IAM dataset

Methods	Features	Experimental setup	mAP (%)	
			Originally reported	BoVW Framework
[37]	SIFT	1539 pages, 5784 queries	54.78	72.16
[46]	Co-HOG	1539 pages, 1000 queries	96.58	82.64
Our approach	Curvature	1539 pages, 2070 queries	--	94.46

4.4. Experiments on Bentham Dataset

The Bentham dataset consists of 50 high qualities handwritten documents written by Jeremy Bentham [44] as well as fair copies written by Bentham's secretarial staff. From this dataset, we have taken correctly segmented 1200 word instances of 12 different classes of words from all the 50 pages of scanned handwritten documents. Each class of words may occur at least two times per document. While performing five-fold cross-validation, we partitioned dataset into five disjoint subsets and each subset consists of 240 word images (from each class of words we have consider 20 word instances). In the training stage, 960 word instances are used and remaining 240 word images are used for testing. Table 6 shows the qualitative results of our approach for the Bentham dataset. It is noticed that our approach achieves the highest accuracy with mAP 91.84. The result confirms that the robustness of our approach for a different variation of the same word involves, different writing style, font size, noise as well as their combination.

Table 6. Query word image (first column) and corresponding retrieved word instances from Bentham dataset.

We compared the performance of our approach with existing word spotting method [48] using texture features extracted around the selected keypoints. Table 7 shows the comparison results of our approach with the extended existing method to a BoVW framework. It is observed that, compared to the extended existing method, our approach achieves promising results with mAP 91.84.

Table 7. The performance comparison of our approach with existing word spotting methods for Bentham dataset

Methods	Features	Experimental setup	mAP (%)	
			Originally reported	BoVW Framework
[48]	Texture	50 pages, 3668 queries	68.01	74.36
Our approach	Curvature	50 pages, 1200 queries	--	91.84

Based on the experimental results, we can conclude that our approach efficiently retrieves the handwritten words which are having non-uniform illumination, suffering from the noise and written by different writers. The highest accuracy of our approach is due to the extraction of curvature at corner keypoint describes the geometrical shape of strokes present in handwritten words. The construction of BoVW using curvature features is simple when compared to SIFT or SURF features because it has scalar value. It is identified through evaluation of experimental results on GW, IAM and Bentham dataset, our approach outperforms existing SIFT or SURF based word spotting methods because the curvature is robust with respect to noise, scale, orientation and it preserves the local spatial information of the word shape. The advantage of our approach is that it uses less memory space because of scalar value extracted at every keypoint when compared to SIFT or SURF descriptors where feature vector is extracted at every keypoint. Another advantage involved in our approach is that the codebook size is very small compared to a size of codebook generated using SIFT or SURF features. A benefit of using BoVW representation is that, word retrieval can be effectively carried out because word image is retrieved by computing the histogram of visual word frequencies, and returning the word image, with the closest histogram. Therefore, word images can be retrieved with no delay.

4. CONCLUSION

In this paper, we proposed a novel approach for word spotting in handwritten documents using curvature features in Bag of Visual Words framework. The curvature feature is significant in word shape perception and preserves the advantage of holistic representation of word image. From the experimental results, it is observed that curvature features are more suitable for handwritten word representation and which can improve the performance of proposed word spotting method as compared to existing non BoVW as well as BoVW framework based word spotting methods. The use of BoVW framework has gained attention as a way to represent segmented handwritten words and can lead to a great boost in performance of our approach. BoVW representation can retrieve word instances efficiently, which is difficult with the popular existing methods. It is proven through experimental results evaluated using three datasets such IAM, GW, and Bentham. This motivates the development of proposed approach for word spotting in large collection of handwritten documents to retrieve words and its instances similar to query words.

REFERENCES

[1] Llados, J., Rusinol, M., Fornes, A., Fernandez, D., &Dutta, A. (2012). On the influence of word representations for handwritten word spotting in historical documents. *International Journal of Pattern Recognition and Artificial Intelligence* 26(5), 1263,002.1–1263,002.25.

[2] Plamondon, R., & Srihari, S. (2000). Online and off-line handwriting recognition: a comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(1), pp.63–84.

[3] Madhvanath, S., & Govindaraju, V.(2001). The role of holistic paradigms in handwritten word recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(2), pp.149-164.

[4] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), pp.91-110.

[5] Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-up robust features (SURF). *Computer vision and image understanding*, *110*(3), pp.346-359.

[6] Sivic, J., & Zisserman, A. (2003). Video google: A text retrieval approach to object matching in videos. In *iccv* Vol. 2, No. 1470, pp.1470-1477.

[7] Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., & Gong, Y. (2010). Locality-constrained linear coding for image classification. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference,* pp. 3360-3367.

[8] Schroth, G., Hilsenbeck, S., Huitl, R., Schweiger, F., & Steinbach, E. (2011). Exploiting text-related features for content-based image retrieval. *In Multimedia (ISM)*, IEEE International Symposium, pp.77-84.

[9] Rusinol, M., Aldavert, D., Toledo, R., & Llados, J. (2011). Browsing heterogeneous document collections by a segmentation-free word spotting method. In *Document Analysis and Recognition (ICDAR), International Conference,* IEEE, pp.63-67.

[10] Yalniz, I. Z., & Manmatha, R. (2012). An efficient framework for searching text in noisy document images. In *Document Analysis Systems (DAS), 10th IAPR International Workshop,* IEEE pp.48-52.

[11] Jain, R., & Doermann, D. (2012). Logo retrieval in document images. In *Document analysis systems (das), 10th iapr international workshop on* IEEE, pp.135-139.

[12] Smith, D. J., & Harvey, R. W. (2011). Document Retrieval Using SIFT Image Features. *J. UCS*, *17*(1), pp.3-15.

[13] Shekhar, R., & Jawahar, C. V. (2012). Word image retrieval using bag of visual words. In *Document Analysis Systems (DAS), 2012 10th IAPR International Workshop,* IEEE, pp.297-301.

[14] Rothacker, L., Rusinol, M., & Fink, G. A. (2013). Bag-of-features HMMs for segmentation-free word spotting in handwritten documents. In Document Analysis and Recognition (ICDAR), 12th International Conference IEEE, pp.1305-1309.

[15] Asada, H., & Brady, M. (1986). The curvature primal sketch. *IEEE transactions on pattern analysis and machine intelligence*, (1), pp. 2-14.

[16] Mokhtarian, F., & Mackworth, A. K. (1992). A theory of multiscale, curvature-based shape representation for planar curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *14*(8), pp. 789-805.

[17] Shi, M., Fujisawa, Y., Wakabayashi, T., & Kimura, F. (2002). Handwritten numeral recognition using gradient and curvature of gray scale image. *Pattern Recognition*, *35*(10), pp.2051-2059.

[18] Kannan, B., Jomy, J., & Pramod, K. V. (2013). A system for offline recognition of handwritten characters in Malayalam script.

[19] Jones, G. J., Foote, J. T., Jones, K. S., & Young, S. J. (1995). Video mail retrieval: The effect of word spotting accuracy on precision. *In Acoustics, Speech, and Signal Processing*, International Conference on IEEE. Vol. 1, pp.309-312.

[20] Rath, T. M., & Manmatha, R. (2003). Word image matching using dynamic time warping. *In Computer Vision and Pattern Recognition, Proceedings*. IEEE Computer Society Conference on Vol. 2, pp.II-II.

[21] Marti, U. V., & Bunke, H. (2001). Using a statistical language model to improve the performance of an HMM-based cursive handwriting recognition system. *International journal of Pattern Recognition and Artificial intelligence*, *15*(01), pp.65-90.

[22] Rath, T. M., & Manmatha, R. (2003). Features for word spotting in historical manuscripts. In *Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference,* IEEE, pp.218-222.

[23] Lavrenko, V., Rath, T. M., & Manmatha, R. (2004). Holistic word recognition for handwritten historical documents. In *Document Image Analysis for Libraries, 2004. Proceedings. First International Workshop,* IEEE. pp. 278-287.

[24] Rodriguez, J. A., & Perronnin, F. (2008). Local gradient histogram features for word spotting in unconstrained handwritten documents. *Proc. 1st ICFHR*, pp.7-12.

[25] Zhang, B., Srihari, S. N., & Huang, C. (2004). Word image retrieval using binary features. In *Electronic Imaging,* International Society for Optics and Photonics. pp. 45-53.

[26] Johansson, S., Leech, G., & Goodluck, H. (1978). Manual of Information to Accompany the Lancaster-Olso/Bergen Corpus of British English, for Use with Digital Computers.

[27] Rodríguez-Serrano, J. A., Perronnin, F., Sánchez, G., & Lladós, J. (2010). Unsupervised writer adaptation of whole-word HMMs with application to word-spotting. *Pattern Recognition Letters*, *31*(8), pp.742-749.

[28] Howe, N. R. (2015). Inkball models for character localization and out-of-vocabulary word spotting. In *Document Analysis and Recognition (ICDAR), 13th International Conference on* IEEE pp.381-385.

[29] Khurshid, K., Faure, C., & Vincent, N. (2012). Word spotting in historical printed documents using shape and sequence comparisons. *Pattern Recognition*, *45*(7), pp. 2598-2609.

[30] Rodríguez-Serrano, J. A., & Perronnin, F. (2012). A model-based sequence similarity with application to handwritten word spotting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(11), pp.2108-2120.

[31] Khayyat, M., Lam, L., & Suen, C. Y. (2014). Learning-based word spotting system for Arabic handwritten documents. *Pattern Recognition*, *47*(3), pp.1021-1030.

[32] Gatos, B., & Pratikakis, I. (2009). Segmentation-free word spotting in historical printed documents. In *Document Analysis and Recognition, (ICDAR), 10th International Conference,* IEEE. pp.271-275.

[33] Leydier, Y., Ouji, A., LeBourgeois, F., & Emptoz, H. (2009). Towards an omnilingual word retrieval system for ancient manuscripts. *Pattern Recognition*, 42(9), pp. 2089-2105.

[34] Howe, N. R. (2013). Part-structured inkball models for one-shot handwritten word spotting. In *Document Analysis and Recognition (ICDAR), 12th International Conference on* IEEE pp.582-586.

[35] Zhang, X., & Tan, C. L. (2013). Segmentation-free keyword spotting for handwritten documents based on heat kernel signature. In *Document Analysis and Recognition (ICDAR),12th International Conference on* IEEE, pp. 827-831.

[36] Wshah, S., Kumar, G., & Govindaraju, V. (2014). Statistical script independent word spotting in offline handwritten documents. *Pattern Recognition*, *47*(3), pp.1039-1050.

[37] Almazán, J., Gordo, A., Fornés, A., & Valveny, E. (2014). Word spotting and recognition with embedded attributes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *36*(12), pp. 2552-2566.

[38] Ghosh, S. K., & Valveny, E. (2015). Query by string word spotting based on character bi-gram indexing. In *Document Analysis and Recognition (ICDAR), 13th International Conference on* IEEE pp.881-885.

[39] Toselli, A. H., Vidal, E., Romero, V., & Frinken, V. (2016). HMM word graph based keyword spotting in handwritten document images. *Information Sciences*, *370*, pp.497-518.

[40] Fei-Fei, L., Fergus, R., & Perona, P. (2007). Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Computer vision and Image understanding*, *106*(1), pp. 59-70.

[41] Csurka, G., Dance, C., Fan, L., Willamowski, J., & Bray, C. (2004, May). Visual categorization with bags of keypoints. In *Workshop on statistical learning in computer vision, ECCV* (Vol. 1, No. 1-22, pp. 1-2.

[42] Shi, Z., Setlur, S., Govindaraju, V. (2009). A steerable directional local profile technique for extraction of handwritten arabic text lines. *In: ICDAR*. pp.176–180.

[43] Harris, C., & Stephens, M. (1988). A combined corner and edge detector. In *Alvey vision conference,* Vol. 15, No. 50, pp.10-5244.

[44] Long, D. G. (1981). The Manuscripts of Jeremy Bentham a Chronological Index to the Collection in the Library of University College, London: Based on the Catalogue by A. Taylor Milne.

[45] Almazán, J., Gordo, A., Fornés, A., & Valveny, E. (2012). Efficient Exemplar Word Spotting. *In Bmvc,* Chicago, Vol. 1, No. 2, pp. 3.

[46] Thontadari, C., & Prabhakar, C. J. (2016). Scale Space Co-Occurrence HOG Features for Word Spotting in Handwritten Document Images. *International Journal of Computer Vision and Image Processing (IJCVIP)*, *6*(2), pp.71-86.

[47] Marti, U. V., & Bunke, H. (2002). The IAM-database: An English sentence database for offline handwriting recognition. *International Journal on Document Analysis and Recognition*, 5(1), pp. 39–46.

[48] Zagoris, K., Pratikakis, I., & Gatos, B. (2014). Segmentation-based historical handwritten word spotting using document-specific local features. In *Frontiers in Handwriting Recognition (ICFHR), 14th International Conference* IEEE, pp.9-14.

[49] Lowe, D. G. (1999). Object recognition from local scale-invariant features. In Computer vision, The proceedings of the seventh IEEE international conference, IEEE, Vol. 2, pp. 1150-1157

Authors

Thontadari C. Received his M.Sc.degree in computer Science from K uvempu University, Karnataka, India in 2010, He is currently pursuing his Ph.D. in Kuvempu University, Karnataka, India. His research interests are document image processing, Computer Vision and Machine Vision

Prabhakar C.J. Received his Ph. D. degree in Computer Science and Technology from Gulbarga University, Gulbarga, Karnataka, India in 2009. He is currently workin g as Assistant Professor in the department of Computer Science and M.C.A, Kuvempu University, Karnataka, India. His research interests are pattern recognition, computer vision and machine vision and video processing