Local Binary Pattern for Word Spotting in Handwritten Historical Document

Sounak Dey[1(✉)], Anguelos Nicolaou[1], Josep Llados[1], and Umapada Pal[2]

[1] Computer Vision Center, Universitat Autonoma de Barcelona, Bellaterra, Spain
{sdey,anguelos,josep}@cvc.uab.es
[2] CVPR Unit, Indian Statistical Institute, Kolkata, India
umapada@isical.ac.in

Abstract. Digital libraries store images which can be highly degraded and to index this kind of images we resort to word spotting as our information retrieval system. Information retrieval for handwritten document images is more challenging due to the difficulties in complex layout analysis, large variations of writing styles, and degradation or low quality of historical manuscripts. This paper presents a simple innovative learning-free method for word spotting from large scale historical documents combining Local Binary Pattern (LBP) and spatial sampling. This method offers three advantages: firstly, it operates in completely learning free paradigm which is very different from unsupervised learning methods, secondly, the computational time is significantly low because of the LBP features, which are very fast to compute, and thirdly, the method can be used in scenarios where annotations are not available. Finally, we compare the results of our proposed retrieval method with other methods in the literature and we obtain the best results in the learning free paradigm.

Keywords: Local binary patterns · Spatial sampling · Learning-free · Word spotting · Handwritten historical document analysis · Large-scale data

1 Introduction

A lot of initiatives has been taken to convert the paper scriptures to digitized media for preservation in digital libraries. Digital libraries store different types of scanned images of documents such as historical manuscripts, documents, obituary and handwritten notes. The challenges in this area become diverse as more and more types of images are considered as input for archival and retrieval.

Documents of different languages are also been archived which is an another challenge. Traditional Optical Character Reader (OCR) techniques cannot be applied generally to all types of imagery due to several reasons. In this context, it is advantageous to explore techniques for direct characterization and manipulation of image features in order to retrieve document images containing textual and other non-textual components. A document image retrieval system

A. Robles-Kelly et al. (Eds.): S+SSPR 2016, LNCS 10029, pp. 574–583, 2016.
DOI: 10.1007/978-3-319-49055-7_51

asks whether an imaged document contains particular words, which are of interest to the user, ignoring other unrelated words. This is also known as **keyword spotting** or simply 'word-spotting' with no need for correct and complete character recognition. Word spotting technique in terms of pattern recognition can be defined as classification of word images.

The problem of word spotting, especially in the setting of large-scale datasets, is the balancing engineering trade-offs between number of documents indexed, queries per second, update rate, query latency, information kept about each document and retrieval algorithm. In order to handle such large scale data, computational efficiency and dimensionality are critical aspects which are effectively taken care by the use of LBP in word spotting.

Word spotting is fundamentally based on appearance based features. In this work we would like to explore the textural features as an alternative representation offering a richest description with minimal computational cost. Moreover the nature of the handwritten words suggests that there is a stable structural pattern due to the ascender and descenders in the words. In this paper, our aim is to propose an end-to-end method which can improve the performance for word spotting in handwritten historical document images. The specific objectives are: (1) To develop a word spotting method for large scale un-anotated handwritten historical data. (2) Apply texture feature like LBP to capture the fine grained information about the handwritten words which is computationally cheap. Converting the text to meta-information. (3) Combine the spatial knowledge using a Quad tree spatial structure [19] for pooling.

We use LBP as a generic low level texture classification features, that don't incorporate any assumptions specific to a task. Out here we consider every textblock as a bi-modal oriented texture.

2 State of the Art

2.1 Taxonomy of Methods

The state of the art word spotting techniques can be classified based on various criteria: (1) Depending on whether segmentation is needed or not i.e. segmentation-free or segmentation-based. (2) Based on possibility on learning: learning-free or learning-based, supervised/unsupervised. (3) Based on usability: Query-By-Example (QBE) or Query-By-String (QBS).

Methods on Segmentation-Free or Segmentation-Base: In the segmentation-based approach, there is a tremendous effort towards solving the word segmentation problem [8,15]. One of the main challenges of keyword spotting methods, either learning-free or learning-based, is that they usually need to segment the document images into words [10,15] or text lines [6] using a layout analysis step. In critical scenarios, dealing with handwritten text and highly degraded documents [11] segmentation is highly crucial. Any segmentation errors have a cumulative effect on subsequent word representations and matching steps.

The work of Rusinol et al. [18] avoids segmentation by representing regions with a fixed-length descriptor based on the well-known bag of visual words (BoW) framework [3]. The recent works of Rodriguez et al. [17] propose methods that relax the segmentation problem by requiring only segmentation at the text line level. In [7], Gatos and Pratikakis perform a fast and very coarse segmentation of the page to detect salient text regions. The represented queries are in the form of a descriptor based on the density of the image patches. Then, a sliding-window search is performed only over the salient regions of the documents using an expensive template-based matching.

Methods on Learning-Free or Learning-Based: Learning-based methods, such as [5,6,17], use supervised machine learning techniques to train models of the query words. On the contrary, learning-free methods, use dedicated matching scheme based on image sample comparison without any necessary training process [9,15]. Learning-based methods are preferred for applications where the keywords to spot are a priori known and fixed. If the training set is large enough they are usually able to deal with multiple writers. However, the cost of having a useful amount of annotated data available might be unbearable in most scenarios. In that sense methods running with few or none training data are preferred. Learning-based methods [14] employ statistical learning methods to train a keyword model that is then used to score query images. A very general approach was recently introduced in [14], where the learning-based approach is applied at word level based on Hidden Markov Models (HMMs). The trained word models are expected to show a better generalization capability than template images. However, the word models still need a considerable amount of templates for training and the system is not able to spot out-of-vocabulary keywords. In the above work holistic word features in conjunction with a probabilistic annotation model is also used. In [5] Fischer et al. used nine features. The first three were the features regarding the cropped window (height, width and center of gravity) and the rest were the geometric features of the contours of the writing. Peronin et al. [14] presented a very general learning-based approach at word level based on local gradient features. In our case we use LBP which is much faster and can be calculated at the run-time. In this paper we use the LBP for the first time to do a fast learning free word spotting schematic. The learning free method, unlike unsupervised methods, can be used without any kind of tuning to any database.

Methods Based on Query-By-Example (QBE) or Query-By-String (QBS). The query can be either an example image (QBE) or a string containing the word to be searched (QBS). In QBS approaches, character models typically HMMs have been pre-trained. At query time the models of the characters forming the string are concatenated into a word-model. Both approaches have their advantages and disadvantages. QBE requires examples of the word to be spotted, whereas, QBS approaches require large amounts of labeled data to train character models. The work of Almazan [1] and Rusinol et al. [18],

where the word images are represented with HOG and SIFT descriptors aggregated respectively, can successfully be applied in a retrieval scenario. Most of the popular methods either work on QBE or in QBS and the success in one paradigm cannot be replicated in another, as comparison between images and texts is not well defined. We will focus on the QBE scenario.

The LBP explained in the Sec. 3.2 uses the uniformity to reduce the dimensionality to speed up the process of matching the feature vectors in the learning free paradigm unlike other state of the art methods.

3 Proposed Method

In this section the oriented gradient property of the LBP has been used to develop a fast learning free method for information spotting for large scale document database where annotated data is unavailable.

3.1 End-to-end Pipeline Overview

In the pipeline given below we consider segmented words. Flow-diagram of our proposed approach is shown in Fig. 1. We use a median filtering preprocessing technique to reduce noise.

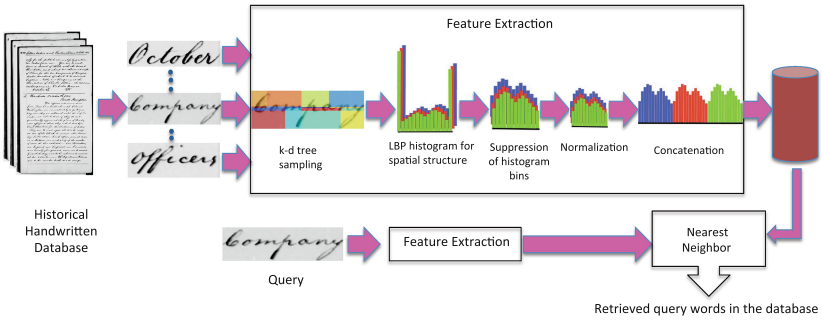


Fig. 1. Flow-diagram of oour proposed approach.

– **LBP Transform:** LBP is a scale sensitive operator whose scaling depends on the sampling rate. In our case this rate is fixed to 8. For each sub-window zone obtained in the spatial sampling state, a uniform compressed LBP histogram is generated. Each histogram is then normalized and weighted by a edge pixel ratio in the sub-window. This was perceived in this way, because the uniform LBP transform contains information regarding the sign transition of gradients which is prominent in case of the edges of the stroke width. We gave importance to the sub-windows with more edge information. The non-uniform pattern is suppressed to reduce the dimensionality of the final feature vector

and also reduce its effect on normalization. It can be seen in purple almost on the medial axis in the Fig. 2(d). The information lost by this exclusion is very less compared to that of the uniform patterns. The final feature is the concatenation of the histogram of each sub-window. Though the dimensionality increases with respect to the number of level in spatial sampling, the texture information becomes more distinctive for that space.

- **Quad Tree Spatial Sampling:** The gray level word images are then used to compute the spatial sub windows zones. This is done based on the center of mass of the image which divides the image in four quadrants. Each quadrant was further subdivided based on the center of mass of those quadrants. This brought about twenty such sub windows for the first two level. The levels are experimentally fixed using the train set. The spatial information is embedded in the final feature vector using this technique. LBP histogram is pooled over the zone created by this sampling technique as it gives more weightage to the zones having the pen strokes.
- **Nearest Neighbor:** The feature thus obtained is compared to that of the query using Bray-Curtis (BC) dissimilarity matching as shown in Eq. 1.

$$BC(a,b) = (\sum_i \mid a_i - b_i \mid)/(\sum_i a_i + b_i) \tag{1}$$

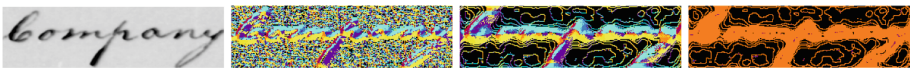
where a_i and b_i are the i-th elements of the histograms. We then use the width ratio which is the ratio between the width of the query and the images as an additional bit of information with the distance matrix. The coefficient of the width ratio was experimentally decided using the training set. Finally, the images with least distances are ranked chronologically. The performance of the system was measured by well established mean average precision, accuracy, precision and recall.

3.2 Local Binary Patterns

The LBP is an image operator, which transforms an image into an array or image of integer labels describing small-scale appearance of the image [12]. It has proven to be highly discriminating and its key points of interest, namely its in-variance to monotonic gray level changes and computational proficiency, make it suitable for demanding image analysis tasks. The basic LBP operator, introduced by Ojala et al. [13], was based on the assumption that texture has locally two complementary aspects, a pattern and its strength. LBP feature extraction consists of two principal steps: the LBP transform, and the pooling of LBP into histogram representation of an image. As explained in [13] gray scale in-variance is achieved because of the difference of the intensity of the neighboring pixel to that of the central pixel. It also encapsulates the local geometry at each pixel by encoding binarized differences with pixels of its local neighborhood:

$$LBP_{P,R,t} = \sum_{p=0}^{P-1} s_t(g_p - g_c) \times 2^P, \tag{2}$$

where g_c is the central pixel being encoded, g_p are P symmetrically and uniformly sampled points on the periphery of the circular area of radius R around g_c, and s_t is a binarization function parameter by t. The sampling of g_p is performed with bi-linear interpolation. t, which in the standard definition is considered zero, is a parameter that determines when local differences are big enough for consideration. In our version, the LBP operator works in a 3×3 pixel block of an image. The pixels in this block are threshold by its center pixel value, multiplied by powers of two and then summed to obtain a label for the center pixel. As the neighborhood consists of 8 pixels, a total of $2^8 = 256$ different labels can be obtained depending on the relative gray values of the center and the pixels in the neighborhood.



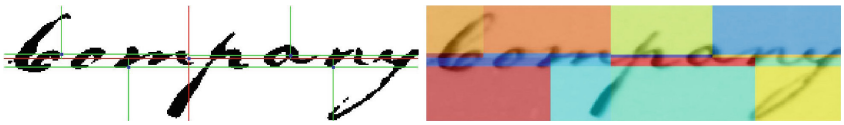
(a) (b) (c) (d)

Fig. 2. (a) Input Image (b) LBP image (c) LBP with median filtering (d) LBP uniformity with median filter.

In [13] a class of patterns called uniform is described as all patterns having at most two bit-value transitions on a clock-wise traverse. These uniform patterns provide the vast majority, over 90 %, of all the LBP patterns occuring in the analysed historical documents. By binning all non-uniform patterns into a single bin, the histogram representation goes to 59 from originally 256 bins. The final texture feature employed in texture analysis is the histogram of the operator outputs (i.e. pattern labels) accumulated over a texture sample. The reason why the histogram of 'uniform' patterns provides better discrimination in comparison to the histogram of all individual patterns comes down to differences in their statistical properties as shown in Fig. 2(d). The relative proportion of 'non-uniform' patterns of all patterns accumulated into a histogram is so small that their probabilities can not be estimated reliably.

3.3 Spatial Sampling

Since LBP histograms disregard all information about the spatial layout of the features, they have severely limited descriptive ability. We consider the spatial sampling as sub-windows on the whole image. These are obtained from the spatial pyramid with two levels. The Quad tree image sampling is based on the center of mass which yields much better results as shown in the Fig. 3. The intuition was that small sub-windows have more discriminating power than others because of their high black pixel density. The black pixel concentration suggests several handwritten letters together. To determine the sub windows, the gray images were considered for center of mass which was calculated on its binarized image obtained using Ostu's technique. On this center point the image was



(a) (b)

Fig. 3. Spatial sampling using Quad tree technique. (a) Quad tree applied based on the center of mass where the red lines is the first level of sampling with blue point as center using sub optimal binarization. The green lines are second level of sampling with blue points as center respectively. (b) The different zones of the sampling is shown by different colors.

divided into four quadrants. The number of hierarchical levels determines the total number of sub-window used. The number of hierarchical levels determines the total amount of sub-windows used, for level one being 4 sub-windows and for level 2, 16 (for each quadrant, 4 more quadrants were generated and so on). In our case, we just used level 2 which was experimentally fixed.

3.4 Nearest Neighbour K-NN

Nearest neighbor search (NNS) is an optimization problem for finding closest (or most similar) points. Closeness is typically expressed in terms of a dissimilarity function: the less similar the objects, the larger the function values. In our case it is the Bray-curtis Dissimilarity as defined in Eq. 1. We made a very naive NN technique for our pipeline.

4 Experiments

4.1 Experimental Framework

Our approach was evaluated on two public datasets (The George Washington (GW) dataset [5] and the Barcelona Historical Handwritten Marriages Database (BHHMD) [4]) which are available online. The proposed algorithm was only evaluated on segmentation-based word spotting scenarios. We have used a set of pre-segmented words to compare our approach with other methods in the literature with the aim of testing the descriptor in terms of speed, compacity and learning independence. The results for all the methods considered all the words in the test pages as queries. The used performance evaluation measures are mean average precision (mAP), precision and rPrecision. Given a query, we label the set of relevant objects with regard to the query as *rel* and the set of retrieved elements from the database as *ret*. The precision is defined in terms of *ret* and *rel* in Eq. 3. rPrecision is the precision at rank *rel*. For a given query, $r(n)$ is a binary function on the relevance of the n-th item returned in the ranked list. The used performance evaluation measures is mean average precision (mAP) which defined in Eq. 3.

$$Precision(P) = \frac{\mid ret \cap rel \mid}{\mid ret \mid}, \; mAP = \frac{\sum_{n=1}^{|ret|} (P@n \times r(n))}{\mid rel \mid} \tag{3}$$

4.2 Results on George Washington Dataset

The George Washington database was created from the George Washington Papers at the Library of Congress. The dataset was divided in 15 pages for training and validation and the last 5 pages for testing. Table 1 shows the performance of our method compared to others. The best results are highlighted in each category. Here, Quad Tree method is an adaptation of [19].

Table 1. Retrieval results on the George Washington dataset

Method	Learning	mAP	Accuracy (P@1)	rPrecision	Speed(secs) (on Test)
Quad-Tree	Standardization	15.5 %	30.14	15.32	44.41
BoVW [18]	Unsupervised, codebook	68.26 %	85.87	62.56	NA
FisherCCA [2]	Supervised	**93.11** %	**95.44**	**90.08**	137.63
DTW [16]	No	20.94 %	41.34	20.3	78095.89
HOG pooled Quad-Tree	No	48.22 %	64.96	43.04	45.34
Proposed method	No	**54.44** %	**72.86**	**48.87**	**43.14**

Some qualitative results are shown in Fig. 4. It is interesting to see that most of the words have been retrieved. This example takes the image *Company* as a query. The system correctly retrieves the first 15 words, whereas the 16th is *observing*, which is very similar to the query word in length and pattern. The next word is again a correct retrieval.

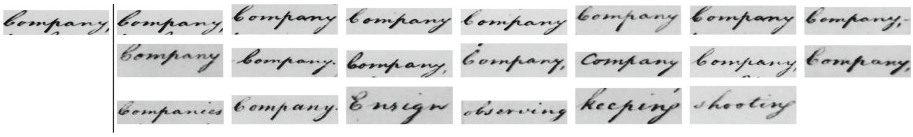


Fig. 4. Qualitative result examples. Query (estreme left) and First 20 retrieved words.

4.3 Results on BHHMD Dataset

This collection consists of marriage registers in the Barcelona area between the 15th and 20th centuries. The ground truth-ed subset contains 40 images. The dataset was divided in 30 pages for train and validation and the last 10 pages for test. Table 2 shows the results. Our method is the second best performing, with best computational time in this dataset. We have also performed tests for cross-dataset evaluation and our method in all categories is just 2 % behind the best state of the art method.

Table 2. Retrieval results on the BHHMD dataset

Method	Learning	mAP	Accuracy(P@1)	rPrecision
Quad-Tree	Standardization	38.4 %	61.92	48.47
FisherCCA [2]	Supervised	**95.40** %	**95.49**	**94.27**
HOG pooled Quad-Tree	No	66.66 %	80.59	62.35
DTW [16]	No	7.36 %	4.69	2.99
Proposed method (LBP)	No	**70.84** %	**84.13**	**70.44**

5 Conclusion

We have proposed a fast learning free word spotting method based on LBP-representations and a k-d tree sampling approach. The most important contribution is that the proposed word spotting approach has been shown to be the best among the learning free ones in terms of performance. The computational speed measured with the same benchmark for the proposed method is best compared to other state of the art methods. We have shown that LBP based on uniformity can be stable under the deformations of handwriting. For the pooling approach, the main contribution of the proposed framework is the pooling of the LBP based on the Quad Tree zones. The LBP has been defined as the textural feature which we use as oriented texture recognition. A sampling architecture has been designed to maximize the usage of the pen strokes and preserve the LBP patterns specific to the region. In terms of a retrieval problems, competitive mAP was obtained. The time complexity of this indexation is linear with the number of words in the database. This was reduced to the order of $logN$ by using a k-NN approach. It leads us to conclude that a feature extraction scheme as it is proposed here is very useful to compute inexact matching in large-scale scenarios. We have demonstrated that compact textural descriptors are useful information for handwritten word spotting, despite the variability of handwriting. The experimental results demonstrates that our approach is comparable to other statistical approaches in terms of performance and time requirements.

Future work will focus on the evaluation of the stability of LBP-based representations in large multi-writer document collections.

Acknowledgment. This work has been partially supported by the Spanish project TIN2015-70924-C2-2-R, and RecerCaixa, a research program from ObraSocial La Caixa. We are thankful to the authors of the compared methods for their support.

References

1. Almazán, J., Gordo, A., Fornés, A., Valveny, E.: Efficient exemplar word spotting. In: BMVC, vol. 1, p. 3 (2012)
2. Almazán, J., Gordo, A., Fornés, A., Valveny, E.: Segmentation-free word spotting with exemplar svms. Pattern Recogn. **47**(12), 3967–3978 (2014)

3. Csurka, G., Dance, C., Fan, L., Willamowski, J., Bray, C.: Visual categorization with bags of keypoints. In: Workshop on Statistical Learning in Computer Vision, ECCV, Prague, vol. 1, pp. 1–2 (2004)
4. Fernández-Mota, D., Almazán, J., Cirera, N., Fornés, A., Lladós, J.: BH2M: the barcelona historical, handwritten marriages database. In: 2014 22nd International Conference on Pattern Recognition (ICPR), pp. 256–261. IEEE (2014)
5. Fischer, A., Keller, A., Frinken, V., Bunke, H.: Lexicon-free handwritten word spotting using character hmms. Pattern Recogn. Lett. **33**(7), 934–942 (2012)
6. Frinken, V., Fischer, A., Manmatha, R., Bunke, H.: A novel word spotting method based on recurrent neural networks. IEEE Trans. Pattern Anal. Mach. Intell. **34**(2), 211–224 (2012)
7. Gatos, B., Pratikakis, I.: Segmentation-free word spotting in historical printed documents. In: 10th International Conference on Document Analysis and Recognition, ICDAR 2009, pp. 271–275. IEEE (2009)
8. Ghosh, S.K., Valveny, E.: A sliding window framework for word spotting based on word attributes. In: Paredes, R., Cardoso, J.S., Pardo, X.M. (eds.) IbPRIA 2015. LNCS, vol. 9117, pp. 652–661. Springer, Heidelberg (2015). doi:10.1007/978-3-319-19390-8_73
9. Leydier, Y., Ouji, A., LeBourgeois, F., Emptoz, H.: Towards an omnilingual word retrieval system for ancient manuscripts. Pattern Recogn. **42**(9), 2089–2105 (2009)
10. Liang, Y., Fairhurst, M.C., Guest, R.M.: A synthesised word approach to word retrieval in handwritten documents. Pattern Recogn. **45**(12), 4225–4236 (2012)
11. Louloudis, G., Gatos, B., Pratikakis, I., Halatsis, C.: Text line and word segmentation of handwritten documents. Pattern Recogn. **42**(12), 3169–3183 (2009)
12. Nicolaou, A., Bagdanov, A.D., Liwicki, M., Karatzas, D.: Sparse radial sampling lbp for writer identification. arXiv preprint arXiv:1504.06133 (2015)
13. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
14. Perronnin, F., Rodriguez-Serrano, J., et al.: Fisher kernels for handwritten wordspotting. In: 10th International Conference on Document Analysis and Recognition, ICDAR 2009, pp. 106–110. IEEE (2009)
15. Rath, T.M., Manmatha, R.: Features for word spotting in historical manuscripts. In: Proceedings of the Seventh International Conference on Document Analysis and Recognition, pp. 218–222. IEEE (2003)
16. Rath, T.M., Manmatha, R.: Word image matching using dynamic time warping. In: Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 521–527. IEEE (2003)
17. Rodriguez-Serrano, J., Perronnin, F., et al.: A model-based sequence similarity with application to handwritten word spotting. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11), 2108–2120 (2012)
18. Rusinol, M., Aldavert, D., Toledo, R., Lladós, J.: Browsing heterogeneous document collections by a segmentation-free word spotting method. In: 2011 International Conference on Document Analysis and Recognition (ICDAR), pp. 63–67. IEEE (2011)
19. Sidiropoulos, P., Vrochidis, S., Kompatsiaris, I.: Content-based binary image retrieval using the adaptive hierarchical density histogram. Pattern Recogn. **44**(4), 739–750 (2011)