

A Coarse-to-Fine Word Spotting Approach for Historical Handwritten Documents Based on Graph Embedding and Graph Edit Distance

Peng Wang^{1,2}, Véronique Eglin¹,
Christophe Garcia¹
¹LIRIS, CNRS UMR5205
INSA-Lyon,
F-69621, Villeurbanne, France
peng.wang@liris.cnrs.fr

Christine Largeton²
²LAHC, CNRS UMR5516
Université Jean Monnet
F-42023, Saint Etienne, France
christine.largeton@univ-st-etienne.fr

Josep Lladós³, Alicia Fornés³
³Computer Vision Center,
Universitat Autònoma de Barcelona,
08196, Bellaterra, Spain
{josep, afornes}@cvc.uab.es

Abstract—Effective information retrieval on handwritten document images has always been a challenging task, especially historical ones. In the paper, we propose a coarse-to-fine handwritten word spotting approach based on the graph representation. The presented model comprises both the topological and morphological signatures of the handwriting. Skeleton-based graphs with the Shape Context labelled vertexes are established for connected components. Each word image is represented as a sequence of graphs. Aiming at developing a practical and efficient word spotting approach for large-scale historical handwritten documents, a fast and coarse comparison is first applied to prune the regions that are not similar to the query based on the graph embedding methodology. Afterwards, the query and regions of interest are compared by graph edit distance based on the Dynamic Time Warping alignment. The proposed approach is evaluated on a public dataset containing 50 pages of historical marriage license records. The results show that the proposed approach achieves a compromise between efficiency and accuracy.

Keywords—word spotting, coarse-to-fine mechanism, graph-based representation, graph embedding, graph edit distance

I. INTRODUCTION

A lot of libraries contain an enormous amount of handwritten historical documents which are interesting to a great range of people, especially literature experts, historians, scientists from different fields. The demands for efficient and effective information retrieval techniques on such collections are strongly increasing, especially for the large-scale datasets. Word spotting in handwritten document images is one of the expected techniques. As a specific pattern recognition task, the success achieved on the printed document images cannot be directly transplanted since the handwritten word spotting faces many particular challenges such as the high intra-writer and inter-writer variability. Moreover, to realize such an indexing system for the historical documents, it must overcome a large amount of difficulties, such as multi-source degradation, ink show-through, artificial noises and large lexicons in unconstrained environment etc.

Generally, there are two groups of methods addressing

the handwritten word spotting problem. One is dedicated to the detection of predefined words (associated to a training process, like the works presented in [1, 2]), and the other one is retrieval-oriented with a dedicated matching scheme based on image sample comparison without any training process [3, 4]. In both cases and from all works dealing with the access to handwritten content, it has been proved that an accurate and comprehensive representation of the image content is necessary [1, 2, 3, 4]. Moreover, the similarity measure is supposed to be enough robust to tolerate the inner variations in the handwriting.

Graph-based representation is a popular approach to capture and model the structural properties of objects. It has been widely used in the pattern recognition domain, for instance, in molecular compound recognition in chemistry and symbol recognition in document analysis. However, the employment of graph representation in the handwritten word spotting application is still very rare. There are only a few attempts made in handwritten recognition researches with graph representation or relative concepts up to now. Researchers started using graphs in the context of single character recognition, such as the graph-based recognition of Chinese characters reported in [5, 6]. Even though the hierarchical model proposed in [5] reflects the hierarchical nature of handwritten characters, especially of Chinese characters, which has a more complex structure than Latin languages, in general, a more sophisticated representation mechanism is required to distinguish similar characters. Later, Fischer et al. developed a graph representation model based on the skeleton of word images, which contains only vertexes without edges [7, 8]. By adding enough intersection points among keypoints, the structural information is still preserved and meanwhile graph representation is robust to the bad quality of skeleton. However, it also leads to the fact that graphs may contain too many vertexes. Recently, due to the expensive computation of graph matching, the transformation of graphs into feature vectors calls the concern. In [9], a directed acyclic graph is constructed for each subword first, and then is converted into a topological signature vector that

preserves the graph structure. Moreover, Lladós et al. [10] adapted the graph serialization concept in graph matching for handwritten word spotting. By extracting and clustering the acyclic paths of graphs, one-dimensional descriptor, bag of-paths is generated for describing the words. This attempt is interesting but the performance is far away from satisfying.

In our work, the motivation for adapting graph-based representation for handwritten word spotting rather than using the appearance-based representation model is mainly due to the fact that the structure is more stable and relevant to the handwriting than the pixels. Furthermore, topological characteristics are very important for describing the handwriting and the two-dimensional nature of handwriting is difficult to be well described by a single one-dimensional sequential scalar feature vector.

In this paper, we propose an original coarse-to-fine handwritten word spotting approach based on the graph representation. First of all, we use Shape Context labelled graphs to represent the handwriting images. In the coarse selection, graphs are converted into a vector space to realize a fast comparison between the query and the test images. Regions of interest are selected for the further comparison. In the fine selection, we develop a comparison scheme based on the Dynamic Time Warping and block merging with the graph edit distance to measure the difference between the regions of interest and the query.

The rest of the paper is organized as follows. Section II is devoted to illustrate the general idea of the proposed approach. Afterwards, all the preprocessing techniques used in the approach are introduced in Section III. The graph-based representation model proposed for handwriting images is presented in Section IV. The details of the coarse selection and the fine selection are explained respectively in Section V and VI. Section VII shows the experiments that are conducted to evaluate the proposed approach. In the end, the conclusion and the perspectives are given in Section VIII.

I. GENERAL ALGORITHM DESCRIPTION

Given a dataset of handwritten document images, the pages are first segmented into lines. Afterwards, we extract the skeleton of the handwriting and detect three types of structural points based on it. Taking the structural points as the vertexes and the strokes as the edges, the handwriting is represented as graphs that comprise the spatial 2-dimensional relations between handwritten strokes. In order to encode the contextual information into the graph-based representation, the Shape Context descriptor is used as the attributes of the vertexes. In this way, the representation model preserves both topological and morphological signatures of the handwriting.

To make the approach more efficient, the retrieval process is executed in two steps. They are respectively called the coarse selection and the fine selection. The coarse selection is designed to realize a fast comparison between the query and the test images. By adapting the explicit

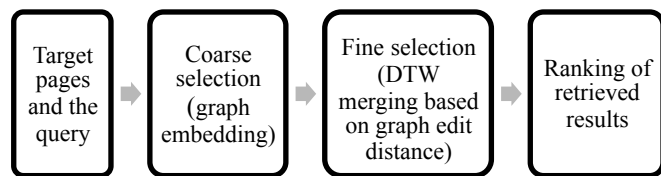


Fig. 1. General workflow of the proposed approach

graph embedding approach, 2-dimensional handwriting is converted into a feature vector in terms of single vertexes and bipartite relations between vertexes. Based on the comparison of graph embedding vectors, regions of interest are selected. Most of the area on the test page is ignored afterwards. The objective of this step is to eliminate the regions that are not similar to the query according to the graph embedding distance so as to reduce the comparison in the next step. Consequently, the recall must be as high as possible after the coarse selection. With regions of interest, the attribute of vertexes in the graph is recalculated in terms of the entire contour inside the region of interest. A more precise and structure-based comparison between the query and the regions of interest is adapted based on the graph edit distance. The similarity distance obtained at this step is used as the final score to rank the candidates. Fig.1 shows a general workflow of the proposed approach.

II. PREPROCESSING

Before building the graph-based representation, several processes are applied. First, the margins of the document are removed. The page is then segmented into lines using projection analysis techniques. The connected components are extracted based on binarized images.

In order to obtain the topological information of the handwriting, we apply the Zhang & Suen's [12] parallel thinning algorithm on binarized images to extract the skeleton of the text, as shown in Fig.2(c). Afterwards, three types of structural points related to the nature of the language and the handwriting style are detected on the skeleton by using the method introduced in [13]. These structural points are respectively starting/ending points, high-curved points and branch points. An example of three types of structural points is given in Fig.2(d). According to the location of structural points, the handwriting is segmented into strokes. A contour tracing approach based on the border following algorithm [14] is also applied to the handwriting as the preparation of the Shape Context descriptor extraction.

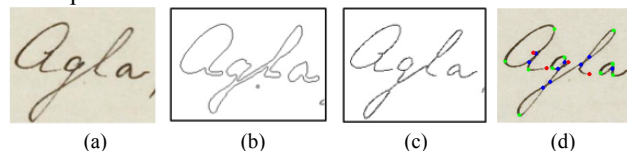


Fig. 2. (a) a handwritten word (b) the contour (c) the skeleton (d) the structural points (red---starting/ending points, green---high-curved points, blue---branch points)

III. GRAPH-BASED REPRESENTATION

The graph representation is a very powerful method of

representing images in terms of preserving structural properties. As mentioned before, our graph-based representation model is established on the skeleton of the handwriting. The vertices of the graph correspond to the structural points and the strokes connecting them are used as the edges (see Fig.3). One connected component can generate an individual graph. The formal definition of the graph is given below.

Definition 1. (Graph) A graph is a four-tuple $g = (V, E, \mu, \nu)$, where

- V is the finite set of vertexes, corresponding to the structural points
- $E \subseteq V \times V$ is the set of edges, corresponding to the strokes
- $\mu: V \rightarrow L$ is the vertex labelling function, corresponding to the Shape Context descriptor
- $\nu: E \rightarrow L'$ is the edge labelling function, corresponding to the length of the stroke

We want to point out that the vertex labelling functions used in the coarse selection and the fine selection are different. Since we do not segment the text into words, it is difficult to apply a global Shape Context extraction. Hence, in the coarse selection, for each vertex, we only extract a local Shape Context descriptor based on the skeletonized strokes, which are corresponding to the edges located in the first level of the neighbourhood of the reference vertex in the graph. Then for the graph representation in the fine selection, we calculate the Shape Context descriptor based on the contour within the scope of the entire region of interest as the labelling function of the vertex.

Usually, a handwritten word is composed of several connected components. Therefore, a word image in our approach is represented by a sequence of graphs. For example, the word “Savay” in Fig. 3 is represented by three graphs respectively corresponding to “S”, “av” and “ay”.

IV. COARSE SELECTION

In the coarse selection, we use the fuzzy C-Means clustering function based on the local Shape Context descriptor as the labelling function of the vertex. Graphs are converted into 1-D vectors with the specific graph embedding approach presented in [15]. A sliding window scans the lines in terms of connected components. According to the width of the query, the size of the sliding window is adjusted automatically to contain a suitable amount of the connected components depending on their sizes. The graph embedding feature is recalculated within the sliding window at each potential location. The distance between the sliding window and the query based on the graph embedding feature is used to grade the candidates. Only half of the candidates are preserved as the regions of interest.

A. Local Shape Context Extraction

The Shape Context descriptor [11] is a point descriptor which captures the distribution over relative positions of the shape points and thus summarizes the global shape in a rich,

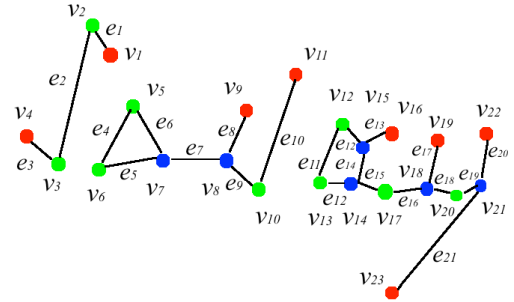


Fig. 3. Graph representation of a word “Savay”

local polar diagram. In order to adapt the Shape Context descriptor to our application, we especially design a dynamic selection of the relative shape points to build the local Shape Context descriptor based on the 1-step neighbourhood of the reference point on the graph.

Originally, according to the graph theory [16], a k -Neighbourhood (kN) of a vertex v is an induced subgraph formed from all the vertexes that can be reached within k steps from v . This induced subgraph is centred around the vertex v and contains all vertexes up to k steps away. In our case, we keep the same definition of the k -Neighbourhood but replace the edges by the corresponding skeleton of the handwritten strokes. It means that the Shape Context extracted in the k -Neighbourhood of a reference point records the relative positions of the skeleton points located in the corresponding area. In practice, we use $k=1$. Fig. 4 shows the 1-Neighbourhood of the blue point and the red point in the example “a”. The polar diagram indicates the contribution of all angular sectors within the 1-Neighbourhood of the reference vertex.

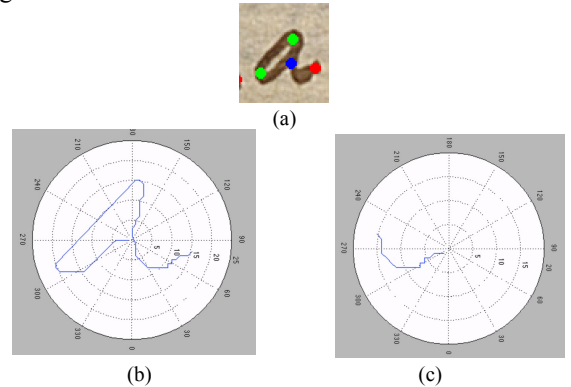


Fig. 4. (a) an example of the letter “a” with the structural points (b) 1-Neighbourhood of the blue point (c) 1-Neighbourhood of the red point

B. Vertex labelling with Fuzzy C-means Clustering

Instead of assigning the exact local Shape Context feature as the attribute of the vertexes, we decide to build a codebook of the local Shape Context features with the size of n classes and use the representative of the corresponding class as the attribute of the vertexes. The total amount of classes is a tunable parameter that influences the effectiveness of the classification. After testing several different values, we use $n=10$ in our experiment which gives the best performance. Moreover, in noisy situations, it can happen quite frequently that a vertex label is between

V. FINE SELECTION

two representatives, and there is no clear rule telling us to which representative the vertex should be assigned. In such case, a soft rather than a hard assignment can be beneficial. Hence, we choose to employ the Fuzzy C-means clustering algorithm [17], which can assign to every vertex a certain degree of belongingness to every cluster, rather than an exclusive assignment. Let $P \subset R^d$ be the set of all vertex labels in all the graphs of G . Furthermore, let $\mathcal{W} = \{w_1, \dots, w_n\}$ be a set of n representatives of all vectors in P . Its main idea is to assign to a point $x \in P$ a degree of belongingness to each cluster in \mathcal{W} , which is inversely proportional to the distance between x and the cluster center. This leads to

$$p_i(x) = \alpha \cdot \left(\frac{1}{\|x - w_i\|_2} \right)^s \quad (1)$$

where α is a constant assuring that $\sum_{i=1}^n p_i(x) = 1$ and s is a parameter that controls the amount of fuzziness that the user is giving to the assignment. The larger s is the more weight is given to points close to the centers. In our experiment, we use $s=2$. With the Fuzzy C-means assignment algorithm, we define the vertex labelling function

$$v \mapsto \lambda_s(v) = (p_1(v), \dots, p_n(v)) \quad (2)$$

where $p_i(v) = P(v \in w_i)$ is the probability of the vertex v being assigned to the representative w_i , while $p_i(v) \geq 0$.

C. Graph Embedding Representation

For the graph embedding, we adapt J. Gibert's [15] approach based on the vertex attribute statistics. It redefines the appearance of a specific label by checking the probabilities of belongingness for all vertexes in the graph. This is given a graph $g = (V, E, \mu)$ and a representative set \mathcal{W} , the frequency of a representative $w_i \in \mathcal{W}$ is

$$U_i = \#(w_i, g) = \sum_{v \in V} p_i(v) \quad (3)$$

Similarly, when there is an edge between two vertexes in the graph $(u, v) \in E$, the fact that vertexes are assigned to representatives according to Eq.2. makes the contribution to all relations between any two representatives. It is defined as:

$$B_{ij} = \#(w_i \leftrightarrow w_j, g) = \sum_{(u,v) \in E} p_i(u)p_j(v) + p_j(u)p_i(v) \quad (4)$$

The formal definition of graph embedding used in our approach is given as follows.

Definition 2. (Graph Embedding) Given a set of vertex representatives $\mathcal{W} = \{w_1, \dots, w_n\}$, we define the embedding of a graph g into a vector space as the vector

$$\varphi_w(g) = (U_1, \dots, U_n, B_{11}, \dots, B_{ij}, \dots, B_{nn}) \quad (5)$$

where $1 \leq i \leq j \leq n$, $U_i = \#(w_i, g)$ and $B_{ij} = \#(w_i \leftrightarrow w_j, g)$.

Since a word image usually contains several connected components, in order to adapt the graph embedding approach to represent the entire word image, we sum up all the graph embedding vectors of the graphs inside the image. The χ^2 distance is used to calculate the distance between graph embedding features.

After obtaining the regions of interest, we adapt a more discriminative and accurate comparison approach developed in our previous work [18]. The graph representation model for the handwriting is almost the same as the one used in the coarse selection step except the labelling function for the vertex. Instead of using the probabilities of belongingness to the representatives based on the local Shape Context obtained from the skeleton, we extract the entire contour of the region of interest and use all of it to build the Shape Context descriptor. We name the Shape Context extracted in this step the global Shape Context. Without any clustering, the global Shape Context descriptor is used directly as the attribute of the vertex.

Since the two-dimensional topological signature is very important, we choose to use an approximate graph edit distance method [19] based on the bipartite matching to further compare two graphs in the consideration of the computational complexity. The details are omitted due to the limited space. Dynamic Time Warping (DTW) is selected to find the assignments among sequences of connected components due to its flexibility. As a major variation of handwriting, the amount of connected components of the same word can vary a lot in different instances as the example shown in Fig. 5. In order to make the similarity measure robust to such variation, based on the matching results, a two-direction merging operation is executed. The graphs assigned to the same connected components are considered as one entire graph. In this way, both the sequences of graphs represent the query and the test image might change. Afterwards, with the new sequences of graph representation, graph matching is performed again. The average distance among corresponding graph matchings is considered as the final distance between two word images.



Fig. 5. Two instances of the word ‘‘Otheos’’ with different amounts of connected components

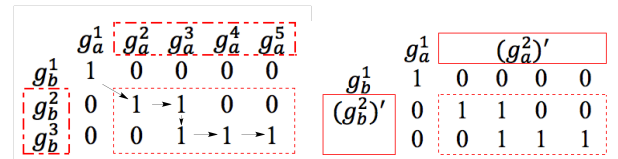


Fig. 6. Illustration of the merging block operation based on the DTW result of the examples in Fig. 5 (the arrows indicate the path of the optimal assignments)

VI. EXPERIMENTS

In this part, the proposed approach is applied to a large collection of historical documents on the marriage licenses. The results are compared to the performance of another coarse-to-fine word spotting approach proposed by J. Almazán et al [20]. The reference approach first uses a pseudo-structure based representation embedded in a

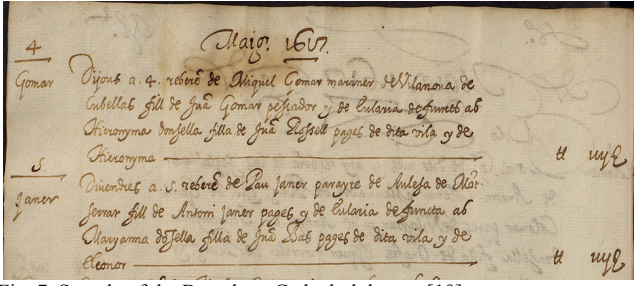


Fig. 7. Sample of the Barcelona Cathedral dataset [19]

hashing structure to coarsely eliminate word images that are unlikely to be instances of the query. Afterwards, they use a SVM to learn a representative model for each query word class based on the HOG description. Differently from the proposed approach, the reference system segments the lines into words using a projection function. Moreover, the reference system needs more than one sample of the query word class and a lot of the negative samples to train the representative model of each word class. We show the evaluation results and the discussion in the latter part of this section.

A. Experimental Data and Evaluation Protocol

The dataset used in the experiment is a collection from the public *5CofM* dataset, which contains scanned marriage licenses of the Barcelona Cathedral between 1451 and 1905 [20]. The resolution of the images is around 20MB per page. The ground truth contains 50 pages from one volume written by the same writer (see Fig. 7).

In order to compare our system with the reference approach, we choose the same query word classes and the same evaluation protocols used in [20]. Some query examples are shown in Fig. 8. For the overall evaluation, we consider that a region is classified as positive if it overlaps more than 50% with the annotated bounding box in the ground truth and negative otherwise. Finally, we combine the retrieved regions of all the documents and rerank them according to their scores. To evaluate the performance of the proposed retrieval system, we use the same measures as the reference approach: the Average Recall (AR) and the Average Precision (AP).



Fig. 8. Some queries used in the experiments

B. Results and Discussion

Table I gives the average recall for the coarse selection corresponding to each query word class. In practice, there are around 800 sliding windows per page in average that are compared to the query with the graph embedding approach. Only half of the regions (~400) are preserved as the regions

of interest for the fine selection. Compared to the reference system, which select approximately 40,000 regions in a single page after the coarse selection, the proposed approach is more efficient and feasible for the large-scale collection, especially when the computational time of the fine selection linearly increases with the number of regions of interest.

TABLE I. AVERAGE RECALL OF THE COARSE SELECTION

Queries	Coarse Selection (%)
barna	77.51
fill	77.54
filla	82.18
habitant	82.81
pages	85.06
viuda	67.21
viudo	82.47
reberé	65.30
mAR	77.51

Table II shows the results of the proposed approach and the reference system with two different amounts of the training samples in terms of the average precision. It can be seen that the proposed approach outperforms the reference system in four word classes. Even though the overall performance of the reference system using 3 training samples is slightly better than our approach, sometimes, it is not feasible for the users to obtain 3 instances of the query in advance in reality. Besides, we can notice that as the amount of instances reduces, the performance of the reference approach decreases. Moreover, the reference approach achieves a good performance also due to an accurate word segmentation which mainly thanks to the regular word separation in the evaluation dataset. However, in reality, it is difficult to produce accurate word segmentation on handwriting images because of the irregular space between words. The hypothesis that the handwriting document must be initially segmented into words is a very hard constraint. Its application to irregular handwritten pages with high uncertainty on inter-words spaces can be compromised. By privileging a segmentation-free method we do not constrain the system with a precise word extraction. Consequently, our system also suffers the imprecision at the extreme locations of the sliding windows. The phenomenon has been encountered in situations where the last characters are represented by isolated connected component. For example, the separate two connected components in "fill a" causes that the word "fill" becomes the distortion.

TABLE II. AVERAGE PRECISION FOR DIFFERENT QUERY CLASSES OF THE PROPOSED APPROACH AND THE REFERENCE SYSTEM WITH DIFFERENT NUMBER OF TRAINING SAMPLES (3 AND 10)

Queries	Proposed	Ref. System 3	Ref. System 10
barna	55.56	40.34	48.84
fill	56.66	42.09	48.15
filla	48.21	57.39	66.66
habitant	55.83	74.10	87.01
pages	70.20	76.55	80.55
viuda	49.67	69.47	77.44
viudo	40.92	33.13	34.30
reberé	75.67	74.78	75.94
mAP	56.59	58.48	64.86

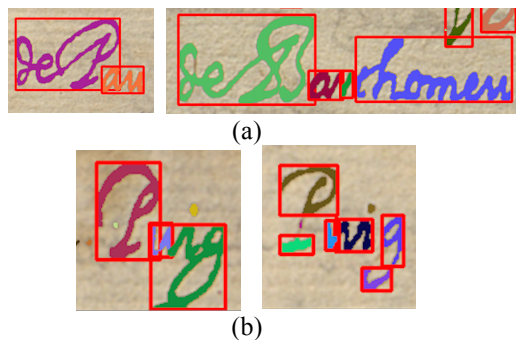


Fig. 9. Examples of limitations of the proposed approach (a) Two words are connected (b) Connected component variation due to the degradation

Another situation of imprecision due to the irregularity of inter-word and inter-character spaces is illustrated in Fig. 9. The example shown in Fig. 9(a) illustrates that the sliding window moving with the interval proportional to the connected component size cannot correctly locate the starting position of the candidates in such scenario. Another reason that can lead to the decrease of the performance is the degradation of the image (see Fig. 7(b)). Both the connected component decomposition and the skeleton of the handwriting become unstable because of the discontinuity of the ink, especially after the binarization.

In a word, our system has proved its efficiency on both short (four characters large) and large queries (containing more than eight characters) and can be applied on a very large diversity of document images without any learning step nor word segmentation. Since we are currently working on the retrieval of words that cannot be easily transcribed on a specific dataset of three hundred thousand documents associated to the CITERE project¹, the proposed approach is more suitable.

VII. CONCLUSION AND FUTURE WORK

In this paper, we propose a new handwritten word spotting approach that takes advantage of a coarse-to-fine selection scheme based on the graph representation. The proposed system first selects some regions of interest based on an off-line graph embedding representation that, for each vertex, models its relations with its closest neighbours. Secondly, an on-line discriminative comparison that lies on the DTW alignment and the graph edit distance between the query and target images is developed to order the candidates by the similarity. The overall performance of the system is evaluated on a public dataset in terms of the precision and the recall. It reveals the efficiency of our segmentation-free and learning-free approach compared to other similar effective methods. It is also a proof of the feasibility and the quality of our exclusively graph-based model exclusively. Some improvements can be proposed to adapt the model to other datasets, especially the dataset of the ANR CITERE project¹ that does not contain any formal ground truth.

Since the approach proposed in this paper only considers the word spotting within the same writing style, to make it more practical, we plan to introduce learning mechanism into the approach in order to cope with different

writing styles and reduce the influence of the variability between styles or writers.

ACKNOWLEDGEMENT

We want to thank Prof. Antony McKenna who offered us images. The research is supported by the French Rhone-Alpes region program, CITERE ANR project and partially supported by the Spanish project TIN2012-37475-C02-02, and by the EU project ERC-2010-AdG- 20100407-269796.

REFERENCES

- [1] A. Fischer, A. Keller, V. Frinken, and H. Bunke, "HMM-based word spotting in handwritten documents using subword models," in Proc. of ICPR, pp. 3416-3419, 2010.
- [2] J. Rodríguez-Serrano, and F. Perronnin, "Handwritten word spotting using hidden markov models and universal vocabularies," PR, vol. 42(9), pp. 2106-2116, 2009.
- [3] T. Rath, and R. Manmatha, "Word spotting for historical documents," in Proc. of ICDAR, vol. 9(2-4), pp. 139-152, 2007.
- [4] Y. Leydier, A. Ouji, F. LeBourgeois, and H. Emptoz, "Towards an omnilingual word retrieval system for ancient manuscripts," PR, vol. 42(9), pp. 2089-2105, 2009.
- [5] S. Lu, Y. Ren, and C. Suen, "Hierarchical attributed graph representation and recognition of handwritten Chinese characters," PR, vol. 24(7), pp. 617-632, 1991.
- [6] M. Zaslavskiy, F. Bach, and J.P. Vert, "A path following algorithm for the graph matching problem," PAMI, vol. 31(12), pp. 2227-2241, 2009.
- [7] A. Fischer, C. Y. Suen, V. Frinken, K. Riesen, and H. Bunke, "A fast matching algorithm for graph-based handwriting recognition," Lecture note in computer science, vol. 7877, pp. 194-203, 2013.
- [8] A. Fischer, K. Riesen, and H. Bunke, "Graph similarity features for HMM-Based handwriting recognition in historical documents," in Proc. of ICFHR, pp. 253-258, 2010.
- [9] Y. Chherawala, R. Wisnovsky, and M. Cheriet, "TSV-LR: Topological signature vector-based lexicon reduction for fast recognition of pre-modern Arabic subwords," in Proc. of the workshop on historical document imaging and processing, pp. 6-13, 2011.
- [10] J. Lladós, M. Rusiñol, A. Fornés, D. Fernández, and A. Dutta, "On the influence of word representations for handwritten word spotting in historical documents," International journal of pattern recognition and artificial intelligence, vol. 26(5), pp. 1263002.1-1263002.25, 2012.
- [11] J. Puzicha, S. Belongie, and J. Malik, "Shape matching and object recognition using shape contexts," PAMI, vol. 24, pp. 509-522, 2002.
- [12] T. Y. Zhang, and C. Y. Suen, "A fast parallel algorithm for thinning digital patterns," Communication of the ACM, vol. 27, pp. 236-239, 1984.
- [13] P. Wang, V. Eglin, C. Largeton, A. McKenna, and C. Garcia, "A comprehensive representation model for handwriting dedicated to word spotting," in Proc. of ICDAR, pp. 450-454, 2013.
- [14] K. Abe, and S. Suzuki, "Topological structural analysis of digital binary images by border following," Computer vision, graphics and image processing, vol. 30, pp. 32-46, 1985.
- [15] J. Gibert, E. Valveny, and H. Bunke, "Graph embedding in vector spaces by node attribute statistics," PR, vol. 45, pp. 3072-3083, 2012.
- [16] N. Dahm, H. Bunke, T. Caelli, and Y. Gao, "A unified framework for strengthening topological node features and its application to subgraph isomorphism detection," in workshop of GBR, vol. 7877, pp. 11-20, 2013.
- [17] J. C. Bezdek, "Pattern recognition with fuzzy objective function algorithms," Plenum Press, 1981.
- [18] P. Wang, V. Eglin, C. Garcia, C. Largeton, J. Lladós, and A. Fornés, "A novel learning-free word spotting approach based on graph representation," in Proc. of DAS, pp. 207-211, 2014.
- [19] K. Riesen and H. Bunke, "Approximate graph edit distance computation by means of bipartite graph matching," Image and Vision Computing, vol. 27(7), pp. 950-959, 2009.
- [20] J. Almazán, D. Fernández, A. Fornés, J. Lladós, and E. Valveny, "A coarse-to-fine approach for handwritten word spotting in large scale historical documents collection," in Proc. of ICFHR, pp. 453-458, 2012.

¹ CITERE ANR Project

<http://citere.hypotheses.org>