Pattern Recognition 42 (2009) 3374–3382

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

A probabilistic method for keyword retrieval in handwritten document images

Huaigu Cao*, Anurag Bhardwaj, Venu Govindaraju

Center for Unified Biometrics and Sensors (CUBS), Department of Computer Science and Engineering, University at Buffalo, Amherst, NY 14260, USA

ARTICLE INFO

Article history:
Received 8 August 2008
Received in revised form 17 January 2009
Accepted 5 February 2009

Keywords:
Word spotting
Information retrieval
Handwriting recognition

ABSTRACT

Keyword retrieval in handwritten document images is a challenging task because handwriting recognition does not perform adequately to produce the transcriptions, specially when using large lexicons. Existing methods build indices using OCR distances or image features for the purpose of retrieval. These alternative methods are complimentary to the traditional approaches that build indices on OCR'ed text. In this paper, we describe an improvement to the existing keyword retrieval (word spotting) methods by modeling imperfect word segmentation as probabilities and integrating these probabilities into the word spotting algorithm. The scores returned by the word recognizer are also converted into probabilities and integrated into the probabilistic word spotting model.

1. Introduction

Keyword retrieval in handwritten document images is a high-level application that relies on document analysis and recognition techniques. There are two common approaches to keyword retrieval from handwritten documents. In the first approach [1–8], image-to-image matching is used. During retrieval, each keyword is converted into a word image. This is done by annotating a small set of word images or collecting the user's handwriting on-line. When a user provides a query word, the similarity between the query and any word image in the database is computed. All of the word images are returned in the decreasing order of the similarities between them and the query. The similarity between two word images is measured as a distance between the two feature vectors computed from the word images. In [1,3], the similarity between the feature vectors of two word images is computed by dynamic time warping (DTW) matching of profile features using various definitions of matching distances [1,9,10,3,11] in the feature space. The GSC-matching method [2,12] is based on bitwise matching of the corresponding GSC features of two word images. Thus, word spotting is a useful alternative when a full fledged handwriting recognition system is not available.

However, word spotting requires on-line matching which is time-consuming. Trade-off between accuracy and speed has to be made in order to scale to large databases. Thus, in order to be fast matching-based indexing approaches are limited in feature selection and the complexity of matching and training methods. This also limits their scope to applications dealing with a single writer or small lexicons.

In contrast, OCR score-based indexing approaches [13–15] do not face the speed problem. In these methods, the indices are built from OCR scores such as posterior probabilities or feature vector observational likelihoods (probability density) obtained from distances returned by word recognizer. These methods [13–15] perform handwriting recognition followed by an indexing step to keep track of the transcription and other useful information (positions and recognition scores of word images). The similarity between the keyword and another word image is computed using the recognition scores, which are usually the likelihood of the feature space, probabilities, or some other distance-based measurements. One question is whether to adopt a word lexicon. The index for fast retrieval can be built on the results of word level recognition in lexicon-driven mode [14,15]. In this mode, any word that is not in the lexicon cannot be retrieved. Ref. [13] performs recognition at the character level and searches for words in a series of character recognition scores. However, this approach is once again difficult and time-consuming which does not scale to larger data sets. We have taken a word-lexicon-driven method and get affected by the out-of-vocabulary (OOV) problem.

We have improved the OCR score-based indexing method by integrating word segmentation probabilities into the retrieval similarity metric. Word spotting methods this far have assumed perfect word segmentation: word images are given by word segmentation algorithm, and the ranks of word images are obtained by sorting the word recognition scores. However it is unrealistic to expect perfect word segmentation in unconstrained handwriting given the variation in the gap sizes between words. The performance of

* Corresponding author.
E-mail addresses: hcao@bbn.com (H. Cao), ab94@buffalo.edu (A. Bhardwaj), govind@buffalo.edu (V. Govindaraju).

0031-3203/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2009.02.003

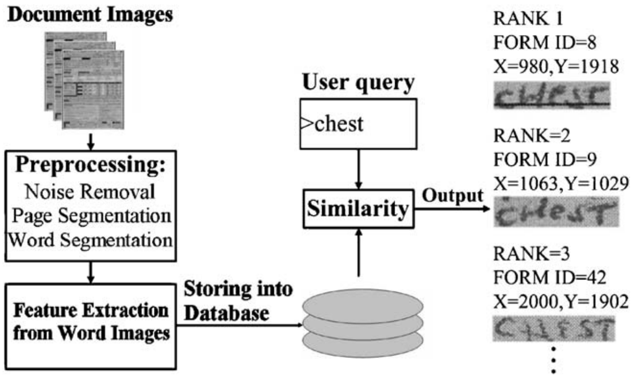

Fig. 1. Diagram of the keyword spotting system.

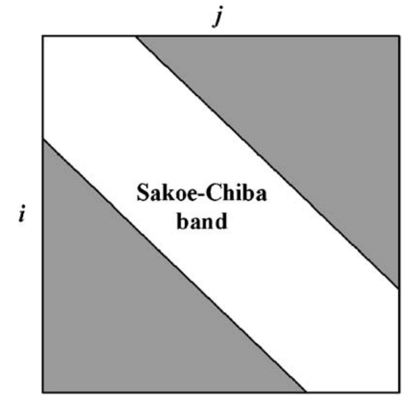

Fig. 2. Sakoe–Chiba band.

word spotting can be improved by modeling the word segmentation probabilities. In this paper, we describe a probabilistic model of word spotting that integrates word segmentation probabilities and word recognition probabilities. The word segmentation probabilities are obtained by modeling the conditional distribution of multivariate distance features of word gaps. The word recognition results are also represented by a probabilistic model. The modeling of the word recognition probabilities is obtained from the distances returned by the word recognizer (Fig. 1).

2. Background in handwritten keyword retrieval

2.1. Image-to-image matching—word spotting

Word spotting was initially proposed as an alternative approach for indexing and retrieving handwritten documents, that is one could search handwritten document images without using a handwriting recognizer. In order to search for a keyword, the user needs to write a copy of the keyword (a word template) and provide the word image as the query. One could also obtain the word templates by labeling a training set. The system executes the query by computing the distance between the query template and each word image in the document images.

DTW-based keyword spotting: In the DTW-based method [1,3,11], the following preprocessing steps are commonly used.

1. Word segmentation is performed and the background of every word image is cleaned by removing irrelevant connected components from other words that reach into the word's bounding box.
2. Inter-word variations such as skew and slant angle are detected and eliminated.
3. The bounding box of any word image is cropped so that it tightly encloses the word.
4. The baseline of word images is normalized to a fixed position by padding extra rows to the images.

A normalized word image is represented by a multivariate time series composed of features from each column of the word image. These features include projection profile, upper/lower word profile, and number of background-to-foreground transitions.

1. Projection profile. The projection profile of a word image is composed of the sum of foreground pixels in each column.
2. Upper/lower profiles. The upper profile of a word image is made of the distances from the upper boundary to the nearest foreground pixels in each column.
3. Background-to-foreground transitions. The number of background pixels whose right neighboring pixels are foreground pixels is taken as the number of background-to-foreground transitions of the column.

Suppose two word images w_A and w_B are represented by $\{f_A(1), f_A(2), \ldots, f_A(l_A)\}$ and $\{f_B(1), f_B(2), \ldots, f_B(l_B)\}$, respectively, where $f_A(i)$ is the feature vector of the i-th column of image w_A, $f_B(j)$ is the feature vector of the j-th column of image w_B, and l_A and l_B are the lengths of w_A, w_B, respectively. Then the DTW matching distance of w_A and w_B is given by the recurrence equation

$$DTW(i,j) = \min \left\{ \begin{array}{c} DTW(i-1,j) \\ DTW(i-1,j-1) \\ DTW(i,j-1) \end{array} \right\} + d(i,j) \quad (1)$$

where $d(i,j)$ is the square of the Euclidean distance between $f_A(i)$ and $f_B(j)$.

The time complexity of the DTW algorithm is in $O(l_A \cdot l_B)$. In order to reduce the computation and prevent pathological warping, a global path constraint like the Sakoe–Chiba band can be applied to force the paths to stay close to the diagonal of the DTW matrix. In Fig. 2, the dynamic programming range of (i,j) is restricted within a band along the diagonal of the (i,j) matrix which is called the Sakoe–Chiba band.

The matching error of $f_A(i)$ and $f_B(j)$ is given by $(1/l)DTW(l_A, l_B)$ where l is the length of the warping path recovered by DTW. The word images are ranked in the increasing order of the matching errors to the template image.

The DTW-based method has been tested on George Washington's manuscripts (CIIR, University of Massachusetts [1,11]). The performance of keyword spotting was evaluated using the mean average precision (MAP) measure [16]:

1. For each query, check the returned word images starting from rank 1. Whenever a relevant word image is found, keep track of the precision of the word images from the one with rank 1 to the current one. The average value of the recorded precisions for the query is taken as the average precision (AP) of the query.
2. The mean value of the AP of all of the queries is the MAP of the test.

A MAP of 40.98% on 2372 word images of good quality and a MAP of 16.50% on 3262 word images of poor quality was reported on George Washington's manuscripts [3].

GSC feature-based keyword spotting: In the GSC feature-based method [2,12], a word image is represented by 1024 bits of the GSC features corresponding to the gradient (192 bits), structural (192 bits) and concavity (128 bits) features. A word image is divided into 32 regions (8 × 4) and 16 binary GSC features are extracted from each region. The gradient features are obtained by thresholding the results of Sobel edge detection in the 12 directions. The structural