A Framework for Efficient Transcription of Historical Documents Using Keyword Spotting

Konstantinos Zagoris
Institute of Informatics and Telecommunications
National Center for Scientific Research 'Demokritos'
Athens, Greece
&
Visual Computing Group
Dept. of Electrical and Computer Engineering
Democritus University of Thrace
Xanthi, Greece
kzagoris@ee.duth.gr

Ioannis Pratikakis
Visual Computing Group
Dept. of Electrical and Computer Engineering
Democritus University of Thrace
Xanthi, Greece
ipratika@ee.duth.gr

Basilis Gatos
Institute of Informatics and Telecommunications
National Center for Scientific Research 'Demokritos'
Athens, Greece
bgat@iit.demokritos.gr

ABSTRACT

Keyword spotting (KWS) has drawn the attention of the research community as the alternative means to solve hard cases of handwriting text recognition. In this paper, a framework is proposed that employs KWS to enhance the efficiency in the manual transcription process, thus, reducing drastically the cost of training data creation. The core principle relies upon the ability of robust document-specific descriptors to produce meaningful similarities between a chosen word image for transcription and the corresponding word images in the full dataset under consideration. In the proposed framework, KWS is coupled with a relevance feedback mechanism which further enhances retrieval performance while being independent to the chosen KWS algorithm. The efficiency of the proposed pipeline is showcased via a user-friendly web-based prototype[1].

CCS Concepts

•**Applied computing** → **Document analysis;** *Document management and text processing;* •**Computing methodologies** → Visual content-based indexing and retrieval;

Keywords

Word Spotting,Handwritten Historical Documents, Relevance Feedback, Transcription

[1]http://vc.ee.duth.gr/ws/

HIP '15 August 22, 2015, Nancy, France

DOI: http://dx.doi.org/10.1145/2809544.2809557

1. INTRODUCTION

In digital libraries, many historical manuscripts are still unexploited due to the lack of proper browsing and indexing tools. For many typical handwritten document images, traditional Optical Character Recognition (OCR) is simply not usable since characters cannot be automatically segmented and recognized very easily. Therefore, holistic Handwriting Text Recognition (HTR) techniques are applied which do not require any explicit character segmentation. Current technology for HTR employs methods as Hidden Markov Models (HMMs) [16] and Neural Networks [23, 6]. Unfortunately, the aforementioned approaches need a considerable amount of training data. The common HMM-Based Handwriting Recognition Systems employ statistical language models which depend on the corresponding language, historical time period, etc. and they are very costly to create as they are requiring huge amount of manually transcribed data.

Ground-truth generation systems such as Aletheia [3] the transcription text is entered manually for each segmented word.

In this paper, a framework is proposed that employs keyword spotting (KWS) to enhance the efficiency in the manual transcription procedure thus, reducing drastically the cost of training data creation.

KWS can be defined as the task of identifying locations on a document image which have high probability to contain an instance of a queried word, without explicitly recognizing it. In this paper, we address the KWS query by example case (QBE) that aims to search a word image from a set of unindexed document images using the image content as the only information source. As final outcome, the system returns to the user a ranked list of document word images. The main advantage of KWS/QBE systems is that they perform word detection without any training data or any language model, and thus it makes them ideal as recommender system for helping the user transcribe the document.

The major achievement of the proposed framework is the reduction in time expenses required to achieve transcription data which could feed a Handwriting Text Recogni-

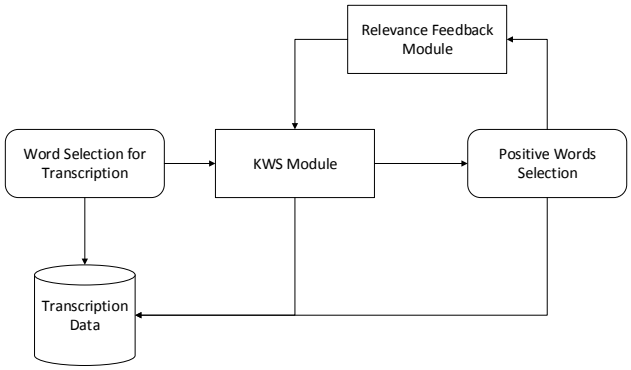

Figure 1: Flow diagram of the proposed transcription process

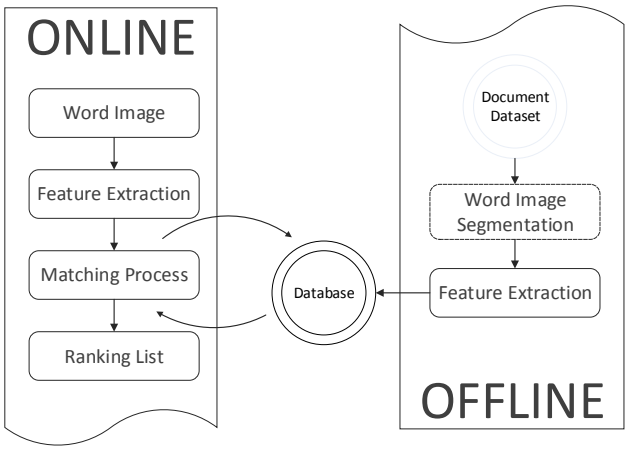

Figure 2: KeyWord Spotting Operational Pipeline

tion engine for training. Furthermore, the keyword spotting pipeline is coupled with a relevance feedback mechanism which introduces the user in the retrieval loop, thus, improving the final retrieval performance.

The remainder of the paper is organized as follows: Section 2 describes the architecture of the proposed transcription framework and its components, Section 3 presents the experimental results while in Section 4 conclusions are drawn.

2. PROPOSED FRAMEWORK

The proposed framework comprises two components that collaborate towards aiding the user to transcribe the document. These components are (a) the KWS module and (b) the Relevance Feedback module. A flow diagram of the proposed transcription process is shown in Figure 1. It is worth to note that a web-based prototype has been implemented[2] that showcases the proposed pipeline.

Initially, the results of a word segmentation process is presented to the user who selects a word to transcribe, thus, the KWS process is activated using the selected word as query resulting in ranked results which are presented to the user (Figure 3a).

Then, the user selects a positive word example from the ranked list and it is transferred to a verified list. The words that are contained in this verified list are linked, thus, having the same transcription text (Figure 3c). Simultaneously, a relevance feedback process is initiated that improves the KWS results.

Figure 3 shows the implemented interface for the above process and Figure 4 shows a snapshot of the corresponding transcription result.

2.1 KWS Module

The word spotting module is based upon a keyword spotting algorithm [26]. It is responsible for the retrieval of the visual similar words.

Although there is an abundance of systems suitable for both modern [7, 24] and historical [9, 12, 28, 10] printed material, very few of these systems are suitable to handwritten documents [13, 25, 24, 26, 14, 22] due to noise sensitivity, character variation and text layout complexity.

The keyword spotting in used [26] is chosen due to its suitability for handwritten historical documents. Despite that, the proposed architecture is suitable for any word spotting algorithm. This generates the ability to change the KWS method that fits better to a specific dataset.

A diagram of the chosen word spotting framework is illustrated in Figure 2. It consists of two distinct steps: the Offline and the Online. At the Offline step, which is executed once, the document images are segmented to the word images for which, document specific local features (DSLF) are extracted and indexed to a database. At the Online step, which is the only visible operation to the user, the DSLF are extracted for the query word image and a matching procedure is addressed between the features of the query and each indexed word image. Finally, a ranking list of all the word images are presented to the user.

2.2 Relevance Feedback Module

The relevance feedback is a post-query process that affects the retrieval results using user involvement in the selection of positive results. The object is to improve retrieval's performance by approximating user's criteria on the concept of similarity in a retrieval task, taking into account user's interaction.

Generally, relevance feedback schemes are divided into two distinct categories, those that aim to modify the initial query [18, 4, 11, 5, 1, 25] and those that mainly intend to alter the similarity measure handling the ranking of the results [20, 27].

The proposed relevance feedback method falls under the query modification category and it is transparent to the underlying KWS algorithm. The novelty introduced here is that instead of directly change the query descriptor, each user selection is executed as a query to the dataset and the results are fused and presented to the user. Figure 5 shows the architecture of the proposed relevance feedback algorithm.

The steps of the proposed relevance feedback are:

Step 1 The system presents an initial retrieved ranked list to the user.

Step 2 The user provides a positive judgement on the displayed results as to whether are relevant to the currently transcribed word by selecting one or more of positive words.

[2]http://vc.ee.duth.gr/ws/

Figure 3: The steps for the proposed transcription method (a) Word selection for transcription : 1. The transcription text. 2. The ranking list from the KWS method. 3. The verified list which contains words that correspond to positive user selected samples. The initial word is only the currently transcribed word. (b) Selecting a word from the ranking list, it is transferred to the verified list. (c) The user can continuously select similar words (d) until there are no other to select from the ranking list.

Step 3 The system queries the database for the more similar words to the user selection.

Step 4 The results are combined based on a fusion strategy and then, the newly created ranking list presented to the user.

In Step 4, multiple basic combination strategies [21, 8] are explored:

- **CombSum:** It combines the relevance listings by adding their corresponding relevance scores.
- **CombMin:** It combines the relevance listings by selecting the minimum relevance score.
- **Probabilistic model:** It combines the relevance listings by multiplying the corresponding relevance scores.

3. EXPERIMENTAL RESULTS

The dataset that is employed for the presented experiments is the *Bentham Dataset* [15]. This dataset is used in ICFHR 2014 Competition on Handwritten Keyword Spotting (H-KWS 2014) [19] and particularly for TRACK I. It

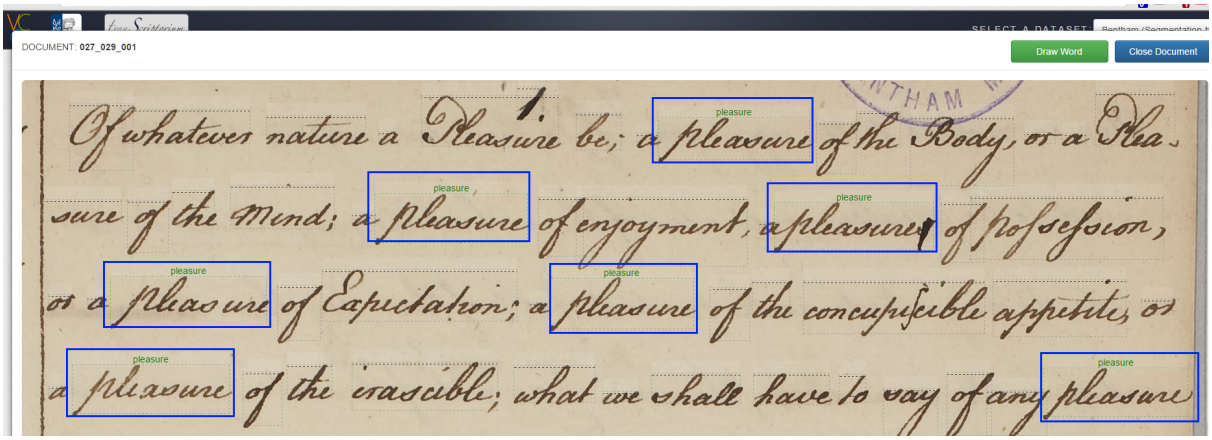

Figure 4: Results from the proposed automatic transcription of the word "pleasure" after the steps shown in Figure 3 (in blue rectangles)

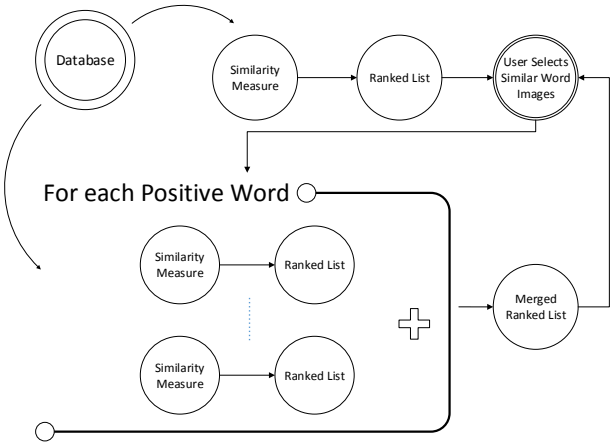

Figure 5: The architecture of the proposed relevance feedback process

consists of 50 high quality (approximately 3000 pixels width and 4000 pixels height) handwritten manuscripts.

Initially, the different fusion strategies are explored. In order to emulate the user involvement, three iterations are applied. For each iteration, two additional positive samples are chosen and the relevance feedback algorithm is applied. The performance evaluation is based on the following two measures:

(i) Precision at Top 5 Retrieved words (P@5)

(ii) Mean Average Precision (MAP)

In particular, Precision and P@k are defined as follows:

$$P@k = \frac{|\{relevant\ words\} \cap \{k\ retrieved\ words\}|}{|\{k\ retrieved\ words\}|} \quad (1)$$

Precision is the fraction of retrieved words that are relevant to the query, while in the case that precision should be determined for the k top retrieved words, P@k is computed. In particular, in the proposed evaluation, P@5 is used which is the precision at top 5 retrieved words. This metric defines how successfully the algorithms produce relevant results to the first 5 positions of the ranking list.

The second metric used is the Mean Average Precision (MAP) which is a typical measure for the performance of information retrieval systems [17, 2]. It is implemented from the Text REtrieval Conference (TREC) community by the National Institute of Standards and Technology (NIST). The above metric is defined as the average of the precision value obtained after each relevant word is retrieved:

$$AP = \frac{\sum_{k=1}^{n} (P@k \times rel(k))}{\{relevant\ words\}} \quad (2)$$

where:

$$rel(k) = \begin{cases} 1, & if\ word\ at\ rank\ k\ is\ relevant \\ 0, & if\ word\ at\ rank\ k\ is\ not\ relevant \end{cases} \quad (3)$$

As shown at Figure 6, at each iteration an improved retrieval performance is achieved for all the fusion strategies. The best performance is achieved with the combMin.

Next, in order to measure the speed up of the transcription time for our proposed architecture, we implement three different transcription strategies for the above historical handwritten dataset:

- **The Conventional Method** by using the Aletheia tool [3]
- **Segmentation guided Conventional Method** by using the implemented web-based prototype without the KWS and its corresponding Relevance Feedback method. This speeds up the transcription procedure as the words are already detected
- **The Proposed Architecture** using the web-based prototype as appears in http://vc.ee.duth.gr/ws/.

Table 1 shows the experimental results that demonstrate time costs improvement in transcription process. Particularly, 80% improvements over the conventional transcription and 55% over the segmentation-based approach.

4. CONCLUSIONS

Table 1: Transcription Time Expenses

	Total Time (hours)	Minutes per Document (m/d)	Seconds per Word (s/w)
Conventional Transcription	116.1	69.66	21.32
Segmentation guided Conventional Transcription	52.1	31.26	9.55
Proposed Transcription Framework	23.2	13.92	4.21

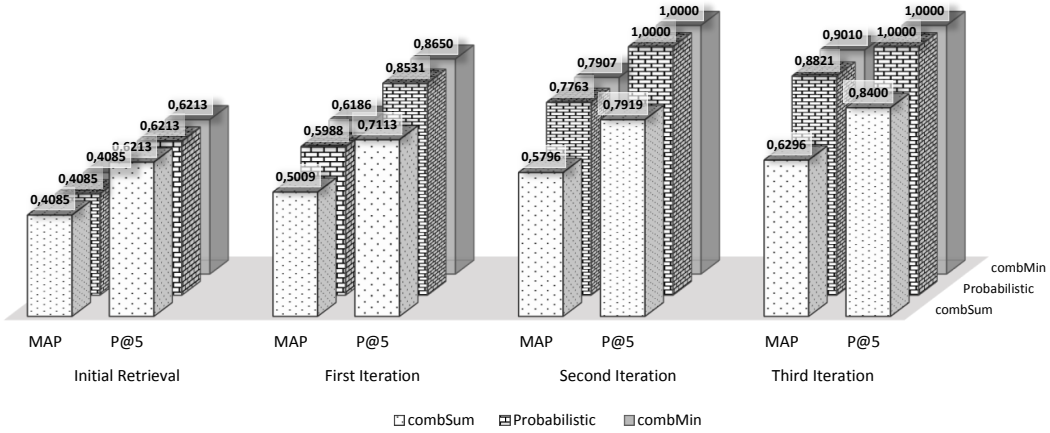

Figure 6: Performance for each relevance feedback fusion strategy

In this paper, a framework is proposed that employs word spotting algorithm in conjunction with a relevance feedback algorithm in order to speed up the manual transcription process and consequently reduce the creation cost of the training data. The experimental results in a publicly available dataset show the improved performance.

5. ACKNOWLEDGEMENT

Work supported by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 600707 - tranScriptorium.

6. REFERENCES

[1] S. Chatzichristofis, K. Zagoris, Y. Boutalis, and N. Papamarkos. Accurate image retrieval based on compact composite descriptors and relevance feedback information. *International Journal of Pattern Recognition and Artificial Intelligence*, 24(2):207–244, 2010.

[2] S. A. Chatzichristofis, K. Zagoris, and A. Arampatzis. The trec files: the (ground) truth is out there. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information*, SIGIR '11, pages 1289–1290, New York, NY, USA, 2011. ACM.

[3] C. Clausner, S. Pletschacher, and A. Antonacopoulos. Aletheia-an advanced document layout and text ground-truthing system for production environments. In *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, pages 48–52. IEEE, 2011.

[4] A. Doulamis, N. Doulamis, and T. Varvarigou. Efficient content-based image retrieval using fuzzy organization and optimal relevance feedback. *International Journal of Image and Graphics*, 3(1):171–208, 2003.

[5] J. French, X. Jin, and W. Martin. An empirical investigation of the scalability of a multiple viewpoint cbir system. In *Image and Video Retrieval*, volume 3115 of *Lecture Notes in Computer Science*, pages 252–260. Springer Berlin Heidelberg, 2004.

[6] A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, and J. Schmidhuber. A novel connectionist system for unconstrained handwriting recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(5):855–868, 2009.

[7] E. Hassan, S. Chaudhury, and M. Gopal. Word shape descriptor-based document image indexing: a new dbh-based approach. *International Journal on Document Analysis and Recognition (IJDAR)*, 16(3):227–246, 2013.

[8] D. F. Hsu and I. Taksa. Comparing rank and score combination methods for data fusion in information retrieval. *Information Retrieval*, 8(3):449–480, 2005.

[9] A. L. Kesidis, E. Galiotou, B. Gatos, and I. Pratikakis. A word spotting framework for historical machine-printed documents. *IJDAR*, 14(2):131–144, 2011.

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

[11] D. Kim and C. Chung. Qcluster: relevance feedback using adaptive clustering for content-based image retrieval. In *Proceedings of the 2003 ACM SIGMOD international conference on Management of data*, pages 599–610, 2003.

[12] T. Konidaris, B. Gatos, S. Perantonis, and A. Kesidis. Keyword matching in historical machine-printed documents using synthetic data, word portions and dynamic time warping. In *Document Analysis Systems, 2008. DAS '08. The Eighth IAPR International Workshop on*, pages 539–545, 2008.

[13] V. Lavrenko, T. M. Rath, and R. Manmatha. Holistic word recognition for handwritten historical documents. In *Document Image Analysis for Libraries, 2004. Proc. 1st International Workshop on*, pages 278–287.

[14] J. Llados, M. Rusinol, A. Fornes, D. Fernandez, and A. Dutta. On the influence of word representations for handwritten word spotting in historical documents. *International Journal of Pattern Recognition and Artificial Intelligence*, 26(05):1263002, 2012.

[15] D. G. Long et al. *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*. The College, 1981.

[16] U.-V. Marti and H. Bunke. 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):65–90, 2001.

[17] T. NIST. http://trec.nist.gov/pubs/trec16/appendices/measures.pdf, 2013.

[18] K. Porkaew and K. Chakrabarti. Query refinement for multimedia similarity retrieval in mars. In *Proc. of the seventh ACM international conference on Multimedia (Part 1)*, pages 235–238, 1999.

[19] I. Pratikakis, K. Zagoris, B. Gatos, G. Louloudis, and N. Stamatopoulos. Icfhr 2014 competition on handwritten keyword spotting (h-kws 2014). In *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*, pages 814–819, Sept 2014.

[20] Y. Rui and T. Huang. Optimizing learning in image retrieval. In *Proceedings of Computer Vision and Pattern Recognition*, pages 236–243, 2000.

[21] J. A. Shaw, E. A. Fox, J. A. Shaw, and E. A. Fox. Combination of multiple searches. In *The Second Text REtrieval Conference (TREC-2*, pages 243–252, 1994.

[22] R. Shekhar and C. Jawahar. Word image retrieval using bag of visual words. In *DAS 2012*, pages 297–301, March 2012.

[23] A. H. Toselli, A. Juan, J. González, I. Salvador, E. Vidal, F. Casacuberta, D. Keysers, and H. Ney. Integrated handwriting recognition and interpretation using finite-state models. *International Journal of Pattern Recognition and Artificial Intelligence*, 18(04):519–539, 2004.

[24] K. Zagoris, K. Ergina, and N. Papamarkos. A document image retrieval system. *Engineering Applications of Artificial Intelligence*, 23(6):872 – 879, 2010.

[25] K. Zagoris, K. Ergina, and N. Papamarkos. Image retrieval systems based on compact shape descriptor and relevance feedback information. *Journal of Visual Communication and Image Representation*, 22(5):378 – 390, 2011.

[26] K. Zagoris, I. Pratikakis, and B. Gatos. Segmentation-based historical handwritten word spotting using document-specific local features. In *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*, pages 9–14, Sept 2014.

[27] X. Zhou and T. Huang. Comparing discriminating transformations and svm for learning during multimedia retrieval. In *Proc. of the ninth ACM international conference on Multimedia*, pages 137–146, 2001.

[28] F. Zirari, A. Ennaji, S. Nicolas, and D. Mammass. A methodology to spot words in historical arabic documents. In *Computer Systems and Applications (AICCSA), 2013 ACS International Conference on*, pages 1–4, 2013.