SEGMENTATION-BASED WORDSPOTTING IN HANDWRITTEN DOCUMENTS

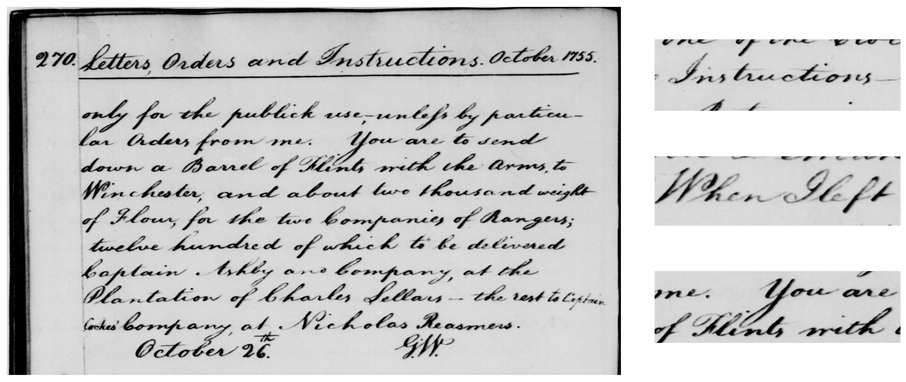
Eugen Rusakov and Gernot A. Fink
(in collaboration with Leonard Rothacker and Hyunho Mo [1])
Department of Computer Science, TU Dortmund University, Germany

technische universität dortmund

Abstract

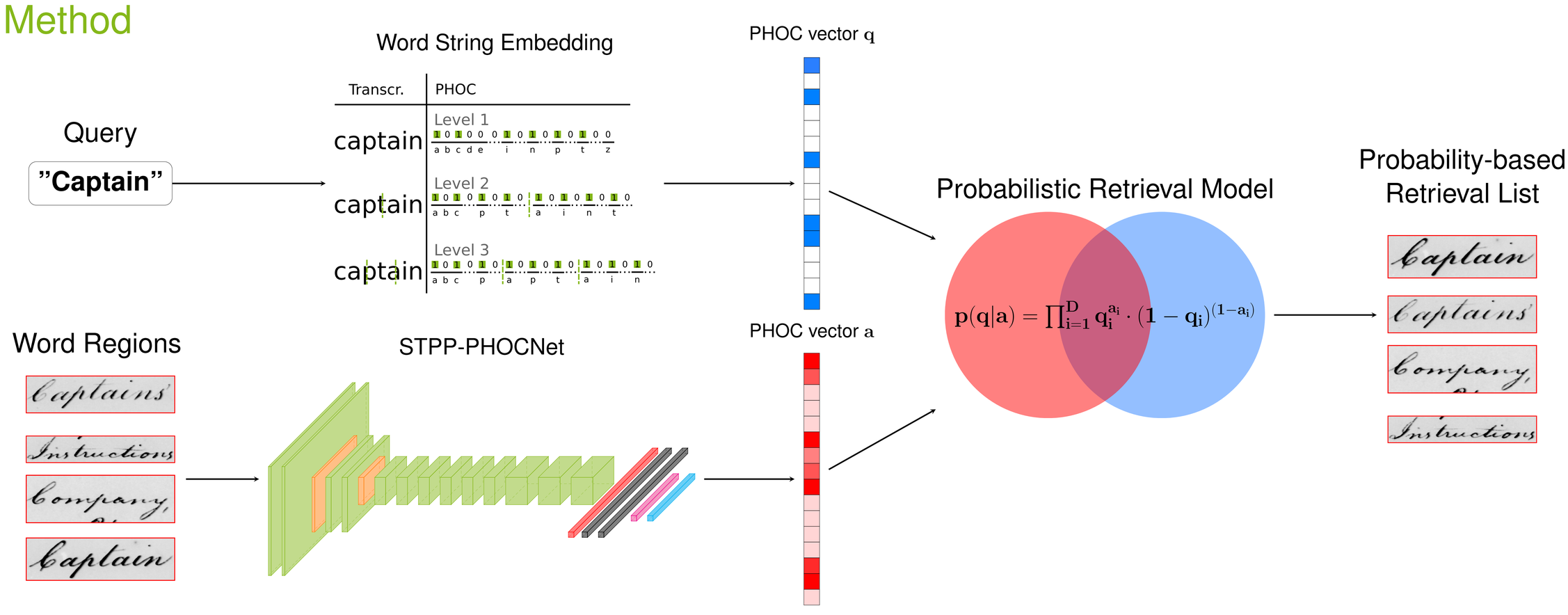
Given a user defined query, the task of word spotting is to retrieve a list containing word images that are relevant with respect to the query. Typically word spotting methods rank all retrieved word images from a given document collection by a certain criterion and sort them by their similarities. These queries can either be word images, defined by a user cropping a snipped from a document page or defining a word string which needs to be retrieved. As dictionaries can contain many thousands of words, present-day word recognition methods require attribute-based classification approaches. Hence, Almazan et al. [2] proposed a word string embedding called *Pyramidal Histogram of Characters (PHOC)*, representing the character occurences as binary attributes. Retrieval is performed by comparing the predicted PHOC vectors using a certain distance measure, for example the cosine similarity. In this work [1], the cosine similarity is replaced by a probabilistic comparison of similarities. Lampert et al. [5] proposed a method for a similarity measure called *Direct Attribute Prediction (DAP)*, comparing the posteriors instead of distances.

Challenges in Handwritten Documents



The example above shows the George Washington data set consists of 20 handwritten document pages. The main challenge in text recognition is the amount of word instances (classes). Thus, a attribute-based classification approach is required to classify many thousands of words. Especially in historical documents, handwritten words can contain faded ink resulting in characters which are not fully visible. As the typography of a single writer can contain different lengths for ascenders and descenders within the same text, different words can overlap.

Method



Future Research

Region Proposal Network

Attribute-based Segmentation

The future research is to move from a segmentation-based to the segmentation-free word spotting scenario. In contrast to segmentation-based word spotting, the words from a document page needs to be detected. Here, the challange is to incorporate the knowledge of words (What is a word?) into the detection methods. A possible approach could be the region proposal networks, proposing regions based on the presumption of the visual apperance of words. Nevertheless, this approach lacks of knowledge about words. Another approach could be the semantic segmentation based on attribute prediction. Instead of predicting a single class per pixel, a attribute vector could be predicted representing the word. The segmentation approach could be combined with a sequential model like conditional random fields, hidden markov models, or recurrent neural networks.

Results

Architecture/Method	George Washington QbE	George Washington QbS	IAM QbE	IAM QbS
STPP-PHOCNet (BCE) [1]	97.47	96.50	88.49	93.03
STPP-PHOCNet (PRM) [1]	97.76	96.89	**89.27**	**95.40**
TPP-PHOCNet (BCE) [3]	97.90	96.73	84.80	92.97
TPP-PHOCNet (CPS) [3]	97.96	**97.92**	82.74	93.42
PHOCNet (BCE) [3]	97.58	95.58	85.50	92.38
PHOCNet (CPS) [3]	97.72	97.44	75.85	91.12
Triplet-CNN [4]	**98.00**	93.69	81.58	89.49

The results are shown on two benchmarks using the *mean Average Precision (mAP)* as evaluation metric for the Query-by-Example (QbE) and the Query-by-String (QbS) scenario. The method proposed in this work achieves state-of-the-art performance on the larger IAM data set. On the George Washington data set all PHOCNet configurations performs nearly equal. As this data set is relative small and all recently proposed methods performs close to 100%, there is not much space for improvement.

References

[1] E. Rusakov, L. Rothacker, Hyunho Mo and G. A. Fink, "A Probabilistic Retrieval Model for Word Spotting based on Direct Attribute Prediction", in *ICFHR*, 2018.

[2] J. Almazán, A. Gordo, A. Fornés, and E. Valveny, "Word Spotting and Recognition with Embedded Attributes", in *TPAMI*, vol. 36, no. 12, pp. 2552-2566, 2014.

[3] S. Sudholt and G. A. Fink, "Attribute CNNs for Word Spotting in Handwritten Documents", in *IJDAR*, 2018, to appear.

[4] T. Wilkinson and A. Brun "Semantic and Verbatim Word Spotting using Deep Neural Networks", in *ICFHR*, 2016.

[5] C. H. Lampert, H. Nickisch, and S. Harmeling, "Attribute-based Classification for Zero-Shot Visual Object Catigorization", in *TPAMI*, vol. 36, no. 3, pp. 453-465, 2014.