https://doi.org/10.1142/9789811203244_0013

Chapter 13

Conclusions and Future Trends

Andreas Fischer, Marcus Liwicki, and Rolf Ingold

Automatic text extraction from document images has been an important research area for nearly half a century. While printed text recognition has reached a level of maturity that allows commercial application for office automation, automatic transcription of handwritten documents remains a major challenge. The additional difficulty handwriting recognition is facing in comparison with printed text recognition can be explained by several reasons. An evident factor is the extreme variability of writing styles related to multiple socio-educative and cultural aspects. Additionally to these interpersonal variabilities, one can also observe significant intrapersonal variabilities related to psychological and aging factors. However, the latter being less important, writer models specifically trained for a given writer, can lead to acceptable transcription results.

Another characterization of handwriting is its cursive nature, which impacts the recognition process in two ways. First, the letter shapes are strongly influenced by the previous and next characters, thus increasing once more the variability. And second, character shape boundaries are hard to be determined; or stated differently, the word segmentation into characters is a non-trivial task reflected by the so-called Sayre's paradox: a character sequence cannot be recognized without being segmented first, and reversely, it cannot be segmented without previous character recognition.

Fortunately, Sayre's paradox can be circumvented by combining the segmentation and recognition tasks. Such approaches can be efficiently achieved by machine learning techniques such as hidden Markov models or recurrent neural networks, which both can integrate character models characterizing

shapes with language models carrying contextual information, either in form of n-grams or of a complete dictionary.

Despite such strategies, universal handwriting recognition is far from being solved. Acceptable solutions can be provided for some restricted applications. This holds for instance for digit or isolated letter recognition or when the used vocabulary is reduced, such as in addresses recognition used for automated mail dispatching.

Another favorable situation occurs when systems can be trained for a specific writer. However, this requires enough training data that has previously been labelled. Fortunately, current modeling techniques do not require a transcription to be aligned at character level; ground-truth alignment at word level or even line level is sufficient. Nevertheless, the practical difficulty to get accurate transcriptions should not be underestimated, notably in case of misspelling or abbreviations that are often corrected or completed by transcribers and therefore do not match the real data. In any case, to build a strong writer model several dozen pages are usually needed and this is often a major hindrance for such recognition applications.

The research community is strongly dealing with all these issues. Regular progresses are reported, not only in terms of higher accuracy, but also by improved learning strategies reducing the size of the training data. A lot of effort is also put into interactive annotation tools, in order to facilitate manual transcription and alignment. Crowdsourcing can be used as a profitable alternative.

As long as automatic transcription is not available, mainly for indexing purposes, keyword spotting can be used as an interesting alternative. Several digital libraries provide search functionalities based on such approaches using as queries either text or sample images.

To further progress in historical document analysis, handwriting recognition is not the only obstacle; several complementary tasks have also to be improved. Complex layout analysis in case of ornaments or glosses is one of them; but we must also mention script classification or scribe authentication. All these techniques must rely on best possible image quality, obtained by adequate image pre-processing and filtering.

In a mid and long term perspective we can expect substantial progress in handwriting recognition. In this perspective, the key element is certainly to consider additional contextual information, including semantic knowledge.

Considering the recent progresses made in natural language processing and machine translation we can anticipate that the integration of the same kind of knowledge with language models will effectively improve text recognition in the future. Combined with the still massive increase of computing power, we can expect a real breakthrough in this domain in the next 5 to 10 years.