Beyond OCR: Handwritten manuscript attribute understanding

He, Sheng

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2017

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
This thesis studies different problems for handwritten manuscript understanding, including writer identification, script recognition, historical document dating and localization. By analyzing different types of features for these four different applications, several conclusions can be obtained as follows.

First, most existing methods are a special case of a general method which may not have been found yet. In fact, science is directed at generalizing observed phenomena to a more general principle, which (therefore) can solve unobserved problems. This thesis provides several examples. The $\Delta^n$Hinge, CoHinge and QuadHinge features are the extension of the Hinge feature and they are not only more powerful than Hinge, but also contain more useful properties, such as the rotation invariance. The general principle here is to make the co-occurrence or joint distribution of patterns explicit in different spaces, such as spatial, attribute or kernel joint distributions. Experimental results in this thesis show that feature methods following the joint-feature distribution principle generally provide improved performance in different applications of understanding handwritten manuscripts.

Second, solving different problems on the same material usually requires different types of features or methods from the pattern recognition field. For example, when considering the writer-identification problem, the system should find the similar writing style in different documents from the same writer. However, when considering the dating problem, the system should find the similar writing style of documents from many different writers in the same period. When a large data set is available, machine learning could handle this problem. However, many methods in machine learning require the estimation of a huge number of parameters (weights). Considering historical collections which usually contain limited number of documents and labels, designing features with different properties is a promising way to solve this problem in a sparse-parametric manner.

Third, providing a rich set of options for end users is always necessary. End users expect to analyse the data in different perspectives, which requires various types of methods. In this thesis, we propose different types of textural-based and grapheme-based features with different properties. $\Delta^n$Hinge is the curvature-based feature which is rotation-invariant. LBPruns and COLD features are curvature-free features which achieve good results on documents where roundness or angularity is not the most descriptive shape characteristic. Juncelets are the basic elements in all handwritten documents with different scripts because junctions and crossings cannot be avoided by a writer. Contour and stroke fragments, on the other hand,
are more powerful for the script recognition and the dating problem. H_{2}OS (Histogram of Orientations of Handwritten Strokes) is more robust than Junclets in historical documents with serious image degradations.

Fourth, any feature extraction (including hand-crafted and transfer-learned features) introduces errors into the recognition process. The difference between the hand-crafted features and learned features is that the errors introduced by hand-crafted features are class-independent or application-independent. Therefore, they are explainable and can be used for any applications. However, learned features always introduce the class-dependent errors during the training, if a supervised training procedure with labels was applied. It is hard to explain what has been learned and applicability to problems that consist of other classes is uncertain. For instance, if a large convolutional network is trained on a large medieval text corpus, how will it perform on contemporary handwriting or another script? Each hand-crafted feature proposed in this thesis is explainable to end users, even to scholars or users who study history.

8.1 Answers to the research questions

Several research questions are proposed in Chapter 1. In this section, we briefly recall them and provide some answers.

Q1: How to design rotation-invariant features for writer identification?

Many features for writer identification are sensitive to rotation changes because they build the feature vector based on some reference directions. For example, the angles in the Hinge kernel are computed based on the horizontal direction, making the final feature vector variant to rotation. One way to achieve the rotation-invariant property is to compute the relative angles, instead of the absolute values, to build the feature vector. The distribution of the relative angles along ink contours can also provide writer-specific information, and combining it with the absolute angles of the Hinge kernel can improve the performance of writer identification.

Q2: How to design efficient grapheme features for writer identification?

Q3: How to perform cross-script writer identification, such as between English and Chinese?

Instead of focusing on the handwritten trace itself, also layout characteristics such as the space between letters/words or the size of the empty area insider a letter/loop is important for writer identification. Such phenomena can be captured by the run length of simple patterns, such as white pixels or background pixels in multiple strands. The run-length of more complicated patterns formed by several scan lines can provide additional discriminative information.
Traditionally, grapheme-based features need character segmentation, which is very challenging on handwritten images because it presupposes a correct “OCR” or manual letter labeling. The junction feature proposed in this thesis does not require any segmentation and is computationally efficient. The performance of the junclets method for cross-script writer identification between Chinese and English is much better than other textural-based features. Another observation that has been found is that English texts written by Chinese people on the CERUG data set contains a large number of long lines (linear pieces of ink trace). The possible reason is that Chinese writers tend to write line strokes affected by the habit of writing Chinese characters which typically consist of linear strokes. Generally, texts wrote by less skilled writers contain more irregular-curvature strokes because the velocity profile of on-line recorded handwriting of non-native speakers shows pauses, as well as a degree of polygonisation [Meulenbroek and Van Galen 1988]. If we approximate the writing contours by a set of line segments, the joint distribution of the orientation and length of these line segments can also characterize the writing style.

Q4: How to design an efficient system to automatically date and localize historical manuscripts based on the handwriting style?

The historical document dating and geographical localization for end users, such as historians or paleographers, requires not only a high performance, but also to be easily visualized. Therefore, grapheme-based features are more promising for this problem. In this thesis, three grapheme-based features (mid-level handwritten patterns) are proposed, such as contour fragments (kCF), stroke fragments (kSF) and H2OS. These features are very easy to compute and contain more information about general handwriting style of handwritten documents wrote in a certain period or location. Combining the kCF and kSF provides much better results than their individual performance. H2OS is similar with the Junction feature, but more robust to images with low-quality. In addition, encoding these mid-level handwritten patterns into a codebook which contains label information can improve the performance.

Q5: What is the general rule or principle to design new features or increase the discriminative of existed features?

Following the joint-feature distribution principle, more discriminative features can be generated based on existing textural-based features, such as the Hinge [Bulacu and Schomaker 2007]. The spatial joint-feature distribution can describe the spatial relationship between different positions and capture more complex local structures with a larger supporting region. Features followed spatial joint-feature distribution usually have a large supporting region. Features followed the attribute joint-feature distribution usually have specific meanings, because they are joint distributions of different properties. When using kernel functions between features with different positions or attributes, the feature has specific properties or is invariant to rotation or scale changes. Recursively using these three principles with proper local features and kernel functions can result in more discriminative and abstract features.
8. Discussion

Figure 8.1: Edge maps computed in different methods. The left image is the input gray-scale image, the middle one is the edge map computed by the method proposed in (Dollár and Zitnick, 2013), and the right image is the edge map computed by a very simple method similar as (Jevnisek and Avidan, 2016).

8.2 Future research

In this section, several remarks, which are related to this thesis and sketch a number of future directions, are discussed as following:

R1: Most features proposed in this thesis depend on binarized images, which is a very challenging problem on handwritten images with a high degradation. How to extract them directly on gray-scale or color images?

One possible solution is to extract features based on ink contours or edges directly from the gray-scale image instead of the binarized image, similar as the edge-Hinge (Bulacu and Schomaker, 2003). Some learning techniques, such as the methods proposed in (Arbelaez et al., 2011; Jevnisek and Avidan, 2016) which detect contours on natural images based on multiple local cues, can be used in handwritten documents. Fig. 8.1 shows an example of the edge maps. The Hinge and proposed Hinge-based features can be extracted on these edge maps with salient stroke text boundaries. The key points detection for the junction feature proposed in Chapter 4 and the H$_2$OS feature proposed in Chapter 6 can be found as the middle point of the line computed by the method proposed in (Epstein et al., 2010), which computes the stroke width based on the line where the gradient directions of two end points much be roughly opposite.

R2: As we mentioned in Chapter 1, the text-dependent method is easy to be visualized for end users. How to date and localize geographically historical documents using text-dependent methods?

Using the Monk system, characters are segmented automatically and labeled by paleographers, including the label of the year. Fig. 8.2 shows the character models from different key years. From the table we can see that not all characters are suitable for dating. For
8.2. Future research

example, characters ‘i’ and ‘j’ contain less stylistic information over time and they are very hard to segment from cursive handwriting. The most informative characters are ‘a’, ‘d’, ‘g’, ‘t’, which are also very easy to segment. One direction for future work is to develop a segmentation method to properly segment individual characters from historical documents with cursive handwriting and model the evolution of handwriting of the same character in different years using some regression or deep learning methods. Fig. 8.3 shows an example of evolution of junction structures in different key years in the MPS data set. From Fig. 8.2 and Fig. 8.3 we can see that the same handwritten pattern can look different across the 11 key years, but such differences might be very subtle. Therefore, other machine learning methods, such as fine-grained recognition method (Deng et al., 2013), need to be developed for

Figure 8.2: Character models of different key years in the MPS data set.

Figure 8.3: The most similar junction structures in different key years in the MPS data set.
Figure 8.4: Junction structures in different languages (from top to bottom): Arabic, Chinese, English written by Chinese writers, English, Bangla. (All the junctions are normalized into a fixed size for better visualization.)

determining the year of each character.

R3: Script is another important attribute of handwritten documents because most OCR methods are script specific and the script should be recognized before applying OCR. Can the proposed methods be used for script identification?

The junction detector proposed in Chapter 4 can handle this problem. The distribution of junctions, which are the basic elements in handwritten documents, are distinct among different languages. The distinctness are from several aspects. Firstly, the percentages of different junction types are distinct among different languages. For example, the percentages of L- and Y-junctions in Chinese are approximately equal. However, the percentage of L-junctions is much higher than L- and X-junctions in Arabic. In addition, Chinese contains more X-junctions than others and Arabic contains more L-junctions. It is well known that Chinese is a logographic writing system and Chinese characters contain lots of crosses. On the contrary, Arabic is an Abjad writing system and Arabic characters is formed a long main stroke (Ghosh et al., 2010). There are top bars and vertical lines in Bangal script, resulting in more Y-junctions than the Latin-based languages. These properties are reflected on the corresponding handwritings and make the percentages of junction types different. Secondly, even for the same type junction, their aperture angle distributions are distinct. The aperture angle is defined as the minimum angle formed by the two legs of L-junctions. For Chinese and English, the aperture angle for most L-junctions is around \([80^\circ, 90^\circ]\). However, for the Arabic and Bangal, the aperture angle for most L-junctions is around \([140^\circ, 160^\circ]\). There are two peaks for the Latin language groups around 90° and 110°.

In addition, the proposed kCF can also be used for script identification, because the character shapes take an important role for script identification (Zhu et al., 2009). In practice,
we have found that $k$CF achieves the best performance among the features proposed in this thesis for script identification on four major scripts: Western script, Arabic, Chinese and Greek.

R4 Recently, it has been shown that deep learning provides much better results than traditional handcraft features with shallow machine learning methods. How to use deep learning for attribute understanding of handwritten documents?

Deep neural networks are successful when trained on a large data set. However, deep learning is still a black box and it is hard to explain results for end users, who may be not only interested in results, but also reasons. Future works include using transfer-learning methods and making the deep neural network explainable. When the scale of the data set is too small to train a deep neural network, transfer learning methods are always used to first pre-train the deep neural network on other large-scale data sets and then fine-train it on the studied data set. Another direction is to mathematically explain the learned neurons. For example, each neuron corresponds to a filter, which can be linearly represented by a set of Gabor or other types of basic filters.