Handwriting Recognition (HR) of Family History Documents using a 2-D Warping-based Word-level HR Approach

Douglas J. Kennard
Brigham Young University
Provo, Utah, USA
kennard@cs.byu.edu

William A. Barrett
Brigham Young University
Provo, Utah, USA
barrett@cs.byu.edu

Thomas W. Sederberg
Brigham Young University
Provo, Utah, USA
tom@cs.byu.edu

ABSTRACT

An enormous amount of handwritten information exists that is potentially very useful for family history research. However, finding information of interest is a daunting task unless the handwriting is transcribed or indexed so that it can be digitally searched. Transcription / indexing is typically done manually because automatic handwriting recognition (HR) is not yet accurate enough to provide reliable transcriptions. Since manual transcription is both costly and time consuming, improvements in HR are very desirable.

In this paper, we describe a novel method of word-level HR that we recently published at the International Conference on Document Analysis and Recognition (ICDAR 2011) and discuss how it can be applied to family history document images. We use an automatic morphing algorithm to generate a 2-D geometric warp that aligns each unknown word to known training examples. Once the word strokes are aligned, a distance map is used to calculate how different the aligned (warped) word is from the training example. The label of the training example that is most similar is used as the digital transcription for the previously unknown word. Our initial results are based on two datasets, each consisting of 1,000 training words and 1,000 test words. For in-vocabulary words, we get 88.77% and 89.33% word recognition accuracy, respectively.

1. INTRODUCTION

Handwritten documents comprise a large portion of the historical records that are of interest to people who research family history and genealogy. Churches, large companies, government archives, and other organizations are undertaking ambitious efforts to scan historical records into digital images and make them accessible. In order to make it more practical for people to find information in the records, the digitized images must be indexed or transcribed so that they can be digitally searched. In addition to providing the ability to perform searches, digital indexes and transcriptions are also necessary to enable new technologies that automatically discover, link, and bring together information from many different sources in a usable manner.

Family History Technology Workshop / RootsTech'12, Feb 2012, SLC, UT

Transcription and indexing are currently done manually because handwriting recognition (HR) is not accurate enough yet for automatic transcription. Manual transcription is very costly, requiring a great deal of time and effort for any sizable project. This limits how fast records can be indexed or transcribed, and also limits how much of the available information can even be considered for transcription or indexing. Often, transcription and indexing projects are prioritized based on which records will be of use to the most people or based on which are more easily indexed (structured forms like census, birth, and death records will usually be done before unstructured free-form handwriting). Other records may also be of significant informational value, but must wait (even indefinitely) just as a matter of practicality. Improvements in HR that help reduce the cost and time required to transcribe / index records would allow more information to be made available (and sooner).

We recently published a novel approach to HR [2] at the International Conference on Document Analysis and Recognition (ICDAR 2011). Our initial results are encouraging, and with additional work that we are currently doing, we believe this approach will be useful in several ways for processing family history documents. In Section 2 of this paper, we describe our approach at a high level. More detail is available in the ICDAR paper[1]. In Sections 3 and 4, we report our initial experiments and the results of those experiments. We then discuss how our HR method might be applied to family history document images in Section 5.

2. METHODS

Instead of trying to recognize individual letters (or even smaller sub-word pieces), our HR approach performs recognition at the whole-word level. For each unknown word that needs to be recognized, we compare it to *training examples* — word images that have been pre-labeled with their corresponding digital transcriptions. The label of whichever training example the unknown word is most similar to is then used as the recognized label for the unknown word.

In Figure 1, we illustrate the main concepts involved in our method of comparing an unknown word to known training examples in order to decide which example is most similar. When comparing any two given words, we compute a numerical *matching cost*. A lower cost means the words are more similar and a higher cost means they are more different.

[1] http://dx.doi.org/10.1109/ICDAR.2011.271

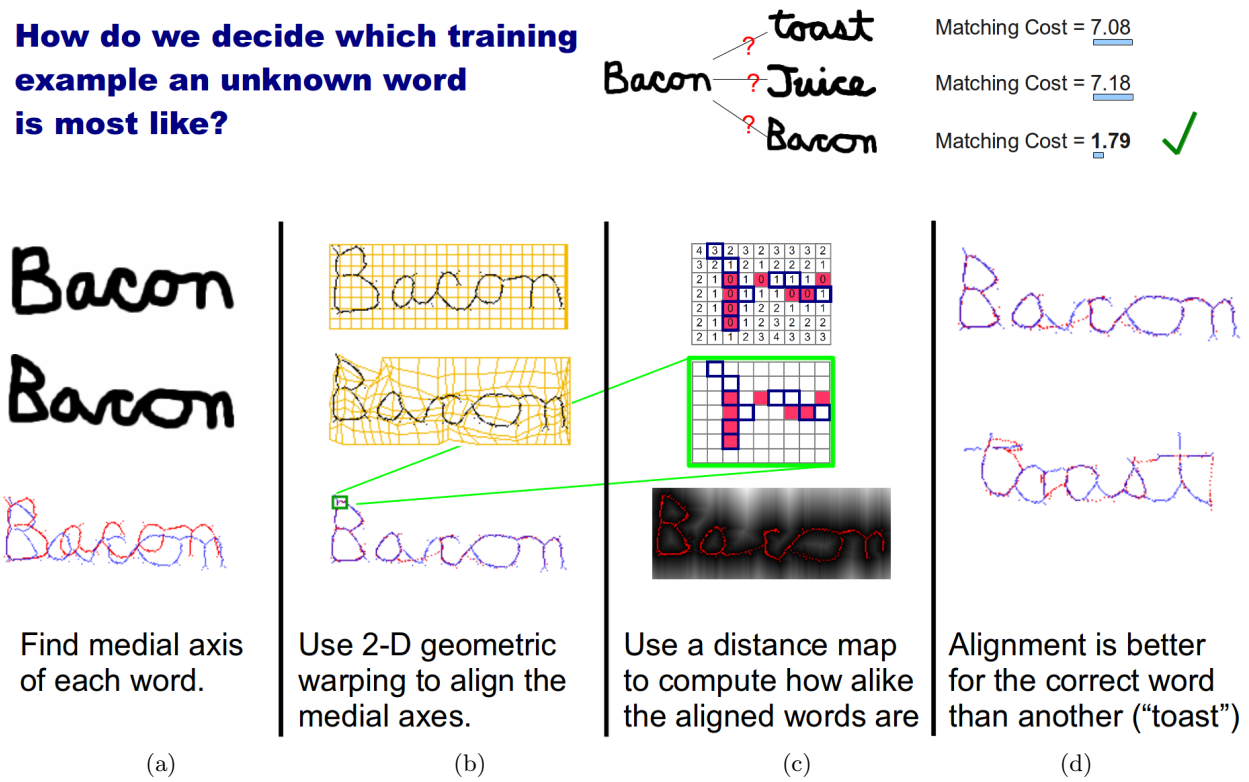

Figure 1: Overview of our recognition method. a) The medial axis (ink center) of the top word is shown in red, the bottom word in blue. b) The automatically-computed warp meshes (top) are used to warp the red medial axis so it aligns with the blue (bottom). c) Distance map values are the number of non-diagonal steps from a red pixel (top); full distance map (bottom) shows larger values as lighter intensity. d) Aggregate distance is smaller for similar words because they align better (top) than incorrect words (bottom).

To compute the cost, we use the *medial axis* pixels — those in the middle of the ink (Figure 1a). This simplifies the algorithm and also reduces the influence of different stroke thicknesses. We compute a 2-D geometric warp to align one medial axis to the other (Figure 1b). To do this automatically, we use an algorithm that we adapted from the work minimization approach to image morphing developed by Gao and Sederberg in [1]. We then compute the matching cost using values from a *distance map*— a map of how far any given position in the image is from the nearest medial axis pixel (Figure 1c). The cost function represents how far the pixels of one medial axis are from the pixels of the other. Words that are similar align better than words that are different, resulting in lower costs calculated from the distance map (Figure 1d).

Since the calculated cost may be different when warping one word to a second word than the cost when warping the second word to the first, we add the costs of warping each direction to get a symmetric result as the final matching cost for a given pair of words.

3. EXPERIMENTS

In our initial experiments, we use two datasets of labeled word images. The first dataset consists of words from a set of 20 pages of George Washington's manuscripts [3]. The second consists of words from pages of Jennie Leavitt Smith's diary[2], downloaded from the "Mormon Missionary Diaries" online collection of the Brigham Young University Harold B. Lee Library, available at http://www.lib.byu.edu/dlib/mmd/. We manually segment and label each word to provide ground truth for our experiments.

For each dataset, we select the first 1,000 word images as training examples for which the recognition system is allowed to look at the labels. We use the next 1,000 words (which are not used as training examples) as test data. We compare each test word with the training examples and assign to it the label from the training word that it most closely matches.

Since we are using relatively small amounts of training data (only 1000 words), many of the test words are *Out of Vocabulary* (OoV) words, meaning there are no training examples that have the same label as their ground truth. As such, we report the recognition accuracy with respect to the number of in-vocabulary words (total test words minus the number of OoV test words for that dataset). The Washington dataset contains 748 in-vocabulary test words (252 OoV). The Smith dataset contains 787 in-vocabulary test words (213 OoV).

[2]Our preprocessed Smith dataset word images are available upon request.

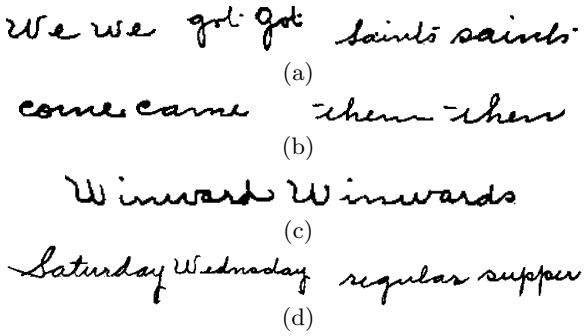

Figure 2: Examples of recognition errors (Smith dataset). Test words followed by the erroneous best match. a) Errors only because of capitalization differences. b) Very similar words: "come" vs. "came" and "them" vs. "then." c) "Winward" vs "Winwards." d) Some more obvious errors.

We assess the recognition accuracy of our method by comparing the ground truth labels with the labels assigned by the recognizer. Recognition accuracy is calculated as the number of test words labeled correctly by the recognizer (the number given the same label as its ground truth), divided by the total number of in-vocabulary test words. The string comparison between the label and ground truth is case-sensitive.

We also assess the recognition accuracy when the correct answer is not the best match, but is found within the top N matches. This gives us a metric of how often our method is "almost" right.

4. RESULTS AND DISCUSSION

For the Washington dataset, in-vocabulary recognition accuracy is 88.77% (664 words are recognized correctly out of 748 in-vocabulary) and for the Smith dataset accuracy is 89.33% (703 out of 787).

Many of the recognition errors that we see are minor, such as differences in case, single letters, or word endings (Figure 2a–2c). Some errors are more blatant (Figure 2d). For many errors, the correct match is ranked very near the top (Table 1). In fact, the correct result is ranked in the top 3 matches more than 94% of the time for both datasets, and within the top 10 almost 97% of the time.

These initial results are very encouraging, and since this is a new recognition method, it is likely that future improvements can be made to increase the recognition accuracy.

Although our experiments so far have been on relatively small, single-author datasets (with fairly good penmanship, as can be seen in Figure 2), we are currently extending the method to handle large, multiple-author datasets.

5. APPLICATION TO FAMILY HISTORY

Besides the obvious application of automatic transcription of handwritten documents, there are other ways our recognition method could be used, even before future improvements in recognition accuracy rates. We mention just a few here.

Table 1: Correct Label Within Top-N Matches

Dataset	Top-1	Top-3	Top-5	Top-10
Wash.	88.77% (664/748)	94.52% (707/748)	96.26% (720/748)	96.93% (725/748)
Smith	89.33% (703/787)	94.28% (742/787)	94.79% (746/787)	96.82% (762/787)

5.1 Temporary Indexes and Transcriptions

Often, great care is taken to assure transcriptions and indexes are as accurate as reasonably possible. On the other hand, collections of images without indexes are also sometimes made available before the indexing can be done. In those cases, it may be possible to automatically provide a temporary index. While imperfect, the index may be accurate enough to allow some of the information of interest to be found through digital searches long before the more reliable index is complete.

5.2 Ranked Search Results

A slightly different solution for the same situation (collections of images that haven't yet been indexed by humans) could be to store the top N matches for each word in the index, instead of just the best match. Then when a user performs a search, the query recall would be higher (more of the relevant information would be returned as a query result). The match costs could be used to help rank the search results so the most likely results appear close to the top.

5.3 Reducing the Indexing Workload

Our HR engine could potentially be used as an automatic first indexer in an indexing pipeline. Currently, two people index the same records and then if they disagree a third person arbitrates. If HR were used before the first human indexer and the HR transcription matched the first human indexer's transcription, there would be no need to have a second person duplicate the work. On the other hand, if the HR transcription didn't match the first human, a second person could transcribe the record, followed by an arbitrator if necessary (just as is currently done) to guarantee a level of index integrity similar to that of the current manual process even when the HR system was wrong. Even an HR system with relatively low accuracy could prevent a significant amount of the duplication workload.

6. CONCLUSION

In this paper, we have reported on the novel HR method that we recently published at ICDAR 2011. We described the methods at a high level and our initial results. More details are available in the ICDAR paper. We also discussed some ways in which our HR method could be applied to family history documents. These suggestions are in addition to the obvious application of automatic transcription of handwritten documents.

7. REFERENCES

[1] P. Gao and T. W. Sederberg. A work minimization approach to image morphing. *The Visual Computer*, 14:390–400, 1998.

[2] D. J. Kennard, W. A. Barrett, and T. W. Sederberg. Word warping for offline handwriting recognition. In *2011 International Conference on Document Analysis and Recognition (ICDAR)*, pages 1349–1353, Beijing, China, Sep. 2011.

[3] V. Lavrenko, T. M. Rath, and R. Manmatha. Holistic word recognition for handwritten historical documents. In *Proc. of the Int'l Workshop on Document Image Analysis for Libraries (DIAL)*, pages 278–287, Palo Alto, CA, Jan. 2004.