Pattern Recognition 48 (2015) 849–861

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

An Arabic handwriting synthesis system

Yousef Elarian [a], Irfan Ahmad [a], Sameh Awaida [b], Wasfi G. Al-Khatib [a], Abdelmalek Zidouri [a,*]

[a] *King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia*
[b] *Qassim University, Qassim 51452, Saudi Arabia*

ARTICLE INFO

Article history:
Received 3 May 2014
Received in revised form
7 August 2014
Accepted 15 September 2014
Available online 28 September 2014

Keywords:
Arabic handwriting
Handwriting synthesis
Training data
Offline text recognition
Statistical model

ABSTRACT

In this paper, we present an Arabic handwriting synthesis system. Two concatenation models to synthesize Arabic words from segmented characters are adopted: Extended-Glyphs connection and Synthetic-Extensions connection. We use our system to synthesize handwriting from a collected dataset and inject it into an expanded dataset. We experiment by training a state-of-the-art Arabic handwriting recognition system on the collected dataset, as well as on the expanded dataset, and test it on the IFN/ENIT Arabic benchmark dataset. We show significant improvement in recognition performance due to the data that was synthesized by our system.

1. Introduction

Handwriting recognition is an active area where researchers are trying various approaches to increase recognition rates [1]. Researchers agree that expanding the training set of a text recognition system is generally beneficial to recognition rates. However, conventional ways of collecting datasets can be time-consuming and may incur a lot of effort, especially for ground-truthing. Hence, researchers proposed the use of synthesized data in expanding training sets of recognition systems [2–6].

Handwriting synthesis refers to the computer generation of online and offline data that resemble human handwriting. It is a reverse process for handwriting recognition as it transforms input text into image samples, whereas recognition maps handwritten samples into digital text.

Handwriting synthesis has become a topic of rapidly increasing interest because of its applications such as the improvement of text recognition systems (in terms of overall performance [2,7], stability [8,9], and speed [10,11]), personalized fonts [12,13], and forgery detection [14,15]. Depending on the application, synthesis methods and their corresponding evaluation methods vary. Personalized fonts, for example, aim at capturing the style of a particular writer and tend to be evaluated subjectively. Whereas synthesized data for text recognition may aim at maximizing style variability within natural limits [3,16,17] and its evaluation is mainly tied to recognition rates.

Handwriting synthesis encompasses *generation* and *concatenation* operations [18,19]. Handwriting generation operations alter samples of handwriting to increase their shape-variability within some closed-vocabulary. Concatenation operations, in contrast, aim at the compilation of new units of vocabulary, such as words, from a smaller pool of basic samples, such as characters. Handwriting generation can be seen as the inverse operation of preprocessing in a text recognition system whereas handwriting concatenation can be regarded as the inverse operation of segmentation.

Synthesized data can improve systems that have deficiencies in their text segmentation accuracy, their recognition features and classifiers, or in the variability of their training data. One advantage of this approach is that it functions on the data level which is system-independent.

Arabic is a widely used language and the Arabic script is used in other languages as well like Urdu and Persian [20]. In Arabic, most characters must connect to their successor within a word. These characters take one of four character-shapes: Beginning (B), Middle (M), Ending (E), and Alone (A); the few characters that do not connect to their successors can only take the (E) or (A) character-shapes. These characters cause Arabic words to break into Pieces of Arabic Words (PAWs). From right to left, a multi-character PAW consists of one (B) character-shape followed by zero or more (M) character-shapes and is terminated by one (E) character-shape. A PAW that consists solely of one character always takes the (A) character-shape.

Characters connect in Arabic via a stroke called the Kashida [21]. Kashida are semi-horizontal strokes that often lie in the

* Corresponding author. Fax: +966 138603535.
E-mail addresses: yarian@kfupm.edu.sa (Y. Elarian), irfanics@kfupm.edu.sa (I. Ahmad), s.awaida@qu.edu.sa (S. Awaida), wasfi@kfupm.edu.sa (W.G. Al-Khatib), malek@kfupm.edu.sa (A. Zidouri).

http://dx.doi.org/10.1016/j.patcog.2014.09.013
0031-3203/

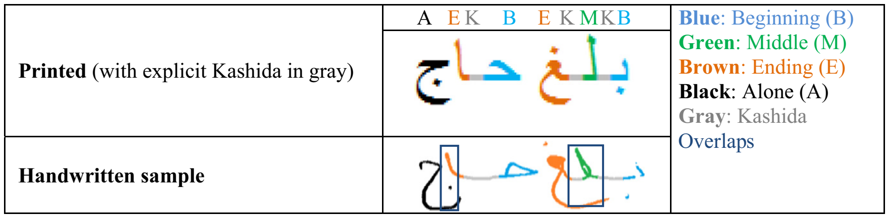

Fig. 1. Arabic printed and handwritten samples colored to distinguish their character-shapes and connection strokes called Kashida.

baseline zone of an Arabic text-line. Unlike printed text, vertical overlaps between PAWs and characters are common in Arabic handwriting. Fig. 1 presents some of the aforementioned features of Arabic writing.

In this paper, we aim at the synthesis of offline Arabic handwriting for its use in improving handwritten text recognition systems. This paper is organized as follows. In Section 2, we discuss related work. In Section 3, we present our concatenation-based synthesis system. Our experiments and results are presented in Section 4. Finally, the conclusion is summarized in Section 5.

2. Related work

There are top-down and bottom-up approaches for handwriting synthesis. Top-down approaches, also referred to as movement-simulation, aim at modeling the neuromuscular actions of writing [22–29]. Bottom-up approaches, also known as shape-simulation, model the written samples themselves [30]. Movement-simulation usually requires the acquisition of online data on tablets; hence, shape-simulation approaches may be more practical for offline data [31].

Synthesis techniques may aim at the generation or concatenation of samples. Generation techniques produce new synthesized images at the same level of the input samples they receive; e.g. new character samples from input character samples. Concatenation techniques, in contrast, produce output images at higher levels than their inputs; e.g. word samples from character samples. Generation encompasses perturbation-based, fusion-based and model-based techniques [32]. Perturbation-based techniques alter a handwritten sample using image-processing tools. Several geometric perturbations are applied on handwritten text-lines to supplement training sets of recognition systems [3,16–18]. Similarly, affine transformations by Wakahara et al. [33] and local perturbations by Keysers et al. [34] were applied for the same goal. Fusion-based, or example-based, techniques merge few samples into new hybrid ones. Among the works adopting fusion-based techniques for the expansion of training sets are those of Zheng and Doermann [32] and Viswanath et al. [35,36]. Research efforts have been little here probably because it lacks established models. Model-based techniques rely on significant numbers of online [31,37,38] or offline [39,40] samples in capturing the writing styles. Except for perturbation-based techniques, the two other generation techniques require shape-matching [32,35]. Hence, perturbation-based generation techniques are easier to implement. However, their results may appear unnatural if parameters are not calibrated carefully [3,41,42].

Concatenation operations can be performed with or without connecting the aligned units. Alignment without connection sets in juxtaposition character groups to form words and lines, as in [43–46]. Direct-connection techniques connect character tails to their heads. Arabic [47,46] and Latin [2,5,48,49] cursive text-lines are synthesized by direct-connection. More sophisticated concatenation was achieved by connection-stroke interpolation which is based on polynomial-models [44,45] spline-models [48,50,51] or probabilistic-models [2,49].

For the Arabic script, shape-simulation was first presented in Elarian et al. [46] for offline handwritten text, where we introduced the idea of sample selection. Another work, but for online recognition, was presented in [6,47]. Dinges et al. [52,53] generate and concatenate Arabic character-shapes represented by Active Shape Models. They synthesize text and modify it by affine transformations and B-Spline interpolation to obtain artificial offline handwriting.

One major application for handwriting synthesis is to improve OCR systems. Some researchers, e.g. [3,7,16], inject synthesized data to improve the original results, whereas others, e.g. [2,6,47], experiment on the synthesized data only without the original data. Bayoudh et al. [7] experiment on online writer-dependent lower-case-character Radial Basis Function Network (RBFN) and Support Vector Machines (SVM) recognizers. They inject 300 synthesized versions of the 26 English characters to the training set. Their best improvement increases the character recognition rate (CRR) by approximately 13 percents. Helmers and Bunke [2] generate data that performs approximately as well as collected data on an offline HMM recognizer whereas Varga and Bunke [3,17] perturb handwritten text-lines to expand their training set and improve their recognition rate. Saabni and El-Sana intent to find Piece of Arabic Word (PAW) recognition rates (PAWRR) for alternative online and offline training sets [6,47]. They evaluate their work on a Dynamic Time Wrapping (DTW) online recognizer [58] and adaptation of it for offline recognition. Similarly, Miyao and Maruyama synthesize a virtual Hiragana dataset that performs comparably to their original dataset [4]. The robustness, speed and performance of the recognition of online gestures are addressed in [9]. We summarize the Related Work section in Table 1.

In our current work, we present shape-simulation synthesis by direct concatenation and statistically-modeled connections to synthesize handwriting samples that look natural to improve recognition system training.

3. Synthesis of Arabic handwriting

We present an approach for synthesis of Arabic handwriting by concatenation techniques. The approach is outlined in the block diagram of Fig. 2. The approach takes character-shape images classified as strictly segmented or extended characters as inputs and concatenates them into synthesized handwriting. Important properties of the dataset are detailed in Section 3.1.

The filled rectangles in the diagram show the four steps of the synthesis procedure while the rounded rectangles indicate the information needed in each step. The connection-point location block receives character-shapes and intends to locate their connection-points so that character-shapes can be aligned to them when concatenated. This step can benefit from contextual information about the characters gathered from the database as discussed in Section 3.2. Then, some features are computed on the connection parts and on the character-shapes to help selecting