Pattern Recognition Letters 74 (2016) 74–81

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Baseline detection of multi-lingual unconstrained handwritten text lines☆

Dibyayan Chakraborty[a,*], Umapada Pal[b]

[a] Advanced Computing and Microelectronics Unit, Indian Statistical Institute, Kolkata
[b] Computer Vision And Pattern Recognition Unit, Indian Statistical Institute, Kolkata

ARTICLE INFO

Article history:
Received 15 March 2015
Available online 15 February 2016

Keywords:
Handwriting recognition
Baseline detection
Multilingual document analysis
Multioriented text processing

ABSTRACT

Many handwritten text recognition systems use the baseline information for better recognition of text line characters. Improper baseline detection reduces the performance of the recognition. In this paper we propose a novel baseline detection scheme for unconstrained handwritten text lines of multilingual documents. For baseline detection of a text line, at first, we detect the set of significant contour points (S-points) of the text line. Every non-singleton subsets of S-points forms a curve. The orientation invariant features of the curve determine whether the curve can construct a probable baseline of the input text line or not. It is determined by an SVM, trained using the orientation invariant features of the curves. The curves classified as probable baselines, are sorted according to their relative positions in ascending order to get the optimal baseline. We tested our method on different handwritten text lines of Bangla(Bengali), English(Roman), Kannada, Oriya, Devnagari and Persian scripts and obtained encouraging results.

1. Introduction

In the past decade, the scope of handwritten text recognition systems has been enlarged from word recognition to block or text-line recognition. The baseline of a text line has been used in many recognition systems. The baseline is the fictitious line which follows and joins the lower part of the character bodies. It may be straight or curved. A handwritten text line is called singly-oriented (multi-oriented) if its baseline is straight (curved). State-of-the-art recognition systems often rely on the position of the baseline since its position is useful for detecting ascenders and descenders. The baseline is also used for de-skewing of handwritten text lines. Therefore, a correct estimation of the baseline is required for better recognition.

Many algorithms have been proposed for baseline detection. In the histogram based methods ([3]), the projection of the writing tracing points according to a predefined direction makes it possible to detect the levels of the baselines that coincide with the local maximum of these histograms. In the Hough transform based methods ([15]), the transposition in the polar coordinates space of the tracing points makes it possible to detect agglomerations of point images intersection defining the angle of writing lines skew. [12] proposed a baseline correction algorithm using Bidirectional recurrent neural network for text lines which are skewed, fluctuating, or both. [6] and [14] proposed an algorithm for baseline detection for Arabic handwritten texts. Both the above mentioned methods use language specific features for their purpose. The baseline detection method due to [1] was based on horizontal projection histogram and directions features of sub-words skeleton for Arabic script. [2] proposed a method for baseline detection of words based on projection histograms and angular corrections. [17] proposed a baseline detection method based on certain properties of isothetic covers tightly enclosing the words in a handwritten document. [9] proposed an approach based on density estimation and level set method for text line boundary estimation. This method can be adopted for baseline detection of text lines. [8] proposed the use of supervised multilayer perceptron for classifying local extrema for baseline detection of text lines. HMM based technique([5]) has also been used for baseline detection purpose.

Preliminary experiments show that the above methods have the following drawbacks: (1) The performance of a method decreases when applied on text lines of different scripts. (2) These methods cannot be applied to curved text lines with multiple turns. (We have shown an example of such text line later in Fig. 9(a)). We address these issues in this work.

Our key contribution can be summarized as follows:

1. We propose a novel algorithm which detects the baseline of multi-lingual unconstrained handwritten text line (including multiple turns) of different scripts.

☆ This paper has been recommended for acceptance by Dr. Y. Liu.

* Corresponding author. Tel.: +91 9062199145.
E-mail addresses: dibyayancg@gmail.com (D. Chakraborty), umapada@isical.ac.in (U. Pal).

http://dx.doi.org/10.1016/j.patrec.2016.02.003
0167-8655/

2. We use machine learning along with rotation invariant features for baseline detection of multi-turn images.

Our approach detects a set of significant contour points (S-points) of the text line. The orientation invariant features extracted from the S-points determine whether the curve constructed by the same S-points can construct a probable baseline of the input text line or not. It is determined by an SVM, trained using the orientation invariant features. The curves classified as probable baselines, are sorted according to their relative positions in ascending order to get the optimal baseline. We use probabilistic methods to argue that the accuracy of our method does not fall in presence of different kinds of noises (see Section 4.5). The results show that the proposed method also provides high accuracy for text lines containing words of different scripts.

The rest of this paper is organized as follows. In Section 2 some properties of baseline and water reservoirs of handwritten text lines are discussed. The proposed algorithm is detailed in Section 3. Different experiments and their results are presented in Section 4. Finally, in Section 5 conclusions are drawn.

2. Properties of baselines and water reservoirs of handwritten text lines

To give the ideas of a baseline, here we provide some definitions like upper line, S-points, n-point curve etc. related to baselines of a handwritten text line. The structural properties of handwritten text lines can be found in [10].

Upper line: Upper line is the fictitious line which follows and joins the upper part of the character bodies in a text line. The contour points lying near the upper line are referred to as upper contour points.

S-points: The contour points which might participate in construction of a baseline are referred to as the S-points (Significant points).

Descender S-points: The S-points lying near a descender of a character are referred to as descender S-points.

n-point curve: A curve constructed by n S-points (referred to as control points) is said to form a n-point curve.

n-point baseline: If a n-point curve is also baseline then it is referred to as a n-point baseline.

Curvature vector of n-point curve: The curvature vector of a n-point curve with control points $C_1, C_2, \ldots, C_n$ is the n-2 dimensional vector whose ith element is $\angle(C_i, C_{i+1}, C_{i+2})$ in radian unit.

Let the sequence of contour points of the input text line image I, sorted in counter clockwise direction be P. Let P_i and P_j be two points of P with $i < j$. $\mathscr{L}_{i,j}$ be the sequence of pixels representing the line segment joining P_i and P_j.

Deviation factor: Intersection of $\mathscr{L}_{i,j}$ with I might create some closed cavity regions. The deviation factor of $\mathscr{L}_{i,j}$ is defined as the total area of those cavities. (see Fig. 1). The deviation factor of a n-point curve can be defined as the average deviation factor of the straight lines joining C_i, C_{i+1} with $1 \leq i \leq n-1$.

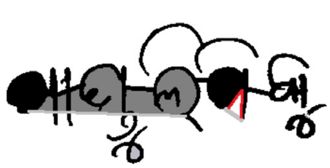

Fig. 1. Illustrations of deviation and coverage factors. The corresponding line segments are shown in light gray. The cavity areas (deviation factor) are shown in dark gray. The covered pixels (with respect to the sequence of boundary pixels) are shown in red. For better visibility please see the pdf file. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Penetration factor: The penetration factor of $\mathscr{L}_{i,j}$ is given as

$$\mathscr{P}_{i,j} = \frac{1}{|\mathscr{L}_{i,j}|} \sum_{P_k \in \mathscr{L}_{i,j}} \mathscr{D}(P_k) \quad (1)$$

where $\mathscr{D}(P_k)$ is the value of the pixel P_k in the distance transform matrix of I. The penetration factor of a n-point curve can be defined as the average penetration factor of the straight lines joining C_i, C_{i+1} with $1 \leq i \leq n-1$.

Coverage factor: Let $V \subseteq P$. A point $P_k \in V$ is said to be covered by $\mathscr{L}_{i,j}$ if $i < k < j$ where P_i and P_j be two endpoints of $\mathscr{L}_{i,j}$. If the number of all covered points in V is Q then coverage factor of $\mathscr{L}_{i,j}$ with respect to V is defined as the ratio $\frac{Q}{|\mathscr{L}_{i,j}|}$. (see Fig. 1). The coverage factor of a n-point curve can be defined as the total coverage factor of the straight lines joining C_i, C_{i+1} with $1 \leq i \leq n-1$.

If water is poured from a side of a component, the cavity regions of the background portion of the component where water will be stored are considered as reservoirs of the component. The details of the water reservoir concept can be found in [16]. However, some of the properties are described here briefly.

Top (bottom) reservoir: By top (bottom) reservoirs of a component, we mean the reservoirs obtained when water is poured from the top (bottom) of the component. A bottom reservoir of a component is visualized as a top reservoir when water is poured from top after rotating the component by 180.

Left (right) reservoir: If water is poured from the left (right) side of a component, the cavity regions of the component where water will be stored are considered as left (right) reservoirs. A left (right) reservoir of a component is visualized as a top reservoir when water is poured from the top after rotating the component by 90 clockwise (anti-clockwise).

Water reservoir area: The area of a reservoir is defined by the area of the cavity region where water will be stored. The number of points (pixels) inside a reservoir is computed, and this number is considered as the area of the reservoir.

Water flow level: The level from which water overflows from a reservoir is called the water flow level of the reservoir (see Fig. 2).

Height of a reservoir: By height of a reservoir, we mean the depth of water in the reservoir.

Minimal water reservoir: The reservoir which does not contain any other reservoir is called minimal reservoirs. Similarly, minimal top water reservoir is the top reservoir which does not contain any other reservoir. (see Fig. 2).

Reservoir depth of a pixel: If a pixel is contained in a reservoir then its depth is defined as the minimum distance between the pixel and the surface of that reservoir.

Minimal reservoir of a pixel: The minimal reservoir of a pixel is that for which the reservoir depth of that pixel is minimum.

Mid-reservoir pixel: Mid-reservoir pixels are those pixels which lie near the mean depth of it's reservoir. Such points are characterised by the ratio of the reservoir depth of the pixel and height of the minimal reservoir of the pixel. Typically, this ratio lies between 0.2 and 0.9.

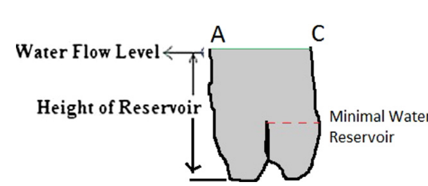

Fig. 2. A top water reservoir, a top minimal water reservoir and its different features are shown. Water reservoir portion is marked in gray.