



HAL
open science

Perceptive Vision for Headline Localisation in Bangla Handwritten Text Recognition

Aurélie Lemaitre, B.B. Chaudhuri, Bertrand Coüasnon

► **To cite this version:**

Aurélie Lemaitre, B.B. Chaudhuri, Bertrand Coüasnon. Perceptive Vision for Headline Localisation in Bangla Handwritten Text Recognition. ICDAR 2007, Sep 2007, Curitiba, Brazil. 10.1109/ICDAR.2007.4376988 . inria-00308561

HAL Id: inria-00308561

<https://inria.hal.science/inria-00308561v1>

Submitted on 31 Jul 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Perceptive Vision for Headline Localisation in Bangla Handwritten Text Recognition

Aur lie Lemaitre
IRISA - INSA
Campus de Beaulieu
35042 Rennes CEDEX
FRANCE
aurelie.lemaitre@irisa.fr

Bidyut Baran Chaudhuri
Indian Statistical Institute
203, B. T. Road
Kolkata - 700 108
INDIA
bbc@isical.ac.in

Bertrand Couasnon
IRISA - INSA
Campus de Beaulieu
35042 Rennes CEDEX
FRANCE
couasnon@irisa.fr

Abstract

In this paper, we propose to give tools for Bangla handwriting recognition. We present a mechanism to segment documents into text lines and words, and more specifically to detect headline position in each word. Indeed, this headline is an horizontal line on the upper part of most of characters, which is characteristic of Bangla writing. Its localisation is a new approach that can improve text recognition quality.

This headline is detected into words inside text lines thanks to a notion of perceptive vision: at a certain distance, text lines appear as line-segments that give the global orientation of words. Watching closer may help to give the exact position of the headline. Consequently, this work is mainly based on applying a segment extractor at different image resolutions and combining extracted information in order to compute the headlines. Our line-segment extractor is based on Kalman filtering.

1. Introduction

Bangla script is one of the major Indian alphabetic systems. Its specificity is its complexity relatively to European Roman-based scripts. Thus, difficulties in recognition of such handwriting are quite specific too. An interesting approach can be the detection of the characteristic headline of words, which is a horizontal line on the upper part of most of the characters.

In this paper, we propose to extract text lines, words and headline of words in handwritten Bangla documents thanks to a notion of perceptive vision: indeed, a global vision will present the global orientation of text lines, whereas a closer look can give a more precise position of elements. Consequently, we propose to combine data extracted at different

vision levels, and their semantic significance, in order to compute a headline for each word.

The paper is organized as follows. We first present specificities of Bangla script and related work for handwriting recognition. Then, we present our line-segment extractor that is our main tool for this work. We expose in section 4 how the use of perceptive vision makes it possible headline detection for each word in a text document. We conclude the paper with some experimental results.

2. Bangla handwriting recognition

2.1. Specificities of Bangla script

Bangla, a morphologically rich language used by 230 million people of eastern India and Bangladesh, was derived from ancient Sanskrit language through various transformations about 1000 years ago. Bangla script, derived from ancient Brahmi, is more complex than Roman-based scripts of Europe. It has about 50 basic vowels and consonants as well as 10 vowel modifiers that may be connected to the basic characters at different positions and create new shapes. Moreover, there are about 250 compound (often fused shapes of two or three consonants) characters in this script and nearly three times newer shapes can be created by attaching the vowel modifiers to the compound characters. The script is not purely alphabetic since the characters do not follow the strict left-right sequence in the temporal flow of utterance. Rather, the written syllables follow a left-right sequence. Thus it may be described as an alphabetic-syllabic script.

One interesting fact for Bangla characters is the existence of a horizontal line at the upper part of the most character, which we shall call as headline. An example of headline is given in figure 8(c). The headline is very prominent in printed text, but not so strong in handwriting, since peo-

ple do not put the headline in 60-70% characters. Yet, a person learned in Bangla writing system can recognize the approximate position of the headline.

2.2. Related work

Automatic recognition of handwritten Bangla script is at an infant stage of research. Pal and Chaudhuri [5] have written an interesting survey that deals mainly with the OCR research of printed Indian scripts. Some work on isolated Bangla character recognition has been reported in recent past [1]. Attempt of reading city names in Bangla from postal documents is also reported [6]. However, we have not come across any paper that deals with the recognition from continuous unconstrained offline Bangla handwriting.

One of the conventional ways of recognizing handwritten text is to identify individual text lines, followed by words in each text line. Then it may be subject to a direct word recognition, or segmented into individual characters before being sent to the feature extraction and to the recognition engine. In the second case, it is very helpful if the headline region of a word is detected, because this line can act as a guiding factor in segmenting the characters. In the former case too, the detection of headline region is useful, since the word-level classifiers like HMM can make a good use of the approximate terminal point of one character and beginning of the following one.

This paper deals with a novel Kalman filter based multi-stage approach to detect handwritten text lines, words as well as headlines in Bangla texts.

3. Line-segment extractor

We use a line-segment extractor, which has been proposed by Leplumey *et al.* in [4], based on Kalman filtering. We recall here the main properties of this approach.

3.1. General principle of Kalman filtering

Kalman filtering is based on a prediction/verification approach. Given a state e_t of an object at instant t , we follow the evolution of this state and predict it during the time, according to real observations.

For line-segment extraction, an observation is composed of a current pixel that belongs to the searched line-segment. The predicted state is the supposed position of next pixel in the image, according to the current slope and the thickness of the searched line-segment.

3.2. Interest of this extractor

We have chosen to use this extractor for three main interesting properties:

1. Possibility to deal with skewed lines.
2. Ability to deal with discontinuities: it is useful to allow locally an absence of point, due to the quality or the nature of extracted object (dotted line, binarization default, noise), as shown in figure 1(a).
3. Possibility to include a "curvature" as a "straight segment": assuming that we can find successive local straight lines inside a curved line, we can decide to keep it (figure 1(b)).

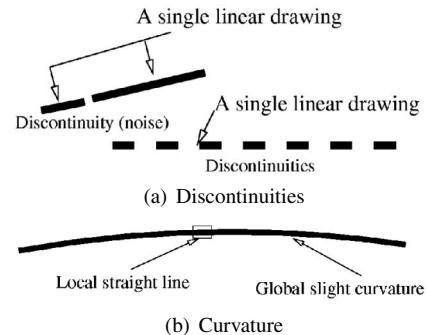


Figure 1. Extractor interesting properties

4. Use of perceptive vision

The main idea for our text line, word and headline extraction is the notion of perceptive vision. This notion is based on the combination of what we can detect at both high and low resolution.

We first present what we can see at low and high resolution and the significance of extracted elements in document structure. Such elements are mainly obtained with our line-segment extractor. Then, we explain how we set up a cooperation process between resolution levels in order to perform a perceptive vision mechanism.

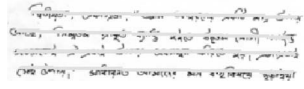
4.1. Low resolution

We call *low* resolution a document of which dimensions have been divided by 8 or 16. We call resolution $-n$ the image which size has been divided by n .

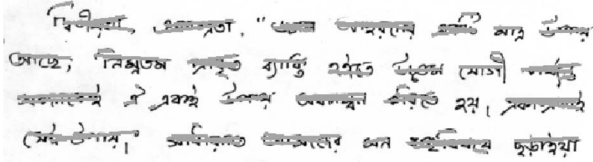
When watching a document at resolution -16 , we can see that text lines appear as line-segments (figure 2(a)). That is why we propose to extract line-segments with Kalman filtering in order to localize text lines. This idea has been developed in [3].

The properties of our line-segments extractor, presented below, are required: indeed, text lines can be skewed and they are seen as dotted lines because of white spaces between words.

With the same principle, we also apply our segment extractor on a -8 resolution image. At that vision level, line-segments correspond to words or small groups of words (figure 2(b)).



(a) Resolution -16



(b) Resolution -8

Figure 2. Low resolution line-segments

We will see that the combination of both resolution levels makes us possible to obtain a reconstruction of text lines.

4.2. High resolution

At high resolution, *i.e.* the initial image resolution, our work is based on three sets of elements: connected components, horizontal and vertical line-segments.

Connected components (figure 3) generally represent letters. They are useful for word grouping.

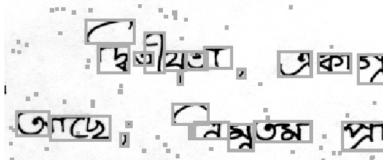


Figure 3. Set of connected component bounding boxes

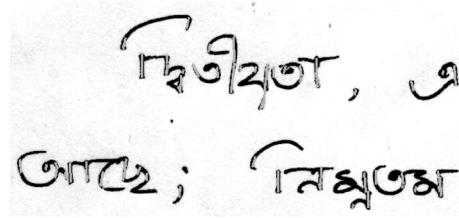
Horizontal and vertical line-segments (figure 4), extracted by Kalman filtering are local small straight parts inside the characters. Those elements will be used to determine the exact position of headline.

This detection requires the interesting properties of our line-segment extractor. Indeed, the ability to deal with skew or slightly curved lines makes it possible to extract the main horizontal and vertical axes of letters as line-segment even if they are not very straight.

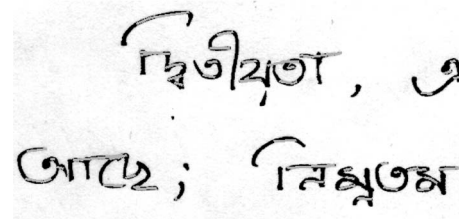
4.3. Cooperation between vision levels

Using all those elements extracted from images at different levels, and their significance in the structure of the doc-

ument, we propose to give a description of text-lines and words and then to detect headline position.



(a) Vertical line-segments



(b) Horizontal line-segments

Figure 4. High resolution line-segments

This implementation has been realized easily, thanks to the method DMOS (Description and MOdification of Segmentation), that is presented in [2], and that we have improved to deal with several resolution images. We have produced a grammatical description of that kind of document in order to automatically generate the associated analyser. This generic system has made us possible to propose a writer-independant analysis, valid for any Bangla text.

Text-line and word extraction Text-line extraction is realized combining segments at resolution -16 and -8 in order to extract a reconstruction for each text line (figure 5). Those two resolution levels are necessary, because a single word text line may not be long enough to be seen at resolution -16. However, it would be detected at resolution -8. We assume that this rebuilt line-segment gives the correct skew of text line and that local headline will follow this slope.

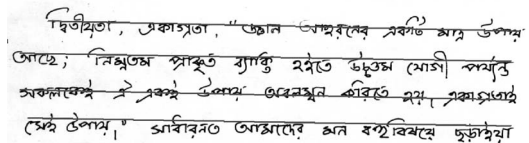


Figure 5. Rebuilt text lines

Then, over each text-line, connected components are grouped into words (figure 6). For each line, we compute every distance between two successive components, using neighbourhood in a Voronoi tessellation based graph. We suppose that we have two sets of distances: intra and inter words. So we find a threshold between those two sets using

k-mean method, $k=2$. This distance threshold gives a way to separate words.

The next step is to find headline position for each word.

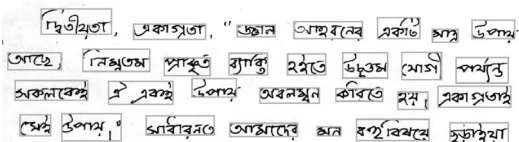


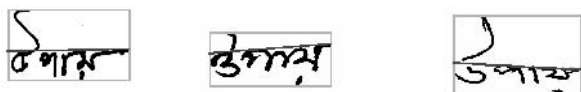
Figure 6. Extracted words

Difficulties in headline positioning Once a word, its bounding box, and headline slope are given, headline positioning is not trivial. Indeed, we could think that headline is globally in upper part of the word bounding box.

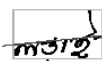
Nevertheless, handwritten characters are varying and the headline will not have the same position from one word to another (figure 7(a)) or with the same word, from a writer to another (figure 7(b)). The position is neither necessarily in the upper part of the word bounding box (figure 7(c)). Consequently, we had to find general rules for headline position.



(a) Between two words from the same writer



(b) For the same word by three writers



(c) Not necessarily in the upper part

Figure 7. Varying headline position relatively to the word bounding-box

Generic description for headline extraction For headline computation, we suppose that the global skew of the text line gives the slope of each headline. Then, in order to find the final headline, we just have to select the correct position along vertical axis.

We propose to have a close watch into words and to take an interest in high resolution line-segments. At that resolution, we make two assumptions:

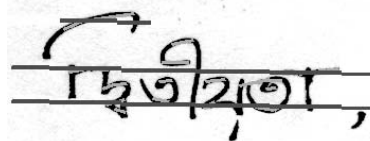
1. The headline is globally located in the top of the word, so it must be at the top of at least one vertical line-segment.
2. The headline is based on horizontally aligned parts of the word, so we suppose that it corresponds to the

longest set of aligned horizontal line-segments, according to the previous condition.

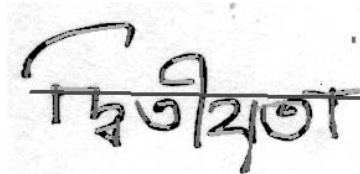
An example of finding position is given in figure 8: we first locate the upper extremities of vertical line-segments (figure 8(a)). Then, we look for long alignments of horizontal line-segments (figure 8(b)). The final correct headline is the longest alignment located at the top of a vertical line-segment.



(a) Vertical position at the top of line-segments



(b) Horizontal interesting position candidates: the upper is too short, the downer is not at top of a vertical line-segment



(c) Final correct headline

Figure 8. Headline final position assuming the two first conditions

For the previously presented image of figure 6, the global result is shown in figure 9.

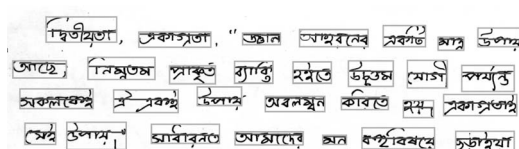


Figure 9. Final extracted headlines in words

5. Results

5.1. Application

We have applied our mechanism on 53 document images of handwritten Bangla text pages, from 26 different writer,

who had to write the same text. Initial images were at a resolution of 300 dpi, with an initial size around 2000*3200 pixels for half documents and 2000*1900 pixels for other part of documents, and stored in TIF. The so called -8 resolution corresponds to 38 dpi images (image size around 250*400 or 250*240 pixels) and resolution -16 corresponds to 19 dpi resolution (image size around 125*200 or 125*120 pixels).

This document base represents 892 text lines to find, that are detected with a rate of 98.9%. The 9 forgotten text lines correspond to cases where lines are too close to each other and merged.

We have manually estimated results on 26 pages, one per writer. Results are presented in table 1. We recognize 1922 words for which headline is correct in 92.87% cases. Errors mainly occurs in small words that are not long enough to present a significant horizontal alignment of line-segments.

Pages	Detected words	Correct headline	Error rate
26	1922	1785 92.87%	137 7.13%

Table 1. Results on the validation base

5.2. Interest of this approach

The main interest of this notion of perceptive vision is the ability to have first a global view before text recognition. This is the key point for skewed images and varied slope inside a single document, which is common for handwritten data. Indeed, the slope is computed independently for each line. Thus, on a document like figure 10, we can obtain text lines with a correct and independent skew.

As shown in figure 7, the headline position varies a lot from one word to another and an analysis using a mere local vision would be difficult. Using a global vision is a main element because we can compute a headline even when the writer did not himself put the headline in his characters. The global context associated to a local vision makes it possible to deal with varying cases.

6. Conclusion

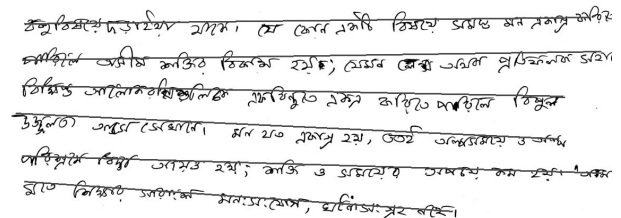
We have presented a new approach for Bangla script recognition based on headline localisation. Headline are specific to Bangla script, and useful for recognition, but not always present in handwritten text. Moreover, their position within a word is very variable.

Using a notion of perceptive vision has made possible a localisation of a headline for each word. Watching the global document at a certain distance gives us text lines and the global slope of each headline. A closer vision makes us

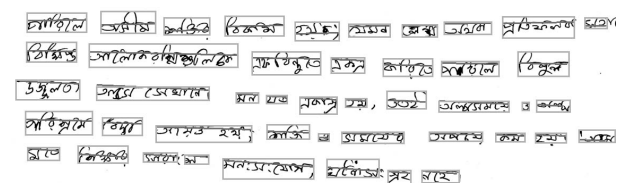
possible to extract words and adjust the vertical position of the headline for each word.

Our approach is generic enough to be writer-independent. We have applied it on a set of 26 writers and obtain a convenient headline in 92.87% cases.

The next step of this work will be to deal with words that are too short to obtain a good headline. Then, the future work will be to take into account this headline in a dedicated OCR and to evaluate the contribution of headline for Bangla handwriting recognition.



(a) Various line skew



(b) Associated extracted words and headlines

Figure 10. Example of varying line skew in documents

References

- [1] U. Bhattacharya, S. K. Parui, and B. Shaw. A hybrid scheme for recognition of Handwritten Bangla basic characters based on HMM and MLP classifiers. In *Proc. International Conference on Advances in Pattern Recognition*, pages 101–106. World Scientific, 2007.
- [2] B. Couasnon. DMOS: A generic document recognition method to application to an automatic generator of musical scores, mathematical formulae and table structures recognition systems. In *Proc. International Conference on Document Analysis and Recognition (ICDAR)*, pages 215–220, 2001.
- [3] A. Lemaitre and J. Camillerapp. Text line extraction in handwritten document with kalman filter applied on low resolution image. In *Document Image Analysis for Libraries (DIAL'06)*, pages 38–45, 2006.
- [4] I. Leplumey, J. Camillerapp, and C. Queguiner. Kalman filter contributions towards document segmentation. In *Proc. International Conference on Document Analysis and Recognition (ICDAR)*, pages 765–769, 1995.
- [5] U. Pal and B. B. Chaudhuri. Indian script character recognition: a survey. *Pattern Recognition*, 37(9):1887–1899, 2004.
- [6] K. Roy, S. Vajda, A. Belaid, U. Pal, and B. B. Chaudhuri. A system for indian postal automation. In *Proc. International Conference on Document Analysis and Recognition (ICDAR)*, pages 1060–1064. IEEE Computer Society, 2005.