

LINE SEGMENTATION ON BANLA HANDWRITTEN DOCUMENT IMAGE

By

Hiranya Debsingha

Supervised by Prof. Kaushik Roy

Department of computer science

West Bengal State University

Introduction

- Line segmentation is precursor of OCR(Optical Character Recognition)
- Design OCR engines are focused on hand "printed" (written in capital letters) text.

1. Print Document

বরফের লাঙলগুলো (রো) প্লাগ কখন ব্যবহার করা হয়? এগুলো কিভাবে কাজ করে? ঝড় শুরু হওয়ার আগেই বরফের লাঙলগুলোকে আগে থেকে জায়গামত নিয়ে রাখা হতে পারে। ২-ইঞ্চি তুষার জমে গেলে বরফের লাঙলগুলো পার্ঠানো হয়। DSNY-এর ট্রাকগুলোর সাথে লাগানো লাঙলের ফলাগুলো রাস্তা থেকে সামান্য উপরে রাখা হয় যাতে লাঙলটি অসমতল রাস্তায় বা ম্যানহোলের ঢাকনায় “ধাক্কা” না খায়। রাস্তার উপরের কালো অংশ দেখা যাওয়ার মত করে বরফ পরিষ্কার করার জন্য লাঙলের ডিজাইন করা হয়নি। মনে রাখবেন, একটি রাস্তাকে লাঙল নিয়ে পরিষ্কার করা যায় যদিও রাস্তার উপরের কালো অংশটি দেখা যাবে না। এটা প্রত্যাশিত। ভারী তুষারপাতের সময় রাস্তাঘাটের উপর লবণ ছিটানো এবং লাঙল ব্যবহার করার সমন্বিত পদ্ধতি সবচেয়ে বেশি কার্যকর হয়ে থাকে।

ফুটপাথ থেকে ভূষাৰ ও বৰফ পৰিষ্কাৰ কৰাৰ দায়িত্ব কাৰ?

ফুটপাথ থেকে ভূষার ও বরফ পরিষ্কার করার দায়িত্ব হলো সংলগ্ন বাড়ির মালিক, লিডগ্রহীতা, ভাড়াটে, বা অন্যান্য বসবাসকারীদের। ভূষার ও বরফ রাস্তায় ঢেলে দেয়া বা পারাপারের জায়গাপ্রত্যেককে ঢেকে দেয়া উচিত নয়। ৩-৪ ফুট প্রশস্ত একটি পথ পরিষ্কার করার চেষ্টা করুন। রাস্তার মোড়ে অবস্থিত বাড়িগুলো পথচারীদের জন্য চালু পথগুলো সহ রাস্তা পারাপারের জায়গায় মাওয়ার পথগুলো পরিষ্কার করা উচিত। যারা এই কাজ করবে না তারা জরিমানার সম্মুখীন হতে পারে। সম্ভব হলে, বয়স্ক ব্যক্তি বা প্রতিবন্ধী প্রতিবেশীদেরকে সাহায্য করার বিষয়টি বিবেচনা করুন।

2. Handwritten Document

[illegible]

Why Line Segmentation?

- Things get more complicated when we consider Indic script like Bangla which contain modified and compound characters along with isolated character like..

compound character

চ্ছ উচ্ছল চ্ছ ব

জ্জ উজ্জল জ্জ ব

ত্ভ রাজত্ভ ত্ভ ব

other form of character

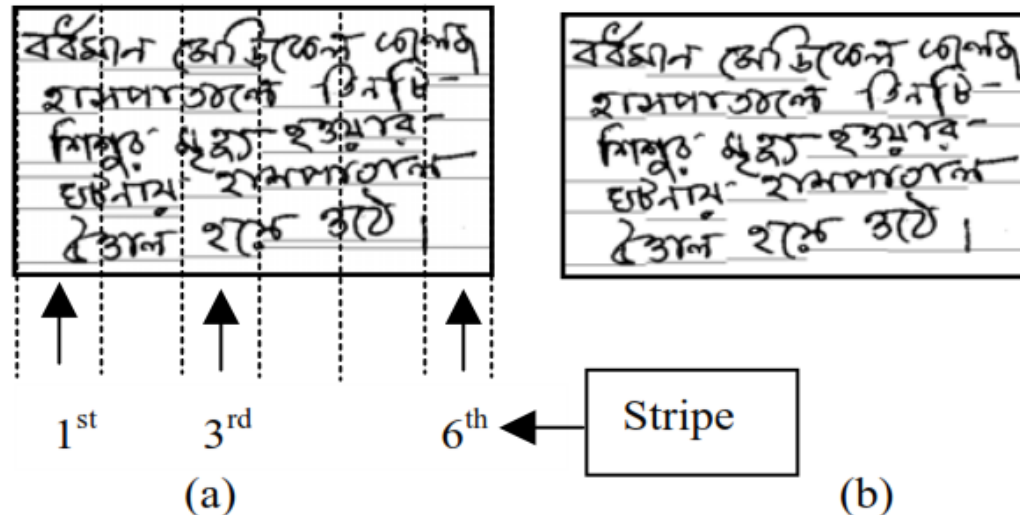
চি

জ্

ত্

Previous Works (1)

- Proposed Approach
 - Divide the text into vertical stripes and determine horizontal histogram projections of these stripes



Previous Works(2)

- Proposed Approach

- smudge the input document image, to blur-out white spaces between words, while preserving gaps between consecutive lines.

বিজয় ভিলাইট মাঁচ চৰা বৈয়া বৰিয়া বুঝিবার চেষ্টা কৰিলে,
তাৰ মূৰে হাত বৰ্ষা বৰ্ষা দিয়া বলিয়া উঠিল, তোমায় .
আমায় কলার কিছু থাকিলে, মাঝনে গঠন কৰুক. নথি-
কবায় সমায় তোমায় নেই- তোমি বাহা-ভালুক-ও নথি,
তাকে মোয়ে খেলৰ না. বাড়ি চাড়াৰ না কেন কৰুক.

তুমিৰ বাহিৰে তোমি না, কিছু কথা কহিলে.
মতোময় ভুয়ে নথি, নিজেৰ ভুয়ে মথ কৰিয়া
বলিলে, বাড়ি চাড়াৰ কথা ছিল না.
তোমায় বাহা হৰিহৰবাবু বুলিছিলে, ওহ ভেতুৰ
ওহে তোমায় বাহা কৰতে মাৰি.

(a)

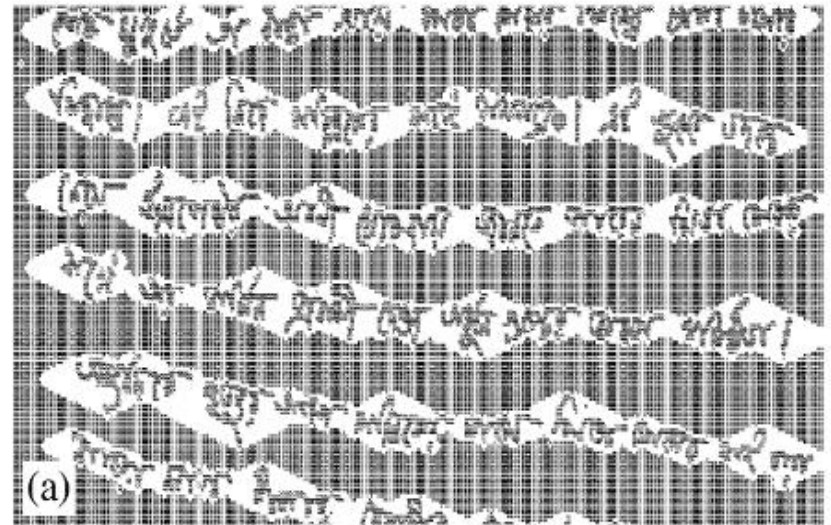
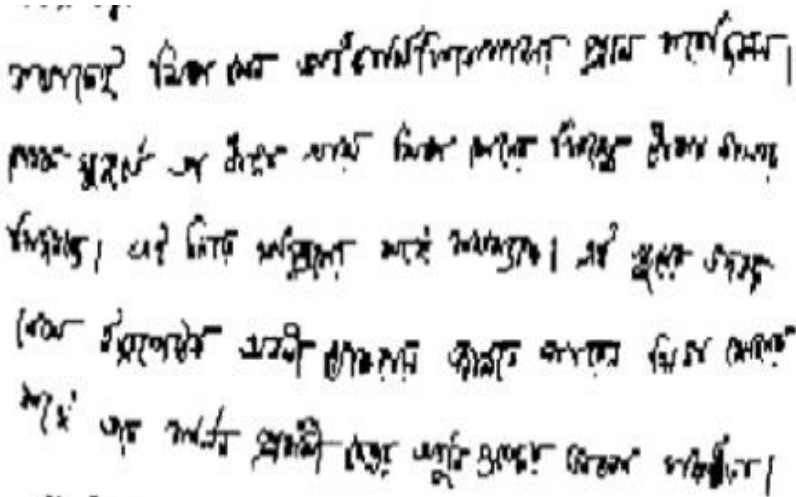
বিজয় ভিলাইট মাঁচ চৰা বৈয়া বৰিয়া বুঝিবার চেষ্টা কৰিলে,
তাৰ মূৰে হাত বৰ্ষা বৰ্ষা দিয়া বলিয়া উঠিল, তোমায় .
আমায় কলার কিছু থাকিলে, মাঝনে গঠন কৰুক. নথি-
কবায় সমায় তোমায় নেই- তোমি বাহা-ভালুক-ও নথি,
তাকে মোয়ে খেলৰ না. বাড়ি চাড়াৰ না কেন কৰুক.

তুমিৰ বাহিৰে তোমি না, কিছু কথা কহিলে.
মতোময় ভুয়ে নথি, নিজেৰ ভুয়ে মথ কৰিয়া
বলিলে, বাড়ি চাড়াৰ কথা ছিল না.
তোমায় বাহা হৰিহৰবাবু বুলিছিলে, ওহ ভেতুৰ
ওহে তোমায় বাহা কৰতে মাৰি.

Previous Works(3)

- Proposed Approach

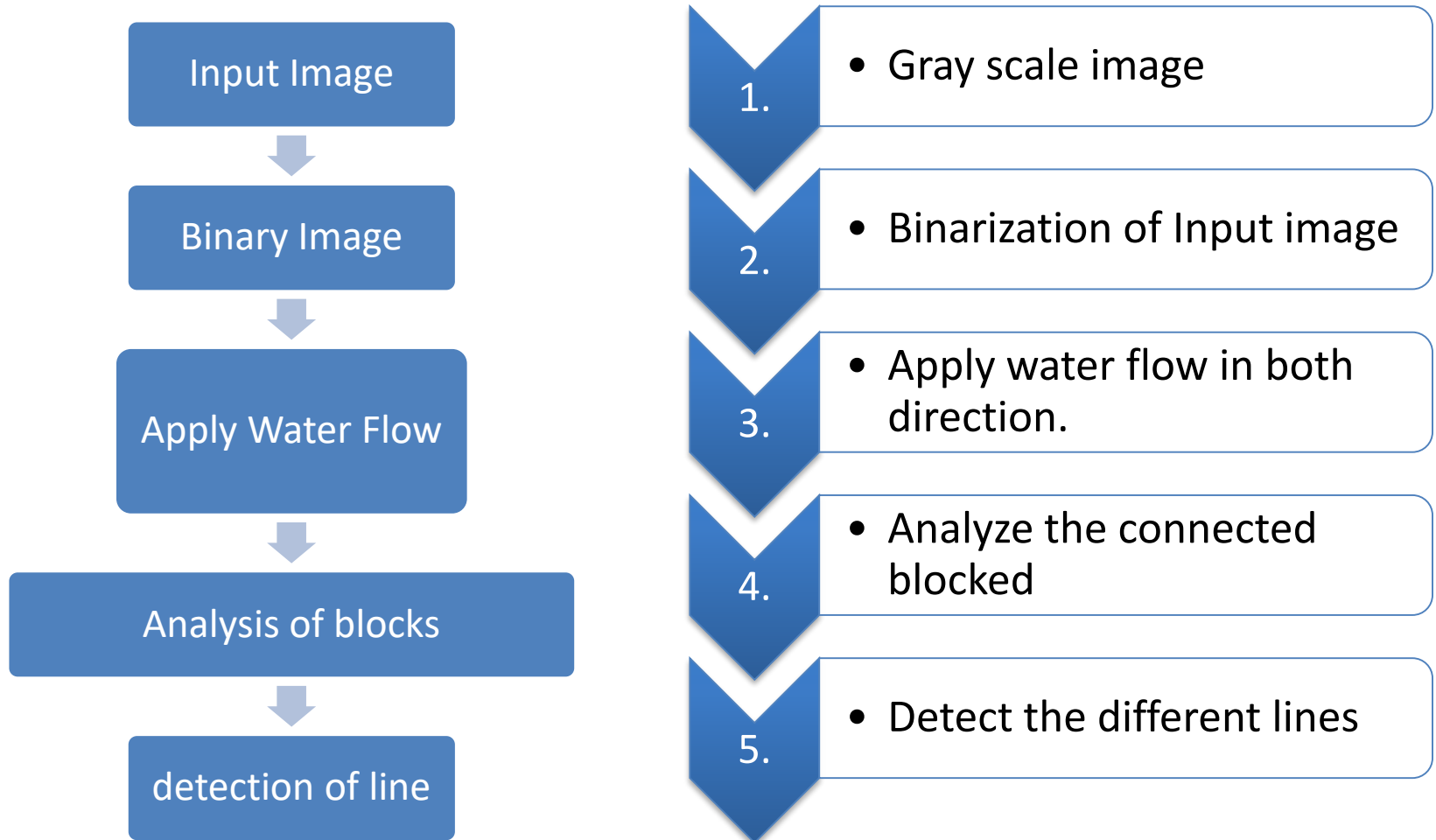
- A flow of water in a particular direction across the image frame
- Water faces obstruction from characters of the text line and changes its flow angle



Problems

- Accuracy
- Robustness
- Application is multi skewed line scenario

Proposed Idea



Future Scope

- Using Convolution Neural Network for implementing this technique in advance way.
- Create a powerful OCR engine which works on different script.

References

- U. Pal and S. Datta, "Segmentation of Bangla unconstrained handwritten text," Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings., Edinburgh, UK, 2003, pp. 1128-1132.
doi: 10.1109/ICDAR.2003.1227832
- S. Basu, C. Chaudhuri, M. Kundu, M. Nasipuri, D.K. Basu, Text "line extraction from multi-skewed handwritten documents" Pattern Recognition, Volume 40, Issue 6, 2007, ISSN 0031-3203, <https://doi.org/10.1016/j.patcog.2006.10.002>.
- K. Mullick, S. Banerjee and U. Bhattacharya, "An efficient line segmentation approach for handwritten Bangla document image," *2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR)*, Kolkata, 2015, pp. 1-6. doi: 10.1109/ICAPR.2015.7050679
- Jiang Yong and Chen Xiaojing, "Text line extraction from multi-skewed handwritten documents," *2008 27th Chinese Control Conference*, Kunming, 2008, pp. 412-415.
doi: 10.1109/CHICC.2008.4604879
- Brodic, Darko & Milivojevic, Zoran. (2011). "A New Approach to Water Flow Algorithm for Text Line Segmentation". *Journal of Universal Computer Science*. 17. 30-47.

Thank You