BANGLA OPTICAL CHARACTER RECOGNITION

A Thesis

Submitted to the Department of Computer Science and Engineering

of

BRAC University

by

Nawsher Ahmed Noor

Student ID: 01201070

In Partial Fulfillment of the

Requirements for the Degree

of

Bachelor of Science in Computer Science

December 2005

DECLARATION

I hereby declare that this thesis is based on the results found by myself and my group partner S. M. Murtoza Habib. Materials of work found by other researcher are mentioned by reference.

Signature of Supervisor

Signature of Author

ACKNOWLEDGMENTS

Special thanks to Dr. Mumit Khan who gave us guidance through our thesis research work and also for his support in the whole process. And also show him our gratitude for considering us as to do the thesis under his supervision.

We thank Sajib Dasgupta for his insight and encouragement for working on this area.

We also thank M. Ashraful Amin for his direction in the area of image processing and neural networks, and also thank Junaed Sattar for his insight in image analysis.

And finally we thank our research institute for providing us with old Bangla books and scanner as a requirement for collection of sample inputs.

ABSTRACT

This report presents an optical character recognition approach for Bangla (national language of Bangladesh) offline printed characters. Sample training data, scanned using a modest scanner, is first translated as gray-scale pixel representation. Pre-processing steps that follows are skew angle detection and correction, noise removal, line, word and character separation. The separated characters are then fed into a two layer feed forward Neural Network for training. Finally this network is used to recognize printed Bangla scripts.

Table of Contents

Page

List of Tables

List of Figures

Chapter 1: Introduction

Optical Character Recognition (often abbreviated as OCR) involves reading text from paper and translating the images into a form (say ASCII codes) that the computer can manipulate. Although there has been a significant number of improvements in languages such as English, but recognition of Bengali scripts is still in its preliminary level. This thesis tries to analyze the neural network approach for Bangla Optical Character Recognition. A feed forward network has been used for the recognition process and a back propagation algorithm had been used for training the net. Before the training, some preprocessing steps were involved of course. Preprocessing includes translating scanned image into binary image, skew detection & correction, noise removal, followed by line, word and character separation. In this report, preprocessing steps were discussed from chapter 3 to chapter 5, and chapter 6 explains recognition. Chapter 2 discuss about some properties of Bangla Character.

Translation of scanned image into binary image, skew detection & correction, noise removal, line and word separation of the pre-processing steps were jointly analyzed and done by my group member and myself.

Character separation was fully done by my group member while I concentrated solely on forming, training and recognition of characters.

Chapter 2: Some properties Bangla character

The writing style of Bangla is from left to right comprising of 11 vowels and 39 consonant characters. These characters may be called as the basic characters (figure 1). The concept of upper/lower case is absent in Bangla script. From Fig. 1 it is noted that most of the characters have a horizontal line at the upper level. This horizontal line is called head line. In Bangla language, we call it 'matra'.

A Av B C D E
F G H I J

(a)

K L M N O P Q R S T
U V W X Y Z _ ` a b
c d e f g h i j k l
m n o p q r s t

(b)

Figure 1: (a) Vowels; (b) consonant.

In Bangla script sometimes a vowel takes a modified shape depending on the position in a word. If the first character of the word is a vowel then it retains its basic shape. Generally a vowel following by a consonant takes a modified shape and placed at the left or right or both or bottom of the consonant (table 1).

Table 1: Example of modified shape of vowel

Vowel	Av	B	C	D	E	F	G	H	I	J
Modified shape	v	w	x				‡	‰	‡ v	‡ Š
K + vowel	Kv	wK	Kx	K	K	K	‡K	‰K	‡Kv	‡KŠ

For two consecutive vowels in a word, the second one remains its basic shape when the first one in modified shape. (figure 2)

L + A¨ + I → L¨I

Figure 2: non-modified Vowel in a word

Again a consonant or vowel followed by a consonant sometimes takes a compound shape which we call as compound character (table 2). There are about 250 compound characters, where most of them are formed by consonant-consonant combination. Compounding of three consonants is also possible. Most interesting thing is if we change the order of same two consonant, the compound character is changed.

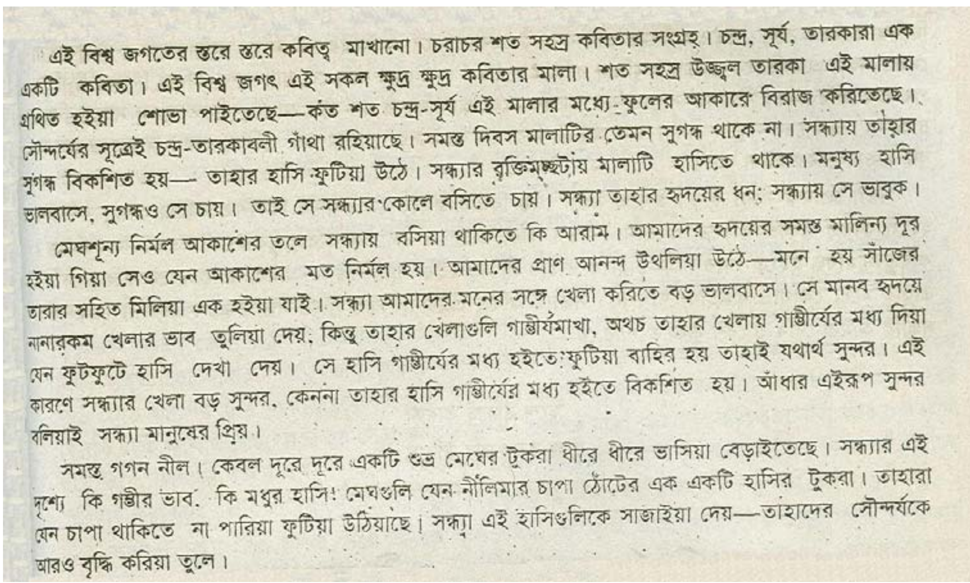
Table 2: Example of compound character

K + K	°		e + e	eŸ
K + Z	³		e + `	ã
K + b	K		P + Q	”Q
K + g	´		P + Q + e	”Q
K + l	¶		R + R	¾
K + l + b	¶		R + R + e	¾
K + l + g	²		R + T	Á
K + j	K		T + R	Ä
j + K	é		0 + K	¼

As we mentioned earlier, most of the basic characters have a head line (matra). Out of 50 basic characters 32 of them have this matra (head line) and most of the compound characters have this matra too.

Chapter 3: Processing digital image

A digital image of text is required which is achieved through scanning a paper or book containing Bangla script; figure 3 is such an image. The quality of the scanned paper in figure 3 was not that good, and thus presence of noise is higher. Now the first task is to convert this image to binary image (black and white). Generally the scanning image is true color (RGB image) and this has to be converted into a binary image, based on a threshold value.


এই বিশ্ব জগতের স্তরে স্তরে কবিত্ব মাখানো। চরাচর শত সহস্র কবিতার সংগ্রহ। চন্দ্র, সূর্য, তারকারা এক একটি কবিতা। এই বিশ্ব জগৎ এই সকল ক্ষুদ্র ক্ষুদ্র কবিতার মালা। শত সহস্র উজ্জ্বল তারকা এই মালায় গ্রথিত হইয়া শোভা পাইতেছে—কত শত চন্দ্র-সূর্য এই মালার মধ্যে ফুলের আকারে বিরাজ করিতেছে। সৌন্দর্যের সূত্রেই চন্দ্র-তারকাবলী গাঁথা রহিয়াছে। সমস্ত দিবস মালাটির তেমন সুগন্ধ থাকে না। সন্ধ্যায় তাহার সুগন্ধ বিকশিত হয়— তাহার হাসি ফুটিয়া উঠে। সন্ধ্যার রক্তিমচ্ছটায় মালাটি হাসিতে থাকে। মনুষ্য হাসি ভালবাসে, সুগন্ধও সে চায়। তাই সে সন্ধ্যার কোলে বসিতে চায়। সন্ধ্যা তাহার হৃদয়ের ধন; সন্ধ্যায় সে ভাবুক।

মেঘশূন্য নির্মল আকাশের তলে সন্ধ্যায় বসিয়া থাকিতে কি আরাম। আমাদের হৃদয়ের সমস্ত মালিন্য দূর হইয়া গিয়া সেও যেন আকাশের মত নির্মল হয়। আমাদের প্রাণ আনন্দ উথলিয়া উঠে—মনে হয় সাঁজের তারার সহিত মিলিয়া এক হইয়া যাই। সন্ধ্যা আমাদের মনের সঙ্গে খেলা করিতে বড় ভালবাসে। সে মানব হৃদয়ে নানারকম খেলার ভাব তুলিয়া দেয়; কিন্তু তাহার খেলাগুলি গাম্ভীর্যমাখা, অথচ তাহার খেলায় গাম্ভীর্যের মধ্য দিয়া যেন ফুটফুটে হাসি দেখা দেয়। সে হাসি গাম্ভীর্যের মধ্য হইতে ফুটিয়া বাহির হয় তাহাই যথার্থ সুন্দর। এই কারণে সন্ধ্যার খেলা বড় সুন্দর, কেননা তাহার হাসি গাম্ভীর্যের মধ্য হইতে বিকশিত হয়। আঁধার এইরূপ সুন্দর বলিয়াই সন্ধ্যা মানুষের প্রিয়।

সমস্ত গগন নীল। কেবল দূরে দূরে একটি শুভ্র মেঘের টুকরা ধীরে ধীরে ভাসিয়া বেড়াইতেছে। সন্ধ্যার এই দৃশ্যে কি গম্ভীর ভাব. কি মধুর হাসি! মেঘগুলি যেন নীলিমার চাপা ঠোঁটের এক একটি হাসির টুকরা। তাহারা যেন চাপা থাকিতে না পারিয়া ফুটিয়া উঠিয়াছে। সন্ধ্যা এই হাসিগুলিকে সাজাইয়া দেয়—তাহাদের সৌন্দর্যকে আরও বৃদ্ধি করিয়া তুলে।

Figure 3: An example of digital image of Bangla text

Here Otsu's Method [1] has been applied to find the threshold value. The method chooses the threshold to minimize the intra class variance of the thresholded black and white pixels. The algorithm is:

Algorithm 1: Otsu's Method

Step 1: count the number of pixel according to color (256 color) and save it to matrix *count*.

Step 2: calculate probability matrix P of each color, P_i = $count_i$ / sum of *count*, where i= 1, 2, … … 256.

Step 3: find matrix *omega*, $omega_i$ = cumulative sum of P_i , where i= 1, 2, … … 256.

Step 4: find matrix *mu*, mu_i = cumulative sum of P_i *i , where i= 1, 2, … … 256 and *mu_t* = cumulative sum of P_{256} * *256*

Step 5: calculate matrix sigma_b_squared where,

$$sigma\_b\_squared_i = \frac{(mu\_t \times omega_i - mu_i)^2}{omega_i - (1 - omega_i)}$$

Step 6: Find the location, *idx*, of the maximum value of *sigma_b_squared*. The maximum may extend over several bins, so average together the locations.

Step 7: If maximum is not a number, meaning that sigma_b_squared is all not a number, and then *threshold* is 0.

Step 8: If maximum is a finite number, *threshold* = (*idx* - 1) / (256 - 1);

To get a binary image, this RGB format image has to be converted to grayscale format, and then by using the threshold value (found by Otsu’s method) this grayscale image is converted to binary image.

A RGB image is converted by eliminating the hue[1] and saturation[2] information while retaining the luminance[3]. The easy technique to do obtain this is to multiply 0.2989 with red color, 0.5870 with green color and 0.1140 with blue color of a particular pixel and put the summation of these three values to that particular location.

Y = Red * 0.2989 + Green * 0.5870 + Blue * 0.1140

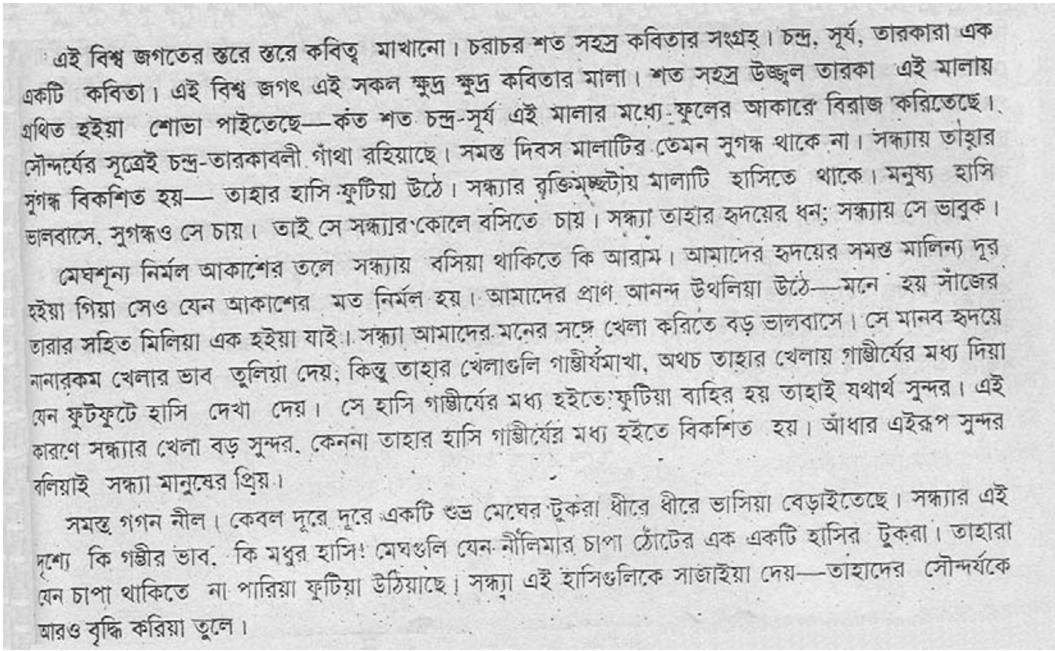
---

[1] A hue refers to the degree of color within the optical spectrum, or visible spectrum, of light. "Hue" may also refer to a particular color within this spectrum.

[2] Saturation refers to the intensity of a specific hue.

[3] Luminance describes the amount of light that passes through or is emitted from a particular area.

The grayscale version of the image (figure 3) is shown in figure 4.



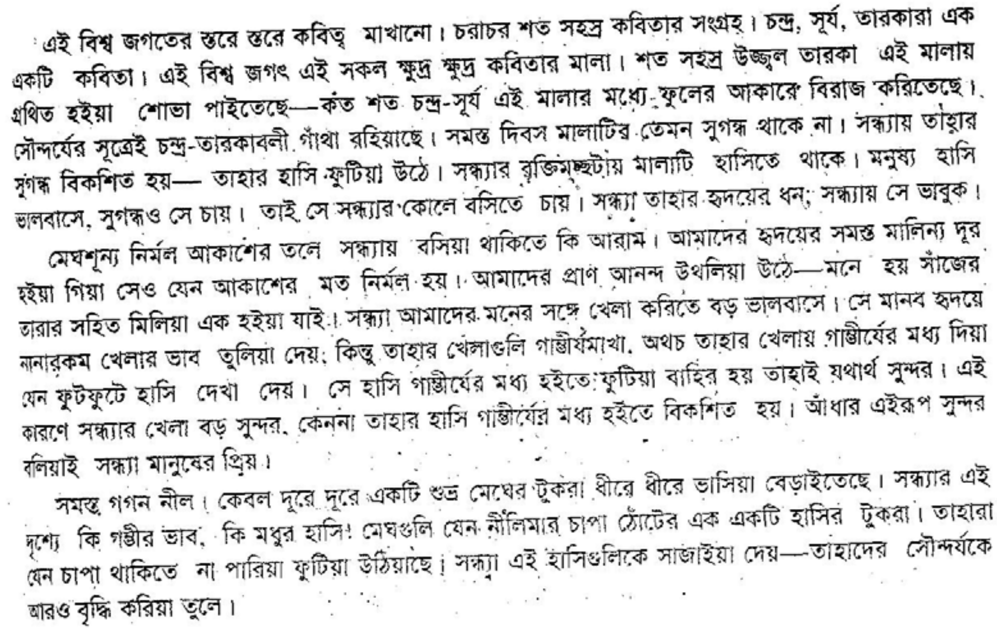
এই বিশ্ব জগতের স্তরে স্তরে কবিত্ব মাখানো। চরাচর শত সহস্র কবিতার সংগ্রহ। চন্দ্র, সূর্য, তারকারা এক একটি কবিতা। এই বিশ্ব জগৎ এই সকল ক্ষুদ্র ক্ষুদ্র কবিতার মালা। শত সহস্র উজ্জ্বল তারকা এই মালায় গ্রথিত হইয়া শোভা পাইতেছে—কত শত চন্দ্র-সূর্য এই মালার মধ্যে ফুলের আকারে বিরাজ করিতেছে। সৌন্দর্যের সূত্রেই চন্দ্র-তারকাবলী গাঁথা রহিয়াছে। সমস্ত দিবস মালাটির তেমন সুগন্ধ থাকে না। সন্ধ্যায় তাহার সুগন্ধ বিকশিত হয়— তাহার হাসি ফুটিয়া উঠে। সন্ধ্যার রক্তিমচ্ছটায় মালাটি হাসিতে থাকে। মনুষ্য হাসি ভালবাসে, সুগন্ধও সে চায়। তাই সে সন্ধ্যার কোলে বসিতে চায়। সন্ধ্যা তাহার হৃদয়ের ধন; সন্ধ্যায় সে ভাবুক।

মেঘশূন্য নির্মল আকাশের তলে সন্ধ্যায় বসিয়া থাকিতে কি আরাম। আমাদের হৃদয়ের সমস্ত মালিন্য দূর হইয়া গিয়া সেও যেন আকাশের মত নির্মল হয়। আমাদের প্রাণ আনন্দ উথলিয়া উঠে—মনে হয় সাঁজের তারার সহিত মিলিয়া এক হইয়া যাই। সন্ধ্যা আমাদের মনের সঙ্গে খেলা করিতে বড় ভালবাসে। সে মানব হৃদয়ে নানারকম খেলার ভাব তুলিয়া দেয়; কিন্তু তাহার খেলাগুলি গাম্ভীর্যমাখা, অথচ তাহার খেলায় গাম্ভীর্যের মধ্য দিয়া যেন ফুটফুটে হাসি দেখা দেয়। সে হাসি গাম্ভীর্যের মধ্য হইতে ফুটিয়া বাহির হয় তাহাই যথার্থ সুন্দর। এই কারণে সন্ধ্যার খেলা বড় সুন্দর, কেননা তাহার হাসি গাম্ভীর্যের মধ্য হইতে বিকশিত হয়। আঁধার এইরূপ সুন্দর বলিয়াই সন্ধ্যা মানুষের প্রিয়।

সমস্ত গগন নীল। কেবল দূরে দূরে একটি শুভ্র মেঘের টুকরা ধীরে ধীরে ভাসিয়া বেড়াইতেছে। সন্ধ্যার এই দৃশ্যে কি গম্ভীর ভাব, কি মধুর হাসি! মেঘগুলি যেন নীলিমার চাপা ঠোঁটের এক একটি হাসির টুকরা। তাহারা যেন চাপা থাকিতে না পারিয়া ফুটিয়া উঠিয়াছে। সন্ধ্যা এই হাসিগুলিকে সাজাইয়া দেয়—তাহাদের সৌন্দর্যকে আরও বৃদ্ধি করিয়া তুলে।

Figure 4: The grayscale image

In a grayscale image there are 256 combinations of black and white colors where 0 means pure black and 255 means pure white. This image is converted to binary image by checking whether or not each pixel value is grater than 255•level (level, found by Otsu's Method). If the pixel value is grater than or equal to 255•level then the value is set to 1 i.e. white otherwise 0 i.e. black. The binary image is shown in figure 5.



এই বিশ্ব জগতের স্তরে স্তরে কবিত্ব মাখানো। চরাচর শত সহস্র কবিতার সংগ্রহ। চন্দ্র, সূর্য, তারকারা এক একটি কবিতা। এই বিশ্ব জগৎ এই সকল ক্ষুদ্র ক্ষুদ্র কবিতার মালা। শত সহস্র উজ্জ্বল তারকা এই মালায় গ্রথিত হইয়া শোভা পাইতেছে—কত শত চন্দ্র-সূর্য এই মালার মধ্যে ফুলের আকারে বিরাজ করিতেছে। সৌন্দর্যের সূত্রেই চন্দ্র-তারকাবলী গাঁথা রহিয়াছে। সমস্ত দিবস মালাটির তেমন সুগন্ধ থাকে না। সন্ধ্যায় তাহার সুগন্ধ বিকশিত হয়— তাহার হাসি ফুটিয়া উঠে। সন্ধ্যার রক্তিমচ্ছটায় মালাটি হাসিতে থাকে। মনুষ্য হাসি ভালবাসে, সুগন্ধও সে চায়। তাই সে সন্ধ্যার কোলে বসিতে চায়। সন্ধ্যা তাহার হৃদয়ের ধন; সন্ধ্যায় সে ভাবুক।

মেঘশূন্য নির্মল আকাশের তলে সন্ধ্যায় বসিয়া থাকিতে কি আরাম। আমাদের হৃদয়ের সমস্ত মালিন্য দূর হইয়া গিয়া সেও যেন আকাশের মত নির্মল হয়। আমাদের প্রাণ আনন্দ উথলিয়া উঠে—মনে হয় সাঁজের তারার সহিত মিলিয়া এক হইয়া যাই। সন্ধ্যা আমাদের মনের সঙ্গে খেলা করিতে বড় ভালবাসে। সে মানব হৃদয়ে নানারকম খেলার ভাব তুলিয়া দেয়; কিন্তু তাহার খেলাগুলি গাম্ভীর্যমাখা, অথচ তাহার খেলায় গাম্ভীর্যের মধ্য দিয়া যেন ফুটফুটে হাসি দেখা দেয়। সে হাসি গাম্ভীর্যের মধ্য হইতে ফুটিয়া বাহির হয় তাহাই যথার্থ সুন্দর। এই কারণে সন্ধ্যার খেলা বড় সুন্দর, কেননা তাহার হাসি গাম্ভীর্যের মধ্য হইতে বিকশিত হয়। আঁধার এইরূপ সুন্দর বলিয়াই সন্ধ্যা মানুষের প্রিয়।

সমস্ত গগন নীল। কেবল দূরে দূরে একটি শুভ্র মেঘের টুকরা ধীরে ধীরে ভাসিয়া বেড়াইতেছে। সন্ধ্যার এই দৃশ্যে কি গম্ভীর ভাব, কি মধুর হাসি! মেঘগুলি যেন নীলিমার চাপা ঠোঁটের এক একটি হাসির টুকরা। তাহারা যেন চাপা থাকিতে না পারিয়া ফুটিয়া উঠিয়াছে। সন্ধ্যা এই হাসিগুলিকে সাজাইয়া দেয়—তাহাদের সৌন্দর্যকে আরও বৃদ্ধি করিয়া তুলে।

Figure 5: The binary image

Chapter 4: Skew angle detection and correction

As mentioned in chapter 2, most of the Bangla character has headline (matra) and so the skew angle can be detected using this matra. Some important statistics of Bangla language found from [2] and [3] are:

1. The average length of Bangla words is about six characters.
2. About 30%-35% of characters are vowel modifiers which, being small in size, contribute very little to the head line of the word.
3. Most basic characters are consonants, as vowels in basic form can appear at the beginning of the word or when two vowels appear side by side.
4. Compound characters are very infrequent, occurring in about 5% of the cases only.
5. In Bangla 41 characters can appear in the first position of a word. Out of these 41 characters 30 of them have head lines.
6. Probability of getting a character with head line in the first position of a word is: $P_1 = \frac{30}{41}$ and getting a character without head line in the first position is: $p_1^{'} = 1 - P_1 = \frac{11}{41}$.
7. In other positions of a word, there are mostly consonants and 28 out of 39 Bangla consonants have head lines.
8. Probability of getting a consonant with head line for other positions in a word is: $P_2 = \frac{28}{39}$ and probability of getting a character without head line in other positions is: $p_2^{'} = 1 - P_2 = \frac{11}{39}$.
9. Thus, probability of all four characters without head line in a word is $(1 - P_1)(1 - P_2)^3 = 0.00601$ (assuming that all characters are equally likely and independently occurring in a word). Hence, probability that a word will have at least one character with head line is 1 - 0.00601 = 0.99399. The practical situation is better than these estimates since characters

are not equally likely in a word and most frequently used characters have head lines.

In Bangla, head line connects almost all characters in a word; therefore we can detect a word by the method of connected component labeling [5] (Glossary [a]). As mentioned in [2], for skew angle detection, at first the connected component labeling is done.

At the time of component labeling, for each labeled component its bounding box (minimum upright rectangle containing the component) is defined. The mean b_m and standard deviation b_s of the bounding box width are also computed. Next, components having boundary box width greater than or equal to b_m and less than $b_m + 3b_s$ are preserved. By threshold at b_m the small components like dots, punctuation marks, isolated characters and characters without head line are mostly filtered out while by threshold at $b_m + 3b_s$ big components that may represent graphs and tables are also filtered out (Figure 6). Because of these filtering processes the irrelevant components can not create error in skew estimation.

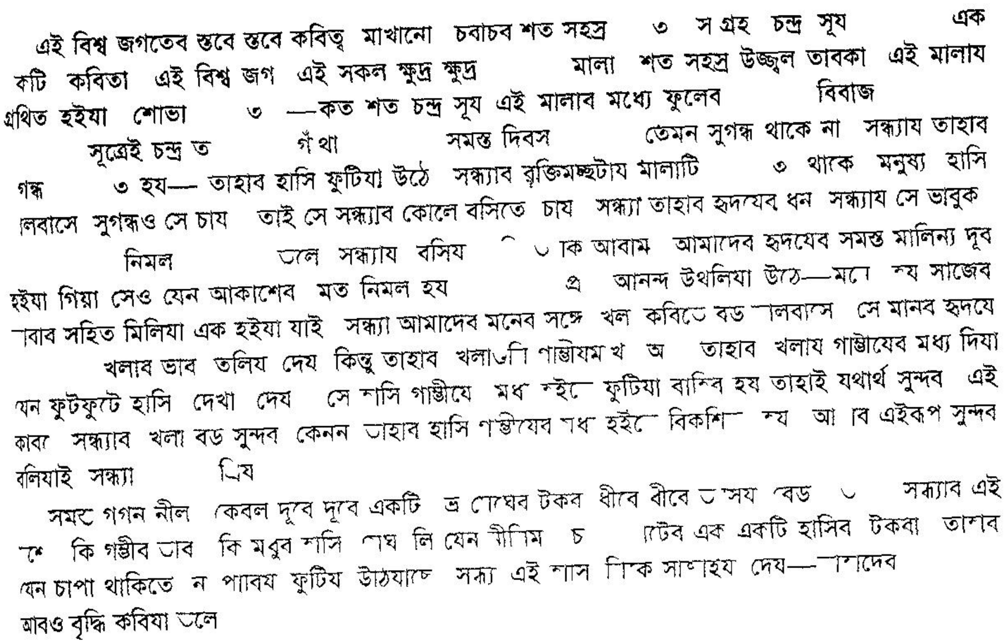


Figure 6: Selected connected components

Then *upper envelope* of the selected components is found as mentioned in [2]. From each pixel of the uppermost row of a bounding box, a vertical scan is performed until a pixel is found. The set of pixels obtained in this way denotes the upper envelope of the component. In figure 7 we show the Upper Envelope of figure 6. Note that in most of the cases the upper envelope contains the head line. In this way irrelevant data could be filtered out for further processing. Here Radon transform (glossary [b]) technique is applied on the upper envelopes for skew estimation.

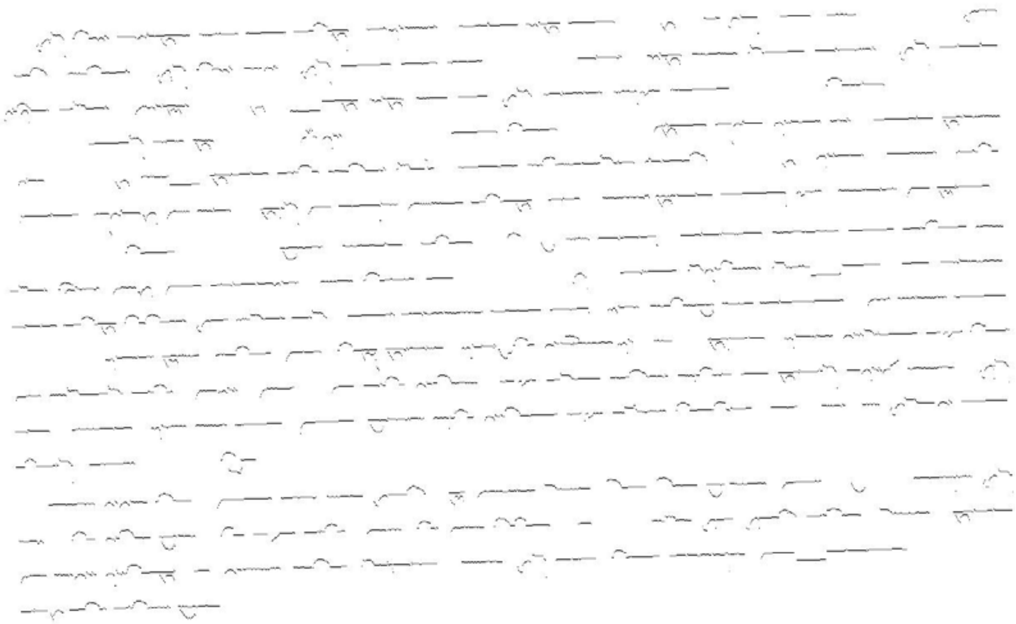


Figure 7: Upper Envelope

The Radon transform of a function f(x,y) is defined as the integral along a straight line defined by its distance *P* from the origin and its angle of inclination θ, a definition very close to that of the Hough transform (glossary [c]) and requires a lot of processing power in order to be able to do its work in a reasonably finite time. Now-a-days high processing power is not a problem. All processor speed in the market is now more than one GHz and main memory is also very cheap. So, to use Radon Transform will not be a problem.

The Radon transform can detect a line in any angle. It is considered that all the line in a single image has same skew angle and the range of this angle is from -10^o to 10^o. Here Radon transform will detect the angle from the upper envelope. If the skewed angle is more than 10^o or less then -10^o the upper envelope will contain 2 lines in different direction. An example is shown in Figure 8 having 50 degree of skewed angle which may create a problem.

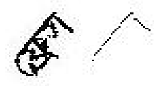


Figure 8: Upper envelope containing 2 lines in different direction

Radon transform gives the angles and distance of lines from the origin. The Radon transform of Upper Envelope (Figure 7) is shown in Figure 9.

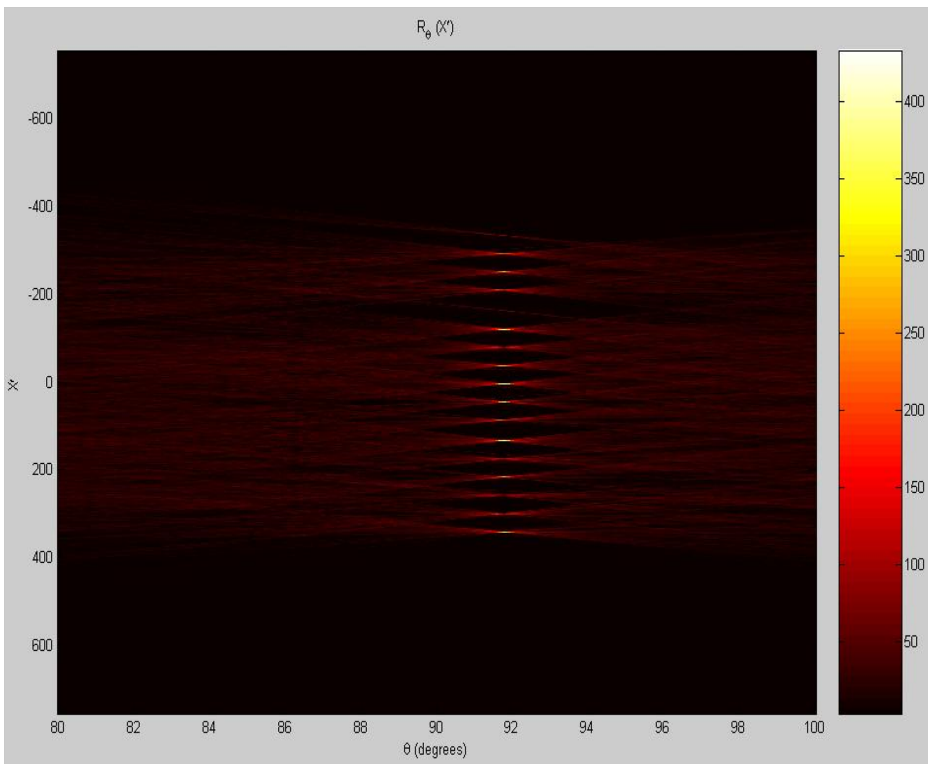
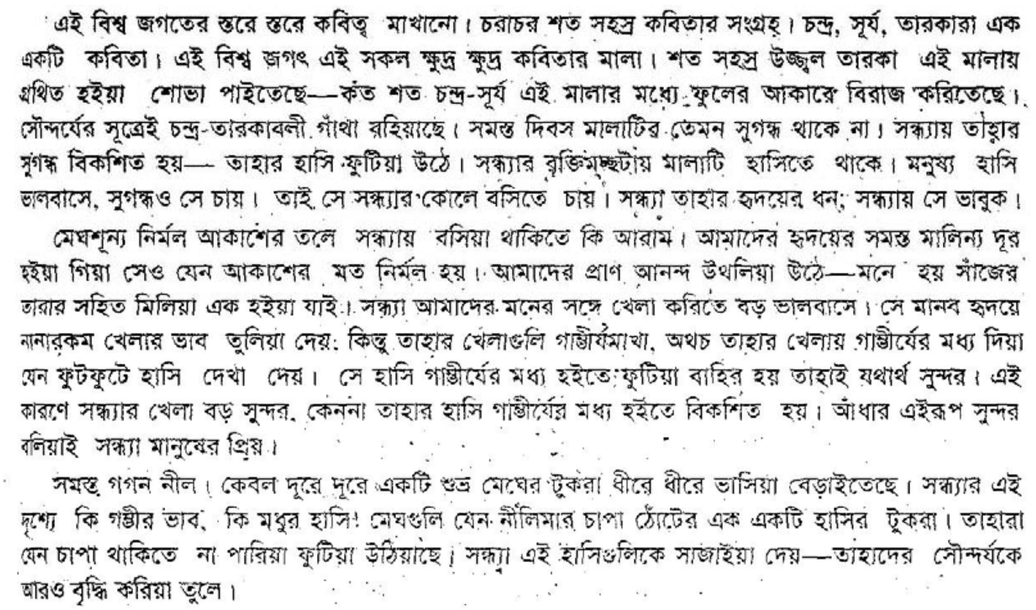


Figure 9: Radon transform of Upper Envelope (Figure 7)

In Figure 9, the point which has a high intensity represents the angle of a straight line. Here one can see that all the straight lines have the same degree of angle which is 91.8. To find the desired angle, this angle is subtracted from 90 degree and thus the desired skewed angle is 90 – 91.8 = –1.8. So, if the digital image (Figure 5) is rotated in degree –1.8, then the desired de-skewed image (Figure 10) is found. It should be mentioned that up to one digit after decimal place has been considered.


এই বিশ্ব জগতের স্তরে স্তরে কবিত্ব মাখানো। চরাচর শত সহস্র কবিতার সংগ্রহ। চন্দ্র, সূর্য, তারকারা এক একটি কবিতা। এই বিশ্ব জগৎ এই সকল ক্ষুদ্র ক্ষুদ্র কবিতার মালা। শত সহস্র উজ্জ্বল তারকা এই মালায় গ্রথিত হইয়া শোভা পাইতেছে—কত শত চন্দ্র-সূর্য এই মালার মধ্যে ফুলের আকারে বিরাজ করিতেছে। সৌন্দর্যের সূত্রেই চন্দ্র-তারকাবলী গাঁথা রহিয়াছে। সমস্ত দিবস মালাটির তেমন সুগন্ধ থাকে না। সন্ধ্যায় তাহার সুগন্ধ বিকশিত হয়— তাহার হাসি ফুটিয়া উঠে। সন্ধ্যার রক্তিমচ্ছটায় মালাটি হাসিতে থাকে। মনুষ্য হাসি ভালবাসে, সুগন্ধও সে চায়। তাই সে সন্ধ্যার কোলে বসিতে চায়। সন্ধ্যা তাহার হৃদয়ের ধন; সন্ধ্যায় সে ভাবুক।

মেঘশূন্য নির্মল আকাশের তলে সন্ধ্যায় বসিয়া থাকিতে কি আরাম। আমাদের হৃদয়ের সমস্ত মালিন্য দূর হইয়া গিয়া সেও যেন আকাশের মত নির্মল হয়। আমাদের প্রাণ আনন্দ উথলিয়া উঠে—মনে হয় সাঁজের তারার সহিত মিলিয়া এক হইয়া যাই। সন্ধ্যা আমাদের মনের সঙ্গে খেলা করিতে বড় ভালবাসে। সে মানব হৃদয়ে নানারকম খেলার ভাব তুলিয়া দেয়; কিন্তু তাহার খেলাগুলি গাম্ভীর্যমাখা, অথচ তাহার খেলায় গাম্ভীর্যের মধ্য দিয়া যেন ফুটফুটে হাসি দেখা দেয়। সে হাসি গাম্ভীর্যের মধ্য হইতে ফুটিয়া বাহির হয় তাহাই যথার্থ সুন্দর। এই কারণে সন্ধ্যার খেলা বড় সুন্দর, কেননা তাহার হাসি গাম্ভীর্যের মধ্য হইতে বিকশিত হয়। আঁধার এইরূপ সুন্দর বলিয়াই সন্ধ্যা মানুষের প্রিয়।

সমস্ত গগন নীল। কেবল দূরে দূরে একটি শুভ্র মেঘের টুকরা ধীরে ধীরে ভাসিয়া বেড়াইতেছে। সন্ধ্যার এই দৃশ্যে কি গম্ভীর ভাব, কি মধুর হাসি! মেঘগুলি যেন নীলিমার চাপা ঠোঁটের এক একটি হাসির টুকরা। তাহারা যেন চাপা থাকিতে না পারিয়া ফুটিয়া উঠিয়াছে। সন্ধ্যা এই হাসিগুলিকে সাজাইয়া দেয়—তাহাদের সৌন্দর্যকে আরও বৃদ্ধি করিয়া তুলে।

Figure 10: After correction the skew angle

Chapter 5: Line and word separation

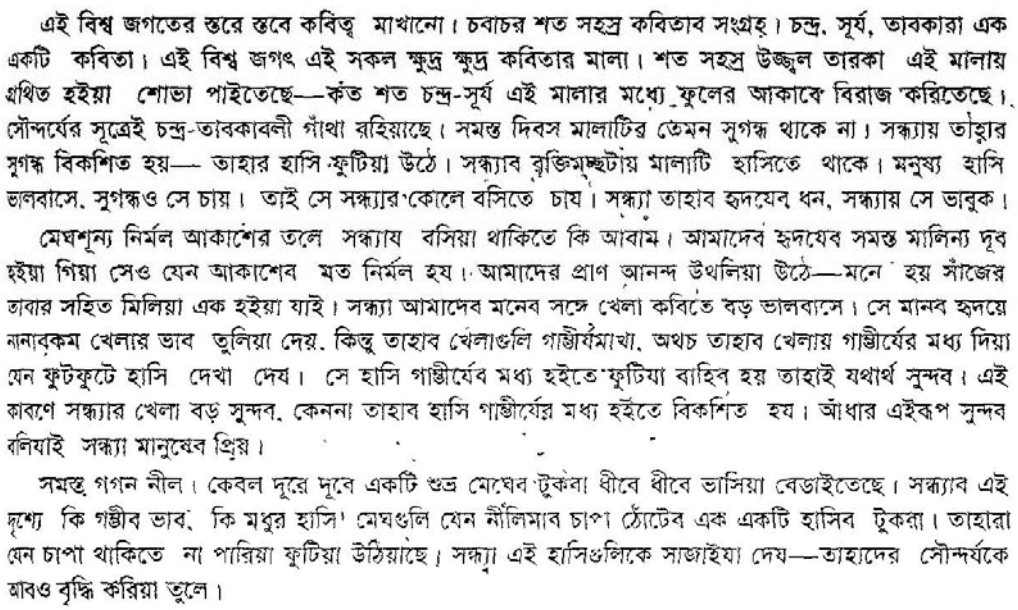
Before separation of lines, some noise must be removed. There are lots of dots in the image (figure10) which may create problem. When removing a dot, one should remember that there are some bangle characters that contain dot. Figure 11 shows such characters. For this it is very difficult to locate the noise dots.

i q o

Figure 11: Dots in bangle character

5.1 Noise removal

Generally a noise dot is very small in size. Here a dot having size 4 by 4 or less is removed. The image after removal of small dots is shown in figure 12.



এই বিশ্ব জগতের স্তরে স্তবে কবিত্ব মাখানো। চবাচর শত সহস্র কবিতাব সংগ্রহ। চন্দ্র, সূর্য, তাবকারা এক একটি কবিতা। এই বিশ্ব জগৎ এই সকল ক্ষুদ্র ক্ষুদ্র কবিতার মালা। শত সহস্র উজ্জ্বল তারকা এই মালায় গ্রথিত হইয়া শোভা পাইতেছে—কত শত চন্দ্র-সূর্য এই মালার মধ্যে ফুলের আকাবে বিরাজ করিতেছে। সৌন্দর্যের সূত্রেই চন্দ্র-তাবকাবলী গাঁথা রহিয়াছে। সমস্ত দিবস মালাটির তেমন সুগন্ধ থাকে না। সন্ধ্যায় তাহার সুগন্ধ বিকশিত হয়— তাহার হাসি ফুটিয়া উঠে। সন্ধ্যাব রক্তিমচ্ছটায় মালাটি হাসিতে থাকে। মনুষ্য হাসি ভালবাসে, সুগন্ধও সে চায়। তাই সে সন্ধ্যার কোলে বসিতে চায়। সন্ধ্যা তাহাব হৃদয়েব ধন, সন্ধ্যায় সে ভাবুক।

মেঘশূন্য নির্মল আকাশের তলে সন্ধ্যায বসিয়া থাকিতে কি আবাম। আমাদেব হৃদযেব সমস্ত মালিন্য দূব হইয়া গিয়া সেও যেন আকাশেব মত নির্মল হয। আমাদের প্রাণ আনন্দ উথলিয়া উঠে—মনে হয় সাঁজের তাবার সহিত মিলিয়া এক হইয়া যাই। সন্ধ্যা আমাদেব মনেব সঙ্গে খেলা কবিতে বড় ভালবাসে। সে মানব হৃদয়ে নানাবকম খেলার ভাব তুলিয়া দেয়, কিন্তু তাহাব খেলাগুলি গাম্ভীর্যমাখা, অথচ তাহাব খেলায় গাম্ভীর্যের মধ্য দিয়া যেন ফুটফুটে হাসি দেখা দেয। সে হাসি গাম্ভীর্যেব মধ্য হইতে ফুটিয়া বাহিব হয় তাহাই যথার্থ সুন্দব। এই কাবণে সন্ধ্যার খেলা বড় সুন্দব, কেননা তাহাব হাসি গাম্ভীর্যের মধ্য হইতে বিকশিত হয। আঁধার এইরূপ সুন্দব বলিযাই সন্ধ্যা মানুষেব প্রিয়।

সমস্ত গগন নীল। কেবল দূরে দূবে একটি শুভ্র মেঘেব টুকবা ধীবে ধীবে ভাসিয়া বেড়াইতেছে। সন্ধ্যাব এই দৃশ্যে কি গম্ভীব ভাব, কি মধুর হাসি! মেঘগুলি যেন নীলিমাব চাপা ঠোঁটেব এক একটি হাসিব টুকরা। তাহারা যেন চাপা থাকিতে না পারিয়া ফুটিয়া উঠিয়াছে। সন্ধ্যা এই হাসিগুলিকে সাজাইযা দেয—তাহাদের সৌন্দর্যকে আবও বৃদ্ধি করিয়া তুলে।

Figure 12: After remove dots having size 4 by 4 or less

In this image there are still some noise dots having size more. If these dots are removed, then some character will be also be lost (e.g. coma, colon etc.).

5.2 Line Segmentation

Line segmentation is performed using the free spaces between each line. Details of this part can be found on the thesis report of my group member.

5.3 Word Segmentation

Word segmentation is also performed using the free spaces; in this case spaces between each word are considered. Details of this part can be found on the thesis report of my group member.

5.4 Character segmentation

This is one of the most challenging parts in pre-processing steps. Generally characters are connected in a word by the head line (matra). To segment characters, first the headline has to be removed. For detection of headlines one of the morphological operations[4], which is thinning, has been used. Details of this part can be found on the thesis report of my group member.

---

[4] Morphological operation is a collection of techniques for digital image processing. Such techniques include closing, shrinking, thinning, thickening, skeletonization, pruning, etc.

Chapter 6: Character Recognition

The whole character itself is used as the neural network input. The size (25 by 25) of the input image is kept fixed. For character images larger than 25 by 25 in size, a resizing operation is performed to convert it to 25 by 25 keeping the aspect ration[5] preserved. On the other hand, for character images less than 25 by 25, the characters are just put into the 25 by 25 grid from top left corner. For example, if the character image was of size 22 by 18, then it will be placed into a 25 by 25 grid from top left corner and the remaining blank grids will be filled by 0s. This is done because if a smaller image is resized to fit a bigger grid, then the image itself becomes too distorted for proper recognition.

The character recognition process is summed up in the figure. The preprocessing steps have been explained in the previous chapters.

[5] The aspect ratio of an image is its displayed width divided by its height

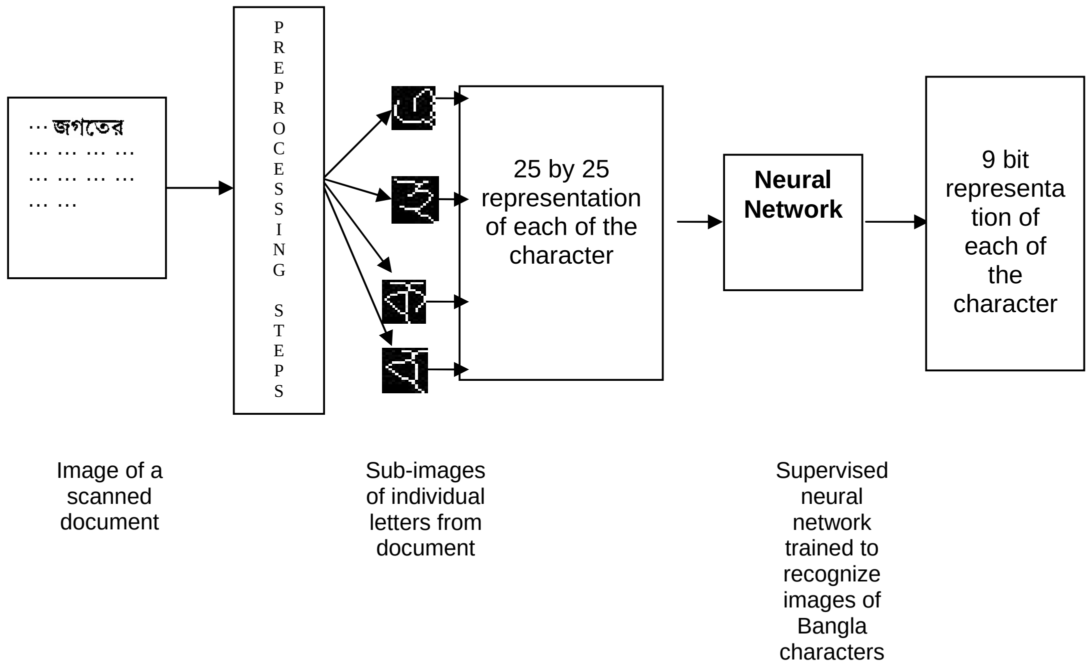


Figure 13: Character recognition process

The network that has been used is a three layer feed forward neural network. Feed forward networks may have a single layer of weights, where the inputs are directly connected to the outputs, or multiple layers with intervening sets of hidden units. Neural networks use hidden units to create internal representations of the input patterns. In fact, it has been shown that given enough hidden units it is possible to approximate arbitrarily closely almost any function with a simple feed forward network. This result has encouraged me to use a multi layer feed forward neural network to solve the problem of recognition for Bangla scripts.

In the feed forward architecture, the activations of the input units are set and then propagated through the network until the values of the output units are determined. The network acts as a vector-valued function taking one vector on the input and returning another vector on the output. For instance, in case of this thesis the input vector might represent the characteristics of one of the letters (say ‘ka’) and the output might be a prediction of whether that character is likely to be ‘ka’. Or the inputs might represent the characteristics of a class member and the output might be a prediction of the class to which that character belongs.

The network functions as follows:

1. Each neuron receives a signal from the neurons in the previous layer, and each of those signals is multiplied by a separate weight value.
2. The weighted inputs are summed, and passed through a limiting function, which scales the output to a fixed range of values.
3. The output of the limiter is then broadcast to all of the neurons in the next layer.

So, to use the network to solve the recognition problem, the input values have been applied to the inputs of the first layer, allowing the signals to propagate through the network, and the output values are read.

Values acquired after preprocessing are fed into the layer downstream the input layer (the hidden layer). Once the neurons for the hidden layer are computed, their activations are then fed downstream to the next layer, until all the activations eventually reach the output layer, in which each output layer neuron is associated with a specific classification category. In a fully connected multi-layer feed forward network, each neuron in one layer is connected by a weight to every neuron in the layer downstream it. A bias is also associated with each of these weighted sums. Thus in computing the value of each neuron in the hidden and output layers first the sum of the weighted sums and the bias is taken and then f(sum), the sigmoid function, is applied to calculate the neuron's activation.

Training basically involves feeding training samples as input vectors through a neural network, calculating the error of the output layer, and then adjusting the weights of the network to minimize the error. Each "training epoch" involves one exposure of the network to a training sample from the training set, and adjustment of each of the weights of the network once layer by layer. Selection of training samples from the training set has been taken in random.

Training can stop when the network error dips below a particular error threshold (a threshold of .01 squared error has been chosen in this case).

Architecture of the network is as follows:

- Input layer: 625 neurons
- 1st Hidden layer: 100 neurons
- 2nd Hidden layer: 50 neurons
- Output layer: 9 neurons (for each character)
- Transfer function: log-sigmoid
- Incremental training algorithm, standard back-propagation method (Glossary [d])

The binary representation that has been used for characters are presented in table 3.

Table 3: Binary reparation for characters

000001001	9	০		000101101	45	ঙ
000001010	10	১		000101110	46	চ
000001011	11	২		000101111	47	ছ
000001100	12	৩		000110000	48	জ
000001101	13	৪		000110001	49	ঝ
000001110	14	৫		000110010	50	ঞ
000001111	15	৬		000110011	51	ট
000010000	16	৭		000110100	52	ঠ
000010001	17	৮		000110101	53	ড
000010010	18	৯		000110110	54	ঢ
000010011	19	অ		000110111	55	ন
000010100	20	আ		000111000	56	ত
000010101	21	ই		000111001	57	থ
000010110	22	ঈ		000111010	58	দ
000010111	23	উ		000111011	59	ধ
000011000	24	ঊ		000111100	60	ণ
000011001	25	ঋ		000111101	61	প
000011010	26	এ		000111110	62	ফ
000011011	27	ঐ		000111111	63	ব

000011100	28	ও		001000000	64	ভ
000011101	29	ঔ		001000001	65	ম
000011110	30	া		001000010	66	য
000011111	31	ি		001000011	67	র
000100000	32	ী		001000100	68	ল
000100001	33	ু		001000101	69	শ
000100010	34	ূ		001000110	70	ষ
000100011	35	ৃ		001000111	71	স
000100100	36	ে		001001000	72	হ
000100101	37	ৈ		001001001	73	ড়
000100110	38	ৌ		001001010	74	ঢ়
000100111	39	¨		001001011	75	য়
000101000	40			001001100	76	r
000101001	41	ক		001001101	77	ং
000101010	42	খ		001001110	78	t
000101011	43	গ		001001111	79	
000101100	44	ঘ				

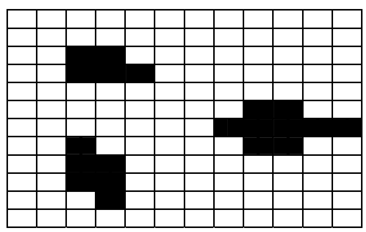
Chapter 7: Conclusion

This thesis tries to suggest an approach for the recognition of Bangla scripts. Approaches suggested from the beginning of scanning a document to converting it to binary image, skew detection and correction, line separation and word segmentation has been successfully stated. One of the challenges faced in the character segmentation part is that two characters are sometimes joined together. There are even cases where a single character breaks apart. Solutions to these challenges are likely to be presented in future. In our current approach, the whole character itself was used as a feature. In future implementation feature extraction will be more comprehensive.

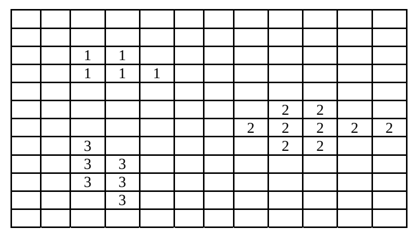
Glossary

[a] Connected Components Labeling

A set of pixels in which each pixel is connected to all other pixel is called a connected component. A component labeling algorithm finds all connected components in an image and assign a unique label to all points in the same component.



(a)



(b)

Sequential Connected Components Algorithm 8-connectivity

Step 1: Scan image left to right, top to bottom.

Step 2: If the pixel is black

(a) If only one of its upper-left or upped or upper-right or left neighbors has a label, then copy the label.

(b) If both have the same label, then copy the label

(c) If both have labels and they are not same, then copy the lowest labels of its neighbors and enter the labels in the equivalence table as equivalent labels.

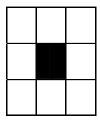
(d) Otherwise assign a new label to this pixel and enter this label in the equivalence table.

Step 3: If there are more pixels to consider, then go to step 2.

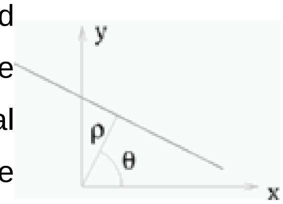
Step 4: Find the lowest label for each equivalent set in the equivalence table.

Step 5: Scan the picture. Replace each label by the lowest label in its equivalent set.

Here 8-connectivity means there are 8 neighbored pixel of a particular pixel shown as image below:



[b] The Radon Transform



In recent years the Hough transform and the related Radon transform have received much attention. These two transforms are able to transform two dimensional images with lines into a domain of possible line parameters, where each line in the image will give a peak positioned at the corresponding line parameters. This has lead to many line detection applications within image processing, computer vision, and seismic.

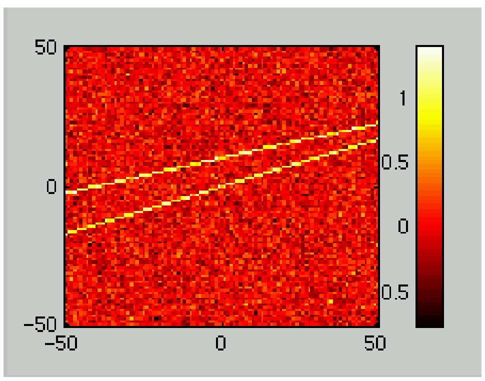
Several definitions of the Radon transform exists, but the are related, and a very popular form expresses lines in the form $\rho = x\cos\theta + y\sin\theta$, where θ is the angle and ρ the smallest distance to the origin of the coordinate system. As shown in the two following definitions (which are identical), the Radon transform for a set of parameters (ρ,θ) is the line integral through the image g(x, y), where the line is positioned corresponding to the value of (ρ,θ). The δ( ) is the delta function which is infinite for argument 0 and zero for all other arguments (it integrates to one).

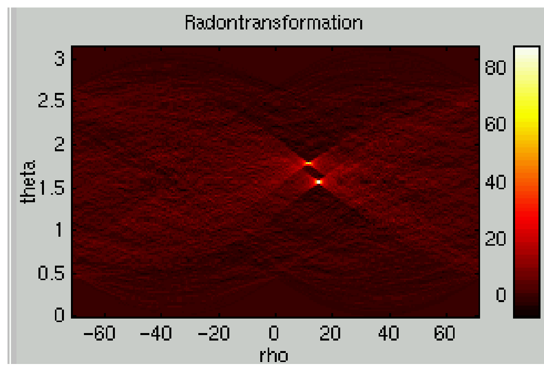
$$\breve{g}(\rho,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\delta(\rho - x\cos\theta - y\sin\theta)\ dx\ dy$$

or the identical expression

$$\breve{g}(\rho,\theta) = \int_{-\infty}^{\infty} g(\rho\cos\theta - s\sin\theta, \rho\sin\theta + s\cos\theta)\ ds$$

Using this definition an image containing two lines are transformed into the Radon transform shown to the right





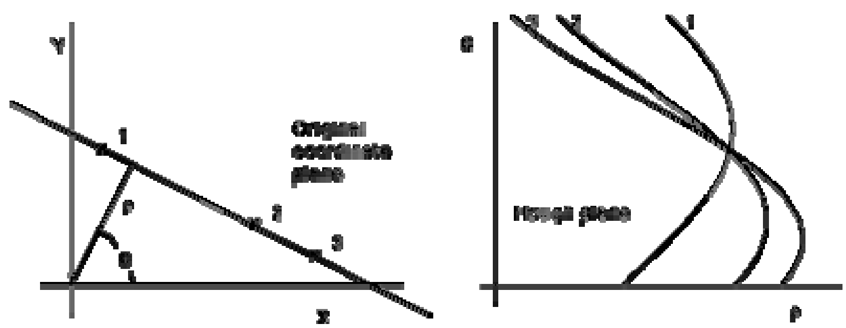
It can be seen that two very bright spots are found in the Radon transform, and the positions shown the parameters of the lines in the original image. A simple threshold algorithm could then be used to pick out the line parameters, and given that the transform is linear many lines will just give rise to a set of distinct point in the Radon domain. In Ph.D. thesis [3], Peter Toft investigated the relationship of Radon transform with the Hough transform, and it is shown that the Radon transform and the Hough transform are related but NOT the same.

[c] Hough transform

The Hough transform is a standard tool in image analysis that allows recognition of global patterns in an image space by recognition of local patterns (ideally a point) in a transformed parameter space. It is particularly useful when the patterns one is looking for are sparsely digitized, have ``holes'' and/or the pictures are noisy.

The basic idea of this technique is to find curves that can be parameterized like straight lines, polynomials, circles, etc., in a suitable parameter space. Although the transform can be used in higher dimensions the main use is in two dimensions to find, e.g. straight lines, centers of circles with a fixed radius, parabolas *y* = *ax*^{2} + *bx* + *c* with constant *c*, etc.

As an example consider the detection of straight lines in an image. We assume them parameterized in the form: $\rho = x\cos\theta + y\sin\theta$, where ρ is the perpendicular distance from the origin and θ the angle with the normal. Collinear points (x_i, y_i), with $i = 1,...,N$, are transformed into *N* sinusoidal curves $\rho = x_i \cos\theta + y_i \sin\theta$ in the (ρ,θ) plane, which intersect in the point (ρ,θ).



Care has to be taken when one quantizes the parameter space (ρ,θ). When the bins of the (ρ,θ) space (it is easy to visualize the transform as a two-dimensional histogram) are chosen too fine, each intersection of two sinusoidal curves can be in a different bin. When the quantization is not fine enough, on the other hand, nearly parallel lines which are close together will lie in the same bin.

For a certain range of quantized values of parameters ρ and θ, each (x_i, y_i) is mapped into the (ρ,θ) space and the points that map into the locations (ρ_m,θ_m) are accumulated in the two-dimensional histogram, IHIST (ρ_m,θ_m), i.e. IHIST (ρ_m,θ_m) = IHIST $(\rho_m,\theta_m)+1$.

If a grey level image $g(x, y)$ is given, and g_i is the grey value at the point (x_i, y_i) the grey values are accumulated: IHIST(ρ_m,θ_m) = IHIST $(\rho_m,\theta_m)+g_i$.

[d] Back Propagation Algorithm:

Step 1: Feed the input vector through the network and compute every unit in the network. This is done by computing the weighting sum coming into the unit and then applying the sigmoid function. The 'x' vector is the activation of the previous layer.

$$o = \sigma(\vec{w}\vec{x})$$
$$\sigma(y) = \frac{1}{1+e^{-y}}$$

The 'w' vector denotes the weights linking the neuron unit to the previous neuron layer.

Step 2: Compute the squared error of the network. This is done by taking the sum of the squared error of every unit in the output layer. The target vector involved is associated with the training sample (the input vector).

$$E(\vec{w}) = \frac{1}{2}\sum_{k \in outputs}(t_k - o_k)^2$$

't' denotes a target value in the target vector, and 'o' denotes the activation of a unit in the output layer.

Step 3: Calculate the error term of each output unit, indicated below as 'delta'.

$$\delta_k \leftarrow o_k(1 - o_k)(t_k - o_k)$$

The error term is related to the partial derivative of each weight with respect to the network error.

Step 4: Calculate the error term of each of the hidden units.

$$\delta_h \leftarrow o_h(1 - o_h) \sum_{k \in outputs} w_{kh} \delta_k$$

The hidden unit error term depends on the error terms calculated for the output units.

Step 5: Compute the weight deltas. 'Eta' here is the learning rate. A low learning rate can ensure more stable convergence. A high learning rate can speed up convergence in some cases.

$$\Delta w_{ji} = \eta \delta_j x_{ji}$$

'x' denotes the unit that's connected to the unit downstream by the weight 'w'

Step 6: Add the weight deltas to each of the weights. Adjusting the weights one layer at a time is preferred. This method involves re-computing the network error before the next weight layer error terms are computed.

$$w_{ji} \leftarrow w_{ji} + \Delta w_{ji}$$

Once finished, proceed back to step 1.

References

Papers presented at conferences (Published)

[1] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62-66, 1979.

[2] B.B. Chaudhuri and U. Pal, “Skew Angle Detection of Digitized Indian Script Documents”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 2, February 1997

[3] B.B. Chaudhuri and U. Pal, “Relational Studies Between Phoneme and Grapheme Statistics in Modern *Bangla* Language,” *J. Acoustical Society of India*, vol. 23, pp. 67-77, 1995.

Thesis

[4] Peter Toft: "The Radon Transform - Theory and Implementation", Ph.D. thesis.Department of Mathematical Modelling, Technical University of Denmark, June 1996. 326 pages.

Books:

[5] Ramesh Jain and Rangachar Kasturi and Brian G. Schunck, Machine Vision, International Edition 1995 McGraw-Hill, Inc.