Handwritten Text Documents Binarization and Skew Normalization Approaches

Subhash Panwar
Malaviya National Institute of Technology, Jaipur
Email: panwar.subhash@gmail.com

Neeta Nain
Malaviya National Institute of Technology, Jaipur
Email: neetanain@yahoo.com

***Abstract*—Handwritten text recognition has been an active research area for many years. Handwritten text recognition needs to perform some preprocessing steps for better recognition. First, we find binary image of given handwritten text document and then after performing the line segmentation task, we normalize it to the segmented lines. There are various normalization tasks such as skew normalization, slant normalization and size normalization. This paper, focuses on the handwritten document binarization and skew normalization and proposes a novel global binarization approach, which is very cost effective. We also propose a new skew normalization approach which is based on orthogonal projection of the segmented line with respect to x-axis. The method has been experimented on various styles of handwritten text documents, and it is found that it detects the exact skew angle, and corrects it efficiently. A comparative study has also been reported to provide a detailed analysis of the proposed methods together with some other existing methods in the literature.**

***Index Terms*—handwritten text document, preprocessing, normalization, binarization**

I. Introduction

Handwritten text recognition has been a subject of research for several decades. Generally a handwritten text recognition includes four main tasks: preprocessing, segmentation, feature extraction, and classification. In preprocessing step researchers normally perform noise removal, binarization, normalization (as skew, slant and size) etc., to enhance the quality of images and correct the distortion so that the recognition rate could be improved. This is followed by segmentation task (to extract the line), where various segmentation methods are available as projection profile, Hough transformation and smearing methods etc. In feature extraction, the features are extracted for segmented image using various techniques as statistical features, structural features etc. Finally for classification, various classification approach are available as neural networks, support vector machine, multiple classifier etc.

Although the recognition rate is dependent upon the proper features extraction for particular word or character and also on proper selection of classification method. But from many observations and experimentations, we analyse that binarization of an original Gray level image improves the recognition rate and also reduces the computation cost.

The main goal of binarization is to convert a Gray scale input image into a binary image, because many vision algorithms and operators only handle the binary image rather Gray scale image. Selection of a global threshold value is a general technique to convert Gray scale image into binary image. Such general methods binarize the entire image using a single threshold value. One simple way is to automatically select the value at the valley of intensity histogram of the image, assuming that there are two peaks in the histogram, one corresponding to the foreground, the other to the background. Poor contrast and strong noise in the input image is also a challenge for selection of proper threshold value because due to the poor contrast, many images do not have such two peaks in the histogram.

Various binarization approaches are used in handwritten text recognition system. In most of the systems researchers generally use the Otsu [8] method for binarization.

In this paper, a novel global binarization approach is proposed which is particularly for handwritten text document. Though it we can handle printed text as well. It is specifically very simple with constant complexity and uses only addition operation, hence it is very efficient for handwritten documents. For example, for banking and postal applications, like cheque detail recognition and postal address recognition etc.

The technique uses the I_{max}-maximum and I_{min}-minimum intensity levels of a given handwritten document. We know that all minimum intensity pixels with some variations $(+I_O)$ (intensity - offset) made by pen, are our foreground pixels and rest is background. Basically two facts are used for this novel approach: (a) for proper visualization or readability of the document a writer make a common sense in selecting a good pen. Which always distinctly differentiate between the background and foreground (handwritten content). If the background is dark then he will use a light pen and if the background is light then he will use a dark pen. So we assume that our text has a single and higher intensity text and rest part is assumed to be background. Taking our example further we can say that generally a writer writes or fill the cheque details in a cheque and fill the postal address in post card by a single pen; (b) when a user uses a single pen then the variation of intensity depends on the type of pen which is used by writer. The variation of intensity is also consider in our approach as average for both dark and light pen.

Skew normalization is required to make a standard form of the character for better classification. The majority of both segmentation and character recognition algorithms are also sensitive to the orientation of the word. Furthermore, the skewed words are very often found in handwritten text. Even

978-1-4673-4369-5/12/$31.00 ©2012 IEEE

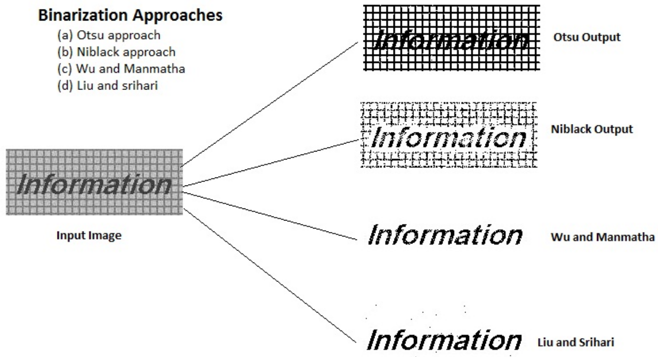

Fig. 1. Comparative outputs of various binarization approaches

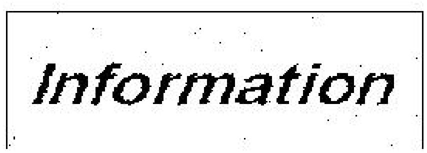

Fig. 2. Output of proposed binarization approach for input image of figure 1

in the case of correctly oriented pages, the handwritten words could present smaller or larger skews. Technically, skew is defined as the deviation of the base lines of the text from the horizontal direction. An example of skewed word image can be found in Figure 4 . At the time of the recognition, handwritten text should be free of skew for good recognition with better computational cost.

Several methods have been proposed for skew detection and correction of document images. Many methods have used projection profiles for skew detection [1]. In general, projection profile methods are limited to estimate skew angle within ± 10 degrees [1]. Some methods are based on Hough Transform [3] [4]. Hough Transform is applied to find the best direction of the straight lines in documents. Another group of skew correction methods, uses k-nearest neighbour clustering on connected components, and they try to find the best overall direction of neighbouring components [6]. See [2] for a survey on skew correction methods. Some methods also entail high computational cost such as Hough transform methods. Some other methods only deal with small skew angles. For example, although projection profiles can be easily constructed, due to high computational cost of exhaustive search, the range of search for skew angles is usually restricted to ± 10 degrees.

Second part of this paper is to proposed a simple and novel approach for Skew normalization for handwritten script. The concept of orthogonal projection with respect to x-axis is used in the skew detection. The orthogonal projection of the word image is maximum with un-skewed or normal text line or word image.

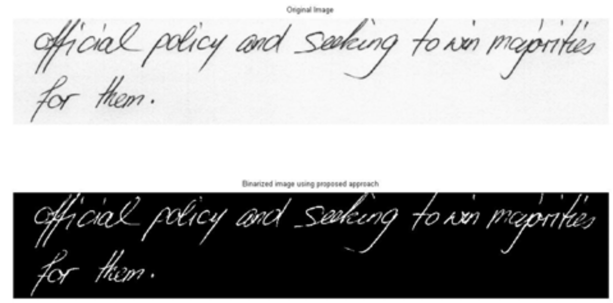

Fig. 3. Handwritten document image and its Binary image as output of proposed approach

As Figure 5 show that the $|OP|$ is the actual length (A_l) of line which is shown with the different skew angles about x-axis as $0, r', r''$ and 90 degrees. The relation between the A_l of segmented text line and projected length (P_l) with respect to skew angle θ is given as

$$P_l = (A_l \times \cos\theta)$$

But in real scenario a segmented text line have some height h and its projectionδ is also included in the projection calculation The value of δ depends on the height of text line h,

$$\delta = h \times \cos(90 - \theta)$$

where, $\delta \in [0, h]$. as shown in Figure 6.
So the actual projection reduced by value of δ.
And in Figure 5 where $\delta = 0$ because $h = 0$
So we can say that,

$$OP = Oq' \times \cos r'$$

If skew angle θ as $r' = 0$, then $\cos\theta = \cos r' = 1$, so

$$P_l = A_l$$

and hence, $Op = Oq'$
If Skew angle is, 90 $degree$, then $\cos 90 = 0$, so

$$P_l = 0$$

So the Figure 5 show the orthogonal projection on the x-axis of each skewed line OP, OP', OP'', OP''', which are same as $|OP|$, is given by $OP, Oq', Oq'', 0$ respectively.

Using this concept, we rotate the input segmented text line image till reach to maximum projected length.

The remaining part of the paper is organized as follows. In section 2 various existing binarization and skew detection and removal approaches are described. Section 3 presents our proposed binarization and skew normalization approaches. Section 4 explain the comparative analysis with existing approaches, followed by conclusions in Section 5.

Fig. 4. Example of a skewed image

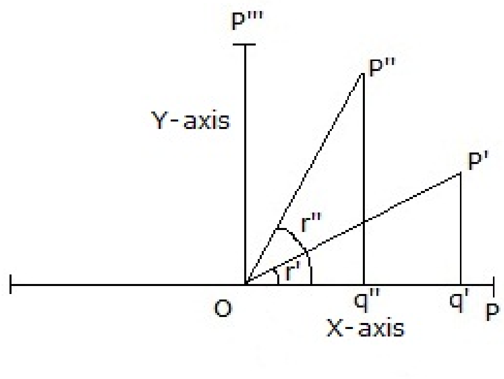

Fig. 5. Orthogonal projection of skewed text line.

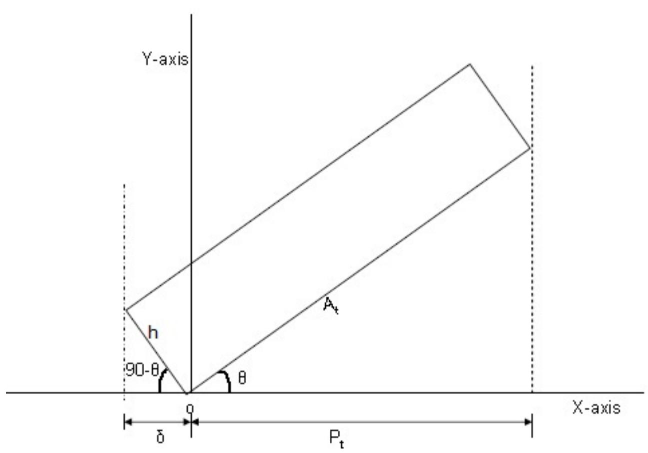

Fig. 6. Orthogonal projection of image of height h with effect of δ introduced in skew angle correction.

II. Previous Approaches

A. Previous Binarization Approaches

This section describes the four algorithms that exhibited the best performances in converting the Gray level into binary image. All those algorithms are based on Gray level image histogram threshold. The true colour to Gray level conversion is performed by:

$$gray = 0.299r + 0.587g + 0.114b$$

All the algorithms presented here take the image histogram and normalize each of its entries by the total number of pixels in the image, yielding a probability distribution provided by relative frequencies. Thus,

$$p_i = \frac{n_i}{N}, 0 \leq i \leq 255$$

$$P_i = \sum_{i=0}^{t} p_i$$

where n_i is the number of pixels with Gray level i, N is the total number of pixels in the image, $\{p_0, p_1, p_2......, p_{255}\}$ is the probability distribution of the pixel of Gray levels taking into accounts their relative frequencies, and P_t is the adding of all probabilities up to entry t.

1) Otsu's Approach: The algorithm given by Otsu [8] is a histogram based global thresholding algorithm. It is a global binarization method. It assume that the image to be threshold contain two groups of pixels(e.g. foreground and background) and it use the Gray level histogram for selection of threshold value and it proposed a criterion for maximizing the variance of the between class of intensity to perform thresholding.

2) Niblack Approach: Niblack approach [9] is based on the calculation of the local mean and of local standard deviation and it calculates the pixel wise threshold by sliding a rectangular window over the grey level image. The threshold T(X, Y) is decided by the formula

$$T(X, Y) = m(X, Y) + K * s(X, Y)$$

where m(X, Y) and s(X, Y) is average values and standard deviation values of all the pixels in the sliding window respectively and K is a empirical constant (value between 0 and 1). The size of sliding window and value of k defines the quality of binarization.

3) Wu and Manmatha Approach: The approach proposed by Wu Manmatha [10] consists two basic steps, in first step, the input image is smoothed using a low-pass (Gaussian) filter. The smoothing operation enhances the text relative to any background text because the background texture normally has higher frequency then text. In the second step, the intensity histogram of smoothed image is computed and a threshold automatically selected as follows. For black text, the first peak of histogram corresponds to text. Thresholding the image at value of valley between first and second peaks of the histogram binarizes the image well.

4) Liu and srihari Approach: Srihari [11] algorithm consists of three steps first, candidate thresholds are produced through iterative use of Otsu's method [8] and then in second step texture features associated with each candidate threshold are extracted from the run-length histogram of the accordingly binarized image and at last the optimal threshold is selected so that desirable document texture features are preserved.

B. Previous Approaches for Skew normalization

Majority of writer fails to write in straight line and write in a particular angle from x-axis. This angle from x-axis is called the slope or skew. As writer may write all text in same slope or may write in multiple slope in same text document. So we can say that slope may be global, multiple and non-uniform. This slope can hurt the effectiveness of later handwritten text recognition algorithms and therefore this slope should be detected and corrected. This is done by estimating the lower baseline and determining its angle relative to horizontal (x-axis). After the slope detection, the word is rotated to the origin.

Here we discuss various approaches of skew normalization. Some of them are dealing with global skew and some of them are for local skew.

1) Approach based on WVD and projection: SKew correction using wigner ville distribution is a two step process

1) Rotate the text line from $-45$to 45(normal skew error range), and calculate horizontal histogram for each angle of rotation.
2) Compute Wigner-Ville [7] distribution of each histogram and the angle with largest distribution intensity, is taken as our skew angle α.

Wigner-Ville distribution [7] is related to time frequency representation of non-stationary signals. In this case, it is affine transformation of an image on intervals of 5^o. For a signal, $s(t)$ with analytic associate $x(t)$, the Wigner-Ville distribution $W_x(t, \omega)$ is defined as

$$W_x(t, \omega) = \int_{-\infty}^{\infty} x(t + \tau/2)x^*(t - \tau/2)e^{-i\omega\tau}d\tau$$

where, the analytic associate $x(t)$ of a signal $s(t)$ is defined as $x(t) = s(t) + \iota H[s(t)]$, where $H[s(t)]$ is the Hilbert transform [7] of the signal $s(t)$. Significance of peaks and troughs are deduced and compared by Wigner-Ville distribution.

2) Bounding Box Method: This method[5] is based on finding the extreme corners of a text image. If the four extreme points form a perfect rectangle,the desired angle could be easily determined. Four final extreme corners of the image are obtained via vertical and horizontal scaling. The advantage of the Bounding Box Algorithm is that if any two of four corner points detected correctly, it will estimate an accurate skewed angle. On the other hand, may be these two of four points do not find out correctly as a problem or fault. The algorithm is as follows:

1) Traverse the image in all the directions one by one to get the four extreme points.
2) Calculate the Euclidean distance between these points.
3) Calculate rotate angle from obtained square.

3) Linear regression: In this approach, we compute the count of foreground pixels in a vector Q using Equation 1.

$$Q = \{q_i = (x_i, y_i) | \text{Lowest black pixel} \in \text{x}_i \text{ column}\} \quad (1)$$

Q_i represents a vector of bottom most pixels of each line for the i^{th} column.
The skew is computed by fitting a straight line
$(L) = y - mx - c$
from the vector Q using least square sense (overall solution minimizes the sum of the square of errors made in the results of every single equation). The slope (m), of the line (L) gives the skew angle θ as: $\tan^{-1}(m)$. To normalize the line segment, for skew error, rotate it by -θ.

4) Approach based on hough transform: The Hough transform is a well known technique in computer vision that has been used to detect lines and curves in digital images. Basically it is a voting algorithm. Each point is handled independently so parallel implementations are possible. It computes the values for the parameters of all the curves of a particular type (e.g., straight lines) that can pass through each black pixel. A comprehensive method for detecting straight line segments in any digital image is accurately controlling both false positive and false negative detections [11]. Hough transform is computationally expensive algorithm and so it is advised preprocessed the document to reduce the number of pixels in order to reduce the processing.

III. Proposed Approaches

A. Proposed Binarization Approach

The intensity histogram computation based approaches [8] [9] [10] [11], generally use two peaks for finding the threshold value, but many images do not have such two peaks in the histogram. The proposed approach is basically global binarization approach and is very effective for those handwritten documents where the text written in input image is having more or less the same intensity value or written by a single pen. Basically it calculates a global threshold. However a unique feature of this algorithm is that it is very simple and performs better than Otsu [8] method and Niblack's[9] method. Assume, that the input image has black foreground pixels and white background pixels (reverse it for the other way around), and perform:

1) Scan each pixel value of image and find I_{max} and I_{min} intensity value of pixels. Assuming that our background is with maximum intensity and minimum value is for text part, and same pen is used by the writer.
2) Set I_{diff} = I_{max} - I_{min}
3) If $I_{diff} > I_{mean}$ (127 for 8 bit images)
   then
   set $I_o = 100$ else $I_o = 20$ Here, if difference between foreground and background intensity is large than average value than possible variation range of text content in also in large range and so offset set to be 100 and for small difference, variation range of text content also be small so offset set to be 20.
4) I_{th} (threshold) = $I_{min} + I_o$

B. Proposed Skew Normalization Approach

According to reviewed approaches in previous section and testing practically with several text word images, we observed that there are several problems. They all are employed for a particular interval of angles properly; and they are time consuming because of enormous volume of their calculations.

In our proposed approach, we assume that given input image of handwritten text is correctly oriented so we can say that

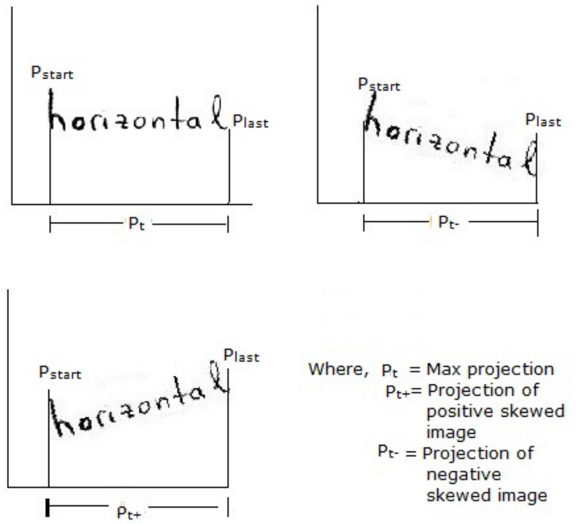

Fig. 7. Example image with positive skewed and negative skewed image version and its orthogonal projections($p_t, P_{t}+, and P_{t}-$) with respect to X-axis

there are only those skew which are introduced by the writer. After applying binarization approach as given in [8], [9], [10], [11] to given handwritten text document, the handwritten text lines are segmented from the given image using various approaches. and then we apply the skew correction approach on segmented line or word image of handwritten text document.

For the calculation of orthogonal projection of segmented image, as shown in Figure 7, we find start and last black pixel values of segmented image P_{start} and P_{last} respectively. and then projection P_l is calculated as $P_{start} - P_{last}$ and corrected by δ value.

Our proposed approach for skew normalization is illustrated in Algorithm 1

Algorithm 1 Text Line Image Skew Correction

Ensure: I-Skew corrected.
Require: I-Skewed image with intensity of background as 0.
$P_l = P_{last} - P_{start} - \delta$ {Calculate the orthographic projection of text line on x-axis.}
$I = I_{(R,+2^)}$
$P_{l_+} = I_{Orthographic}${Orthographic projection of rotated by $(+2^o)$ image}
$I = I_{(R,-2^)}$
$P_{l_-} = I_{Orthographic}${Orthographic projection of rotated (-2^o) image}
Set $\theta = 1${Initialize θ}
if $(P_{l_+} > P_l)$ **then**
    {Positive skew in image.}
    while $(P_{l_+} > P_l)$ **do**
        $I = I_{(R,\theta)}$ Image rotated by angle θ;
        $\theta + = 1$;
        $P_{l_+} = I_{Orthographic}$
    end while
else if $(P_{l_-} > P_l)$ **then**
    {Negative skew in image.}
    while $(P_{l_-} > P_l)$ **do**
        $I = I_{(R,-\theta)}$ Image rotated by angle $-\theta$;
        $\theta + = 1$;
        $P_{l_-} = I_{Orthographic}$
    end while
end if
Return(I)

IV. Results and Comparative Analysis

We have tested the proposed methods on 500 test images (both handwritten, printed, single pen) for our results. Some of the test images were downloaded from internet and some are written by various writers on different backgrounds. Our results are compared with standard algorithms by: Otsu [8], Niblack [9], Wu Manmatha [10], Srihari [11] for comparative analysis.

We are illustrating here some of the boundary cases for comparative analysis, as for general/average cases most of the algorithms gives better results. Figure 1 shows the result of various approaches for a given input image, which has large background noise in the form of mesh of lines with almost similar intensity to the text. For the same input image our proposed approach's result is shown in Figure 2.

Comparing the various results from Figures 1 and 2, we found that our result are qualitatively better than Otsu [8] and Niblack [9], and almost competitive to Wu Manmatha [10] and Srihari [11].

Figure 3 also shows the handwritten text document image and its respective binary image derived from our proposed binarization approach.

We have also tested the proposed skew normalization method on the same set of 500 handwritten text Images after finding binary image through proposed binarization approach. Comparing it with some of the existing approaches, we summarize the results in Table I.

TABLE I
SKEW NORMALIZATION COMPARATIVE ANALYSIS AFTER BINARIZATION USING PROPOSED APPROACH.

Compare methods	**Skew angle (in degrees)**	**Accuracy (%)**	**Complexity**
Bounding Box Method[5]	$0 - 25$	81.30	$m \times n$
Linear Regression[7]	$0 - 360$	82.10	$m \times n$
Hough Transformation[3]	$0 - 180$	95.25	$m \times n$
Proposed Method	$0 - 360$	97	Linear m

V. CONCLUSIONS

It has been observed experimentally that all types of images could be successfully binarized by our proposed global binarization approach and we get better results than others as discussed in Section IV. Our approach is also very fast as it uses only integer addition operation(s), compared to histogram computation by others.

We have also proposed a new approach of skew normalization for handwritten text document with linear time complexity. The proposed approach determines the exact skew and is computationally very efficient compared to existing techniques, as it does simple comparison operations for skew detection. It could handle all skews (360^o), and hence could be adapted to different language writing styles with equal accuracy. For example, we write left to right in most of the languages, but in Urdu language we write from right to left. All such cases could be handled by our approach, where rotation angle varies only from $[0 : 90]$.

REFERENCES

[1] A. Bagdanov and J.Kanai, *"Projection profile based Skew Estimation Algorithm for JBIG Compressed Images "*, ICDAR pp 401–405, 1997.

[2] Hull, J.J,Taylor, S.L. *"Document image skew detection: survey and annotated bibliography"*, Document Analysis System II, pp. 40–64, World Scientific, Singapore 1998.

[3] Srihari, S.N., Govindraju *"Analysis of textual image using the Hough transform"*, Machine Vision Applications Vol. 2 141–153,1989.

[4] A. H. W. Chin and A. jennings *"Skew detection in handwritten scripts."*, Proc. IEEE on Speach and Image Technologies for computation and Telecommunications, PP. 319–322,1997.

[5] M Sarfraj and Z Rasheed *"Skew Estimation and Correction of Text using Bounding Box"*, Fifth IEEE conference on Computer Graphics, Imaging and Visualization, pp. 259–264, 2008.

[6] Z.Shi , V. Govindraju *"Skew Detection for Complex Document Images Using Fuzzy Runlength"*, Proc. of the ICDAR 2003.

[7] Andria, G.; Savino, M.; Trotta, A.; , *"Application of Wigner-Ville distribution to measurements on transient signals,"*,strumentation and Measurement, IEEE Transactions on , vol.43, no.2, pp.187-193, Apr 1994.

[8] N. Otsu, *"A threshold selection method from Gray level histogram"*, IEEE Transaction on system, Man, Cybernetics, Vol. 19, pp 62–66, January 1978.

[9] W. Niblack, *"An Introduction to Digital Image Processing."*, Englewood Cliffs, New Jersey: Prentice-Hall, 1986.

[10] Wu, V. and Manmatha, *"Document Image Clean-Up and Binarization"*, Proceedings of SPIE conference on Document Recognition Vol.3 pp 18–23, January 1998.

[11] Liu, Y. and Srihari, S. N. *"Document Image Binarization Based on Texture Features"*, IEEE Trans. on PAMI Vol. 19(5), pp 540–544, May 1997.

[12] Sarfraz, M.,Mahmoud, S.A and Rasheed, Z. *"On Skew Estimation and correction of Text"*, Computer Graphics, Imaging and Visualization IEEE Computer Society USA pp 308–313, 2007.