Power-law Transformation for Enhanced Recognition of Born-Digital Word Images

Deepak Kumar and A G Ramakrishnan
Medical Intelligence and Language Engineering Laboratory
Department of Electrical Engineering
Indian Institute of Science
Bangalore, 560012, India
Email: deepak,ramkiag@ee.iisc.ernet.in

***Abstract*—In this paper, we discuss the issues related to word recognition in born-digital word images. We introduce a novel method of power-law transformation on the word image for binarization. We show the improvement in image binarization and the consequent increase in the recognition performance of OCR engine on the word image. The optimal value of gamma for a word image is automatically chosen by our algorithm with fixed stroke width threshold. We have exhaustively experimented our algorithm by varying the gamma and stroke width threshold value. By varying the gamma value, we found that our algorithm performed better than the results reported in the literature. On the ICDAR Robust Reading Systems Challenge-1: Word Recognition Task on born digital dataset, as compared to the recognition rate of 61.5% achieved by TH-OCR after suitable pre-processing by Yang et. al. and 63.4% by ABBYY Fine Reader (used as baseline by the competition organizers without any preprocessing), we achieved 82.9% using Omnipage OCR applied on the images after being processed by our algorithm.**

***Index Terms*—power-law transform; binarization; stroke width; word recognition; born-digital image;**

I. Introduction

Object detection and recognition in images is an active research area. We come across several object detection algorithms in the literature. In document imaging community, conventional research primarily focused on digitization of scanned documents. It involved binarization of document image followed by recognition engine. Now, we have several optical character recognition (OCR) engines for the Roman script [3], [11], [14], [4]. When a camera captured image is submitted to such OCR engines, the recognition performance is not necessarily very good. Lucas et.al split the process of word recognition in camera captured images into two parts, namely localization (or detection) and recognition [13]. In International Conference on Document Analysis and Recognition (ICDAR) 2003, a competition was organized for text localization on camera captured images and recognition from the word images extracted by placing a bounding box on the image. Only five entries were received for text localization and none for word recognition. Now, we have several publicly available datasets for text localization [9]. These datasets of camera captured images are known as IAPR TC11 Reading Systems-Datasets. Most of them assume that the bounding box information of a word is sufficient for any OCR to recognize; however we see that the best performing algorithm only has 52% word recognition rate on ICDAR 2003 word image dataset [1]. Recently held ICDAR 2011 Robust Reading challenge 2 reports that the best word recognition rate was 41.2% [2].

Fig. 1. Some word images from Born-digital dataset [6].

Karatzas et.al initiated another robust reading challenge in ICDAR 2011 for born-digital images [6]. Born-digital images are formed by a software by overlaying text on an image. For competition, these images were collected from web pages and email. Most words present in this dataset are oriented horizontally. The reason behind horizontal placement of text may be the simplicity involved in creating the born-digital image using standard softwares. Low resolution of text and anti-aliasing are the main issues to be tackled in born-digital images, whereas illumination changes is a difficult problem in the case of camera captured images. These issues indicate the complexity involved in processing born-digital and camera captured scenic images. Even though we think that the proposed approach may also be applicable for camera captured word images, in this paper we deal only with born-digital word images. ICDAR 2011 received seven entries for text localization task, three for text segmentation and only one for word image recognition. Our OTCYMIST algorithm was the winner of text segmentation task in the competition [7]. We found that the performance of the algorithm submitted for word recognition did not match that of ABBYY Fine Reader. This inspired us to explore as to why the word recognition rate was low. Since the resolution of the given images themselves is very low, the words extracted from born-digital images are still lower in resolution. Some word images from the training dataset are shown in Figure 1.

Recognition of a word image involves two stages: binarization of word image and recognition of the characters. The latter step can be performed using a training dataset or an OCR engine. This latter issue has already been completely researched by the document imaging community. Thus, recognition of properly segmented Roman characters is no longer an issue. The real research issues are improved preprocessing and segmentation of degraded, low resolution or historical documents. Thus, we use Omnipage Professional 16 trial version OCR for recognition of characters in the binarized image [11]. Our task is to develop an algorithm which binarizes a given word image. The major problem in born-digital word images is their resolution. During the binarization process, several characters get merged due to the close placement of characters and this reduces the word recognition rate. The minimum character gap in born-digital dataset is 1 pixel width, which was found during the cross validation process. We present a novel method to prevent the merging of characters, thus increasing the word recognition rate for the born-digital word image dataset.

II. Existing algorithm

This algorithm also focuses on text binarization rather than on the OCR [6]. All word images provided are normalized to the same height using a bi-cubic interpolation method; the height was set to 100. During the text binarization stage, the text polarity is established based on connected component analysis. Then adaptive local binarization is used to create coarse text components followed by morphological opening to separate consecutive connected characters. Noise and bar-style components are filtered out in a post- processing step. For recognition, they use the OCR engine TH-OCR 2005 [8]. OCR text segmentation is guided by the binary image while for recognition, gray scale features are used. The recognition result is achieved without using any language model.

III. Thresholding

An advantage of born-digital images is minimal illumination variation. That means a global thresholding technique can be utilized for binarization. We use Otsu's method for binarization of the image, after it is enhanced by power-law transformation. Since it is one of the global thresholding techniques [10]. In Otsu's method, the threshold calculation is posed as an optimization problem. This thresholding scheme is used in binarization of document images, and works well for images

Fig. 2. Binarization of two sample word images using Otsu's method. In the second example, the binarized image is inverted, since we require polarity of text to be brighter than the background.

of good quality. Any choice of threshold value k splits the histogram into two parts. The optimal threshold value $k*$ is arrived at by maximizing the following objective function:

$$\sigma^2(k*) = \max_k \frac{[\mu_T \omega(k) \ - \ \mu(k)]^2}{\omega(k)[1 \ - \ \omega(k)]} \tag{1}$$

where,

$$\omega(k) = \sum_{i=1}^{k} p_i \tag{2}$$

$$\mu(k) = \sum_{i=1}^{k} i p_i \tag{3}$$

$$\mu_T = \sum_{i=1}^{L} i p_i \tag{4}$$

Here, 'L' is the total number of gray levels and 'p_i' is the normalized probability distribution obtained from the histogram of the image.

Each word image is converted to gray scale and Otsu's threshold is applied. After binarization, components touching the boundary are removed. In this binary image, the polarity of the characters is the opposite of the dominant polarity of the boundary pixels. If the text is darker, the image is inverted before proceeding to the successive stage. This process makes all the binarized word images uniform for the application of power-law transform, where all white pixels form components of a character. Figure 2 shows an example of inversion for an image containing text that is darker than its background.

IV. Power-law transform

In poor contrast images, the adjacent characters merge during binarization. We have to reduce the spread of the characters before applying a threshold to the word image. Hence, we introduce power-law transformation which increases the

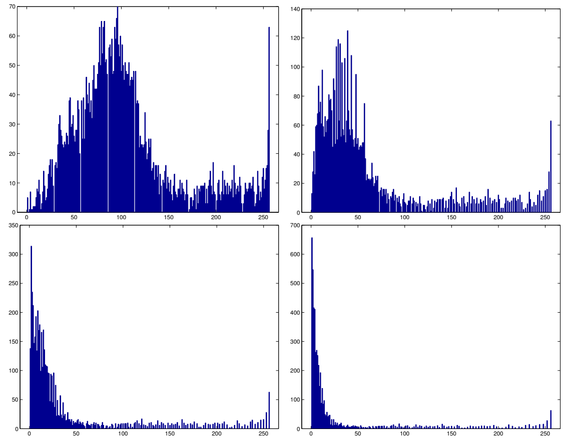

Fig. 3. Histogram plots obtained for the first sample image shown in Fig. 2 after power-law transformation with $\gamma = 1, 2, 3$ and 4, respectively.

contrast of the characters and helps in better segmentation. The basic form of power-law transformation is [12],

$$s = cr^{\gamma} \tag{5}$$

where r and s are the input and output intensities, respectively; c and γ are positive constants. A variety of devices used for image capture, printing, and display respond according to a power-law. By convention, the exponent in the power-law equation is referred to as gamma. Hence, the process used to correct these power-law response phenomena is called gamma correction. Gamma correction is important, if displaying an image accurately on a computer screen is of concern. In our experimentation, γ is varied in the range of 1 to 5. If c is not equal to '1', then the dynamic range of the pixel values will be significantly affected by scaling. Thus, to avoid another stage of rescaling after power-law transformation, we fix the value of $c = 1$.

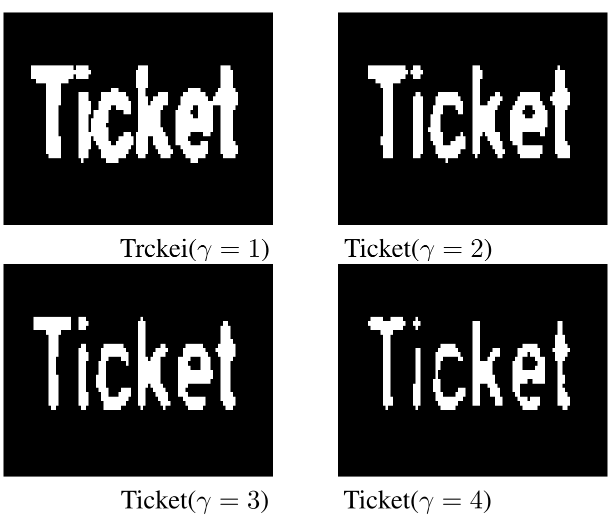

Fig. 4. Binarization outputs and their respective OCR results for different values of γ. To begin with, increasing the value of γ yields proper output and further increase deteriorates the character stroke width.

Fig. 5. Contrast enhancement after power-law transformation with $\gamma = 1, 2, 3$ and 4, respectively.

With $\gamma = 1$, if the power-law transformed image is passed through binarization, there will be no change in the result compared to simple binarization. When $\gamma > 1$, there will be a change in the histogram plot, since there is an increase of samples in the bins towards the gray value of zero. Figure 3 shows the change of histogram with different γ values.

Figure 4 shows the binarization result and the OCR output on a sample input image, for different values of γ. We can clearly observe that the result is erroneous when the characters are merged. With increase in γ value, merged characters split, which results in better OCR output. As discussed, due to low contrast, characters get merged into a single connected component (CC). By applying power-law transform, we are eroding the CC's in a non-linear way. Pixels at the boundary of the CC's are eroded at the earliest. The cause for merging is anti-aliasing, and this effect on the text components in the image is removed. Hence, we call our method as power-law transformation (PLT) method.

Figure 5 shows the gray scale image after power-law transformation with different gamma values. We can clearly identify the gradual reduction of the gradient shadow when the gamma value is increased. Thus the application of power-law transform increases the contrast in the image resulting in improved image binarization.

Each word image responds differently to power-law transform. As we increase γ to a large value, individual text components may split into multiple components. This leads to a poor performance of OCR and reduces the word recognition rate for the word images dataset. Figure 6 shows the plot of word recognition rate as a function of γ, for the entire born-digital image database, with γ shown in log scale and without restriction on stroke width. For $\gamma = 1$ and $c = 1$, we perform OCR on original stroke width of word images. As γ value is increased, there is an improvement in word recognition rate. We observe that the word recognition rate decreases for very high values of γ.

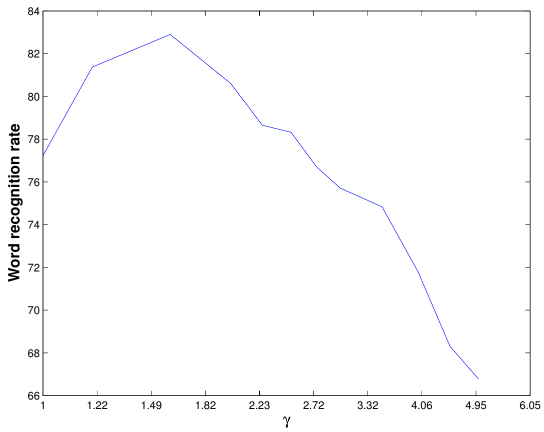

Fig. 6. A plot of word recognition rate against $\log \gamma$ (with no restriction imposed on the stroke width).

V. Proposed Algorithm

Figure 7 shows the block diagram of proposed power-law transform (PLT) algorithm. Due to change in text stroke width in each word image, we cannot fix a specific height for resizing an image. Hence, we resize a word image to three times the existing size, unlike [6]. From the resized image, we obtain gray scale image for the next level of processing. The first block processes for text polarity and inverts the image if necessary. Initially, γ value is set to '1' and the image is binarized using Otsu's threshold.

Estimating the stroke width of a character using median of pixel-by-pixel estimates of stroke widths is computationally intensive [5]. Hence, in our work, we perform horizontal run length measures of white (foreground) pixels for each row in the binarized image. The approximate stroke width is obtained as the average of the non-zero white (foreground) run length values measured across all the rows of the binarized image. In our algorithm, if the mean value of stroke width is less than '8', then we pass the binarized image to OCR block; else γ value is increased in steps of '0.2'. Further, if any original word image which has stroke width less than '8' or if γ has reached a maximum threshold value, then too the image is passed to OCR block.

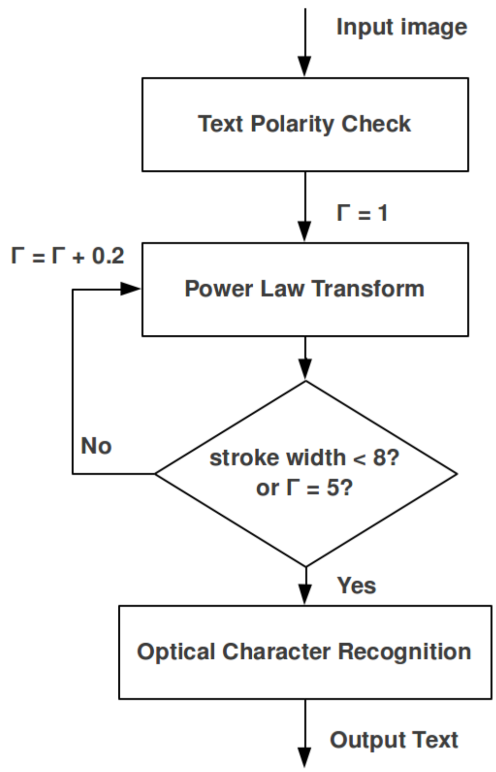

Fig. 7. Block diagram of proposed (PLT) algorithm showing text polarity, power-law transformation and OCR engine stages with feedback based on stroke width of binarized image.

VI. Recognition

Binarized word image is fed to the recognition engine. There are several OCR engines available such as Tesseract [14], OmniPage [11], Adobe Reader [4] and Abby Fine Reader [3]. These engines are used in digitization of old hard bound documents. The quality of OCR performance is good for all the above engines and hence any of them can be used to recognize words from binarized word image. In our experiment, we use Omnipage Professional 16 OCR trial version. The OCR result is used for the performance evaluation of our algorithm.

VII. Performance Evaluation

A standard edit distance metric is used for the evaluation of the recognized word [6]. Equal weights are given for additions, substitutions and deletions. Normalized edit distance is calculated between the ground truth transcription and algorithm's transcription. The number of words recognized is also provided for quantitative analysis of the algorithm performance. Figure 8 shows the plot of word recognition rate for the entire born-digital image database, as a function of the choice of stroke width threshold.

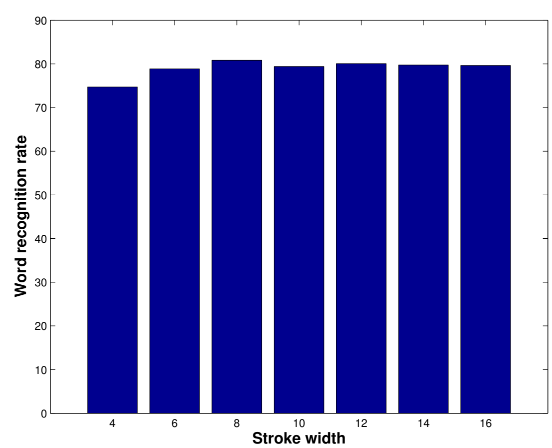

Fig. 8. Plot of word recognition rate for the entire born-digital image database against different stroke width thresholds.

TABLE I
PERFORMANCE EVALUATION OF PROPOSED (PLT) ALGORITHM ON BORN-DIGITAL WORD IMAGE DATASET [6].

Algorithm	Edit distance	Word recognition rate
PLT algorithm	108.7	82.90%
TH-OCR	189.9	61.54%
Baseline method	232.8	63.40%

VIII. EXPERIMENTAL RESULTS

The training dataset contains 3583 word images extracted from born-digital images. Testing dataset contains 918 word images. During cross validation, we observed from the training set that some characters in the word images are only one pixel wide. We applied clustering on the resized image and this resulted in merging or loss of characters due to one pixel width. Hence, we avoided clustering and chose thresholding technique for binarization of an image. Table I compares the edit distance measure and word recognition rate of the proposed (PLT) algorithm with those of other methods. ABBYY Fine Reader was used as baseline method for word recognition by choosing the option of low resolution input image for the OCR. The algorithm that participated in the competition has a lower recognition rate than the baseline method. The structural element used in morphological operation is not known. Fixing the size of structural element is not advisable since the text stroke width varies in these word images. Figure 9 shows a set of word images where proposed (PLT) algorithm fails to correctly recognize words. Low contrast of some of the characters and the handwriting-like font are the reasons for failure in these images.

IX. CONCLUSION

We have proposed an algorithm for improving the recognition of born-digital word image dataset. We observe from experimentation (see Table I) that for a born-digital image, even the correct detection of the bounding box of text, is not sufficient for an algorithm to accurately recognize the text. Using power-law transform, we have proposed a novel way to prevent the merging of characters during binarization. We can utilize our algorithm to improve the precision and recall rates of text localization task for the born-digital image dataset. Our results are of a general nature, and hence can be effectively used for improved segmentation and consequent enhanced recognition of other types of low resolution images such as scenic word images.

Fig. 9. Images, where proposed (PLT) algorithm fails to recognize the words.

REFERENCES

[1] A. Mishra, K. Alahari and C. V. Jawahar, "An MRF Model for Binarization of Natural Scene Text", Proc. 11th *International Conference of Document Analysis and Recognition*, pp. 11–16, September 2011, 2011.

[2] A. Shahab, F. Shafait and A. Dengel, "ICDAR 2011 Robust Reading Competition - Challenge 2: Reading Text in Scene Images", Proc. 11th *International Conference of Document Analysis and Recognition*, pp.1491–1496, September 2011, 2011.

[3] Abbyy Fine reader. http://www.abbyy.com/

[4] Adobe Reader. http://www.adobe.com/products/acrobatpro/scanning-ocr-to-pdf.html

[5] B. Epshtein, E. Ofek and Y. Wexler, "Detecting text in natural scenes with stroke width transform", Proc. 23rd *IEEE conference on Computer Vision and Pattern Recognition*, pp. 2963–2970, 2010.

[6] D. Karatzas, S. Robles Mestre, J. Mas, F. Nourbakhsh and P. Pratim Roy, "ICDAR 2011 Robust Reading Competition - Challenge 1: Reading Text in Born-Digital Images (Web and Email)", Proc. 11th *International Conference of Document Analysis and Recognition*, pp.1485–1490, September 2011, 2011. http://www.cv.uab.es/icdar2011competition/

[7] D. Kumar and A. G. Ramkrishnan, "OTCYMIST: Otsu-Canny Minimal Spanning Tree for Born-Digital Images", Proc. 10th *International workshop on Document Analysis and Systems*, 2012.

[8] H. Liu and X. Ding, "Handwritten Character Recognition Using Gradient Feature and Quadratic Classifier with Multiple Discrimination Schemes", Proc. 8th *Int. Conf. Document Analysis And Recognition*, pp. 19–25, 2005.

[9] IAPR TC11 Reading Systems-Datasets List, http://www.iapr-tc11/mediawiki/index.php/Datasets

[10] N. Otsu, "A Thresholding Selection Method from Gray-level Histogram", *IEEE Trans. on Systems, Man and Cybernetics*, vol. 9, pp. 62–66, March 1979.

[11] Nuance Omnipage reader. http://www.nuance.com/

[12] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Second Edition, Prentice Hall, 2002.

[13] S. M. Lucas et.al, "ICDAR 2003 Robust Reading Competitions: Entries, Results, and Future Directions", *International Journal on Document Analysis and Recognition*, vol. 7, no. 2, pp. 105–122, June 2005.

[14] Tesseract OCR engine. http://code.google.com/p/tesseract-ocr/