NESP: Nonlinear enhancement and selection of plane for optimal segmentation and recognition of scene word images

Deepak Kumar, M N Anil Prasad and A G Ramakrishnan
Medical Intelligence and Language Engineering Laboratory
Department of Electrical Engineering
Indian Institute of Science, Bangalore, INDIA-560012

ABSTRACT

In this paper, we report a breakthrough result on the difficult task of segmentation and recognition of coloured text from the word image dataset of ICDAR robust reading competition challenge 2: reading text in scene images. We split the word image into individual colour, gray and lightness planes and enhance the contrast of each of these planes independently by a power-law transform. The discrimination factor of each plane is computed as the maximum between-class variance used in Otsu thresholding. The plane that has maximum discrimination factor is selected for segmentation. The trial version of Omnipage OCR is then used on the binarized words for recognition. Our recognition results on ICDAR 2011 and ICDAR 2003 word datasets are compared with those reported in the literature. As baseline, the images binarized by simple global and local thresholding techniques were also recognized. The word recognition rate obtained by our non-linear enhancement and selection of plance method is 72.8% and 66.2% for ICDAR 2011 and 2003 word datasets, respectively. We have created ground-truth for each image at the pixel level to benchmark these datasets using a toolkit developed by us. The recognition rate of benchmarked images is 86.7% and 83.9% for ICDAR 2011 and 2003 datasets, respectively.

Keywords: nonlinear enhancement, power-law transform, text polarity inversion, binarization, evaluation, threshold, recognition, normality test

1. INTRODUCTION

A fundamental issue in most image processing problems is image segmentation for object detection and recognition. Traditionally research in document analysis systems was focussed on binarization and recognition of scanned documents. Application of document analysis systems is not limited to digitization of books, but also extends to recognition of vehicle number plates, street names and contents of hoardings. Documents captured with a camera have different kinds of degradations such as perspective deformation, page curl, non-uniform illumination and blur. The performance of optical character recognition (OCR) engines on camera captured images is poor due to these degradations. Hence, text localization, segmentation and word recognition algorithms for camera captured images have become major research areas in document analysis systems. Lucas et.al organized separate competitions for text localization in camera captured images and recognition of words from word images in International Conference on Document Analysis and Recognition (ICDAR) 2003[9] and 2005.[10] In the literature, we come across different methods proposed for word recognition.[12,19,22] IAPR hosts several publicly available datasets as TC-11 standards.[30] Some samples from ICDAR 2011 dataset are shown in Figure 1.

Recently held ICDAR 2011 Robust Reading challenge 2 reported that the top word recognition rate was 41.2%.[21] Shahib et.al[21] explain that the cropping of word images was done based on ground-truth word bounding boxes, to evaluate the performance of word recognition, independently from text localization accuracy . Text localization and word recognition tasks are assumed to be independent of each other. The harmonic mean of text localization task, in ICDAR 2011 competition, was 71.3%. The harmonic mean calculated for the reported text localization and word recognition results, as an end-to-end robust reading system approach, is 52.2%. Thus, the performance of the end-to-end robust reading system is poor from the reported results in the competition.

Further author information: (Send correspondence to Deepak Kumar)
Deepak Kumar: E-mail: deepak@ee.iisc.ernet.in, Telephone: +91 80 2293 2935
M N Anil Prasad: E-mail: anilprasadmn@ee.iisc.ernet.in, Telephone: +91 80 2293 2935
A G Ramakrishnan: E-mail: ramkiag@ee.iisc.ernet.in, Telephone: +91 80 2293 2556

Figure 1. Samples from ICDAR 2011 word images dataset.[21]

As the word recognition rate is low, this reduces the performance of the whole system. There were three entries for word recognition in ICDAR 2011[21] and none in ICDAR 2003.[9]

Recognition from a word image involves two stages: binarization of the word image and recognition of the characters. We propose an algorithm which binarizes a given word image effectively. The word images in ICDAR 2011 image dataset are cropped using ground-truth bounding boxes. This dataset is affected by uneven illumination, low resolution and/or varying character stroke width. Due to these degradations, the word recognition rate in the competition was less than that of text localization result. We handle these degradations, with separate treatments for each degradation such that they have less effect. The novelty of proposed solutions is unique compared to existing methods. The proposed solutions reduce the effect of gradually varying illumination by splitting the image into its constituent parts. In the case of low resolution, the text pixels cannot be captured properly in images, since the number of text pixels is not sufficient for any kind of processing. Hence, we also propose a method to scale the height of an image in size normalization section.

In most of the images in the dataset, the text components touch the borders of the image. This results in ambiguity, when the segmented image is passed to the OCR engine.[33] This problem has been handled by text polarity inversion.

2. RELATED WORK

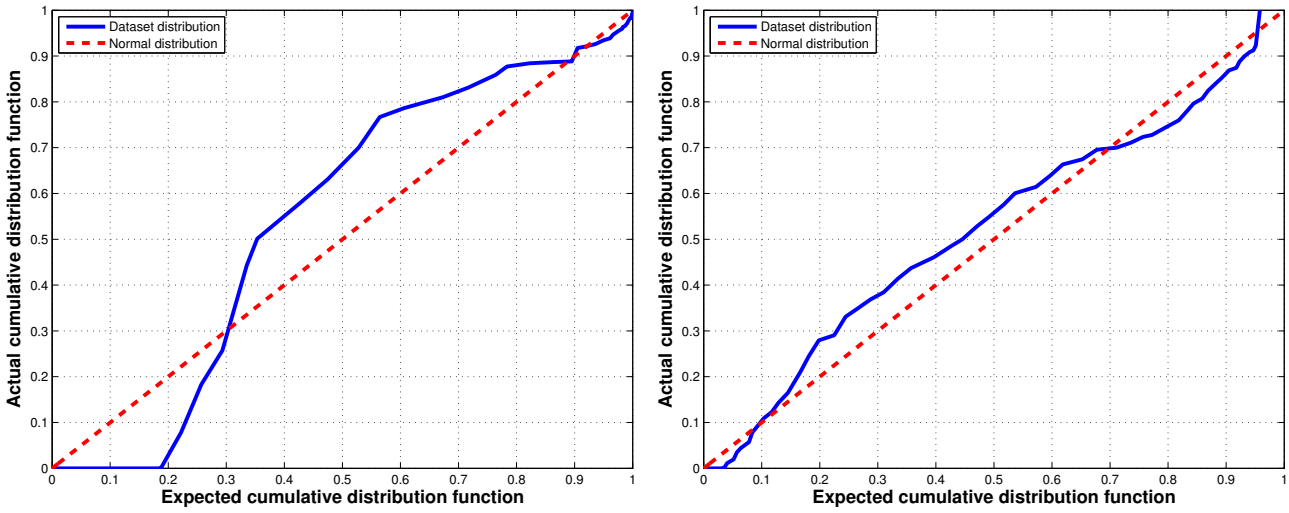
Here, we discuss some of the methods in the literature to know the difficulty in the recognition of word images. In ICDAR 2003 and ICDAR 2011 datasets, we observe variation in the stroke width of the text components. Methods have been proposed for segmenting word images, such as conditional random fields (CRFs) to form super pixels by KAIST AIPR,[32] Maximally Stable Extremal Regions (MSER) by Neumann,[8,18] midline analysis and propgation of segmentation (MAPS) by Kumar et al[27] and Markov random fields (MRFs) by Mishra et.al.[22] During segmentation, Mishra et.al have used Canny edges to seed foreground and background pixels.[3] Other explored methods are clustering and combining different segmentation techniques.[13,15–17,23]

After binarization of word images, recognition is performed using either a standard OCR engine or a classifier built for the purpose. KAIST AIPR system classifies super pixels and passes them to INZI soft OCR engine.[31] Similarly, TH-OCR system uses an OCR engine,[11] Mishra et.al use ABBYY OCR reader,[29] Zeng et.al use OmniPage OCR reader[33] and Neumann et.al[18] classify detected characters in the image using multi-class support vector machines (SVM) applied on character contour feature. Due to variation in the contour feature caused by noise or scaling, they were ranked low in ICDAR 2011: Robust Reading Competition, indicating that better features are required in the classification stage.

Words can be recognized in one shot using a training dataset, which avoids binarization. Wang et.al and Mishra et.al use this approach for word recognition.[19,24] A classifier is trained on selected features, using a training the data available in the dataset. Apart from the training, they provide lexicon for their algorithms (available as a part of the Street View Text (SVT) dataset),[19] as a form of top-down approach. Since ICDAR 2003 and 2011 datasets do not provide any lexicon for text recognition, our method does not use any lexicon as a part of post-processing. The results generated by OmniPage OCR is recorded in the experimental results section.

3. NESP ALGORITHM

Our algorithm, known as nonlinear enhancement and selection of plane (NESP), involves scaling, enhancement, plane selection and selective polarity inversion.

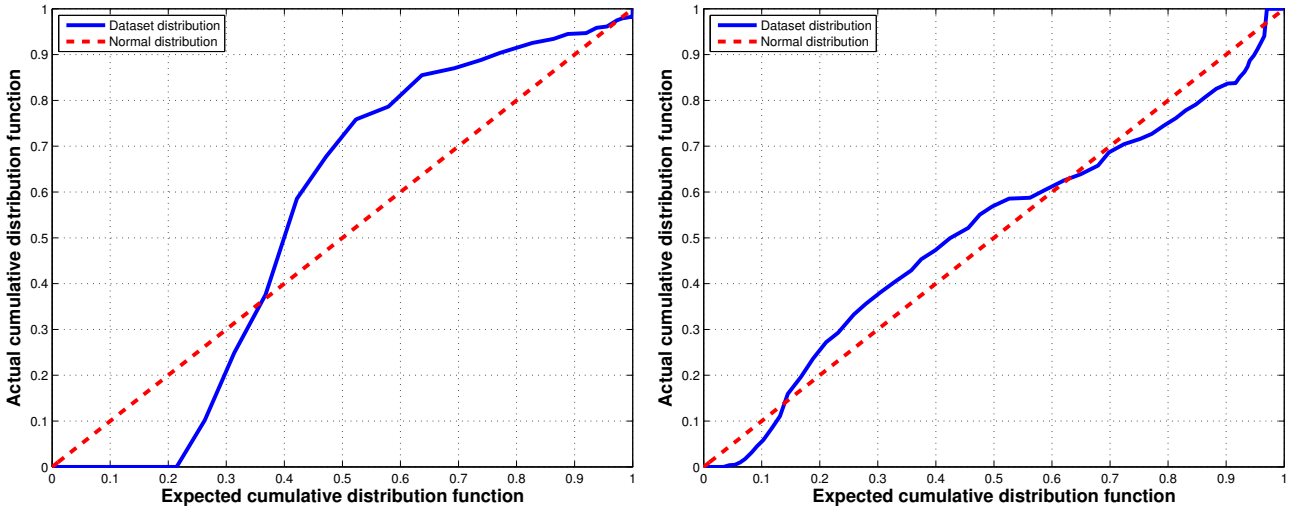


(i) (ii)

Figure 2. Q-Q plots on image heights for ICDAR 2011 dataset. The actual cumulative distribution function is plotted against the expected cumulative distribution. (i) Before height normalization of images. (ii) After height normalization of images.

3.1 Scaling for height normalization

Apriori, we do not have information on the stroke width of characters in the dataset but have access only to the height and width of the word images. A normality test was conducted on image heights in the dataset.[7] Figure 2 shows the plot of actual cumulative distribution against expected cumulative distribution. We observe that the image heights in the dataset are not close to the diagonal plotted in Figure 2(i). So, we perform image scaling to normalize the image height in the dataset. Images are scaled by bi-cubic interpolation preserving the aspect ratio. In order to minimize the variance in the stroke width, we modify the height to lie within a range. The height range is calculated from the normality test. Accordingly the rules for rescaling are:



(i) (ii)

Figure 3. Q-Q plots on image heights for ICDAR 2003 dataset. The actual cumulative distribution function is plotted against the expected cumulative distribution. (i) Before height normalization of images. (ii) After height normalization of images.

- Rule 1: If the height of an image is less than 60 pixels, then it is rescaled by a factor of '3'.
- Rule 2: If the height of an image lies between 60 and 180 pixels, then it is not rescaled.
- Rule 3: If the height of an image exceeds 180 pixels, then it is scaled to a height of '180' pixels.

After height scaling based on these rules. we again show the plot of actual cumulative distribution against expected cumulative distribution in Figure 2(ii). We observe the plot to be approximately linear, which results in a close to normal distribution dataset.

For ICDAR 2003 dataset also, we show the plots of actual cumulative distribution against expected cumulative distribution in Figures 3(i) and 3(ii), before and after height scaling, respectively.

The edges in a low resolution image become smooth during interpolation and we cannot perform any edge operations, since those kind of images will produce spurious edges. Zeng et.al[23] report that they have removed such images from ICDAR 2003 dataset, since they degrade the overall performance of the algorithm on the entire dataset. However, in our method, edge operations are not performed. Hence, the above issue is avoided.

3.2 Nonlinear enhancement of color planes

We split the word images into Red, Green, Blue, Gray and Lightness (CIE Lab) components,[6] which we refer to as planes. When a color image is converted to a gray image, the foreground and background gray levels may not be distinct due to the effect of illumination. However, they may be separable in one of the other planes.

Kumar and Ramakrishnan use power-law transformation on born-digital word images to improve recognition.[25] Extending the idea to camera captured images, in our work, each colour plane is enhanced using the non-linear transformation,

$$f_{out}(i,j) \;=\; C \;*\; f_{in}^{\gamma}(i,j) \tag{1}$$

where $f_{in}(i,j)$ and $f_{out}(i,j)$ are the pixel intensities of the input and output images, respectively; C and γ are positive constants. In our experimentation, C is fixed as 1.0. We vary the value of γ from 1.0 to 2.0 in steps of 0.2. The recognition results for different values of γ are reported in the section on experimental results. Figure 4 shows the values of the discrimination factor obtained before and after power-law transformation for each plane. In the bottom panel, we show the images obtained by binarizing the plane with the highest discrimination factor, before and after the proposed non-linear transformation. The enhancement in binarization achieved by the power law transformation is obvious.

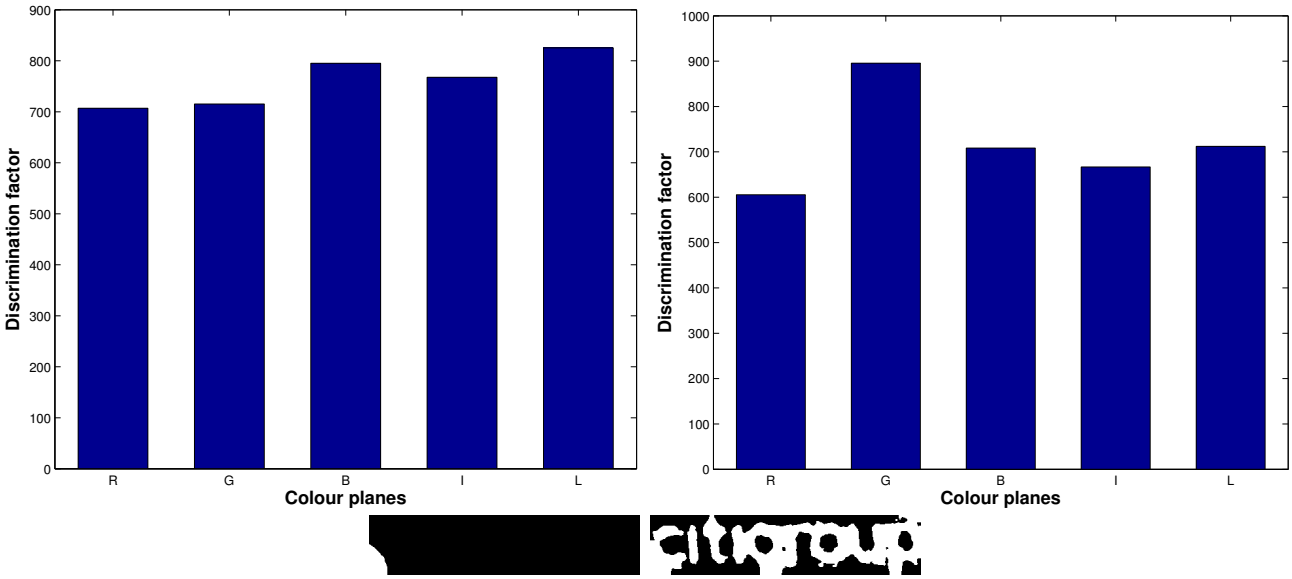


Figure 4. Top row: Discrimination factor computed by our algorithm for different colour planes without and with power-law transformation for 'citigroup' word image shown in Fig 1. Bottom row: Results of binarizing the selected plane.

3.3 Selection of plane for binarization

We can segment an image using either a clustering technique[13–17] or a thresholding technique.[1,2,4] We can also use edge information proposed by Canny.[3] During experimentation, we realized the need for a unique, objective metric which can be utilized for optimal segmentation. This metric should be in such a way, that it can be applied across all the planes derived from an image. When we tried to apply Niblack for binarization, the concept of uniqueness was broken due to the local window approach for segmentation. We can use Canny edges to obtain a measure. But, we found the degradations in the image provide undue weighting towards degraded edge pixels in the plane considered. This resulted in the selection of the degraded plane as the "best" plane. To have generality and uniqueness, we applied Fisher discrimination function[14] as proposed by Otsu.[1] Hence, we choose discrimination factor as the fitness measure for a plane.

Further, it is not possible to infer apriori the plane in which, the text is best separable from the background. Therefore each plane is considered to contain two classes of interest and the problem of selection of the best plane is posed as maximization of two class Fisher discriminant function across all the planes. Based on the fitness of the possible split, the plane of segmentation is chosen. The between-class variance used in 'Otsu' global thresholding technique[1] is found to produce good segmentation results on word images amongst known thresholding techniques. We use this measure as the discrimination factor, calculated for each plane, as:

$$d_c^2(k) = \max_k \frac{[\mu_{ct}\omega_c(k) - \mu_c(k)]^2}{\omega_c(k)[1-\omega_c(k)]}, \quad c \in [R,\ G,\ B,\ I,\ L] \tag{2}$$

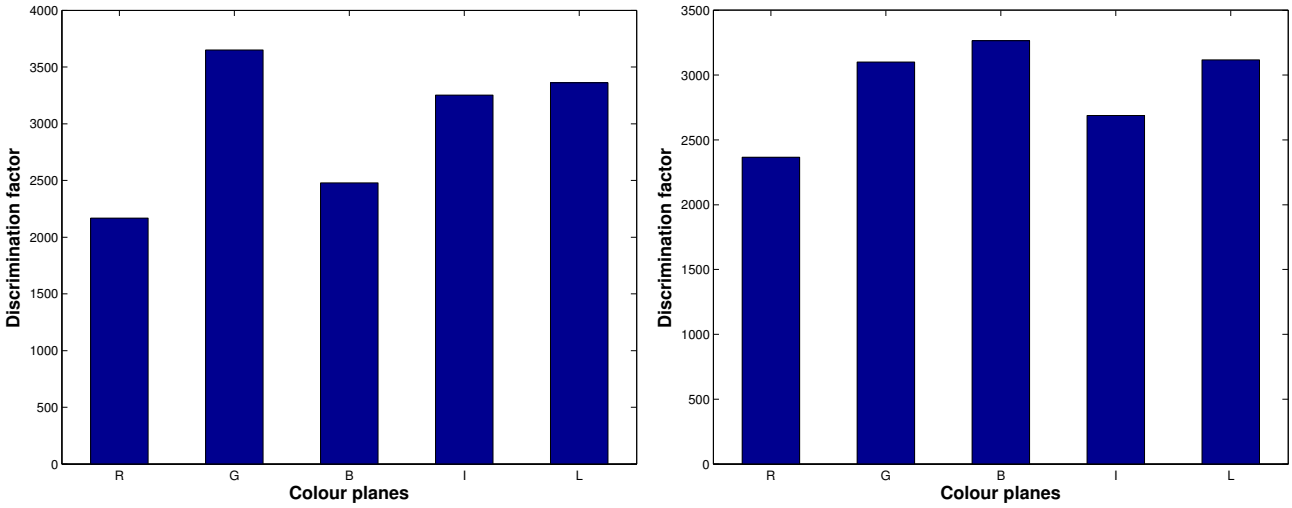


Figure 5. Discrimination factors computed for different colour planes of the 'create' and 'LION' word images shown in Fig 1.

Figure 6. Word images shown in Fig 5, after segmentation by our algorithm using the plane with maximum discrimination factor, computed post non-linear enhancement.

where,

$$\omega_c(k) = \sum_{i=1}^{k} p_{ci} \tag{3}$$

$$\mu_c(k) = \sum_{i=1}^{k} ip_{ci} \tag{4}$$

$$\mu_{ct} = \sum_{i=1}^{N} ip_{ci} \tag{5}$$

Here, 'N' is the total number of gray levels and 'p_{ci}' is the normalized probability distribution from the histogram of the c^{th} colour plane. $\omega_c(k)$ and $\mu_c(k)$ are the zero-th and the first order cumulative moments of the histogram up to k^{th} level, respectively and μ_{ct} is the mean of the complete histogram.

The plane with maximum discrimination factor is binarized using the corresponding threshold. Figures 5 and 6 show the discrimination factor for each plane and segmentation of the word image for the selected plane, respectively, for two sample images from ICDAR 2011 dataset.

3.4 Selective polarity inversion

Figure 7 shows some sample binarized images. Here, black and white pixels denote background (0) and foreground (1) respectively. Even in the images shown in Figure 6, we can observe several text components touching the boundary. Such binarized images result in wrong recognition when passed through an OCR engine, since both foreground and background pixels touch the image boundary. To eliminate the foreground-background ambiguity in dealing with such images, we first detect the polarity of the background and foreground pixels and optionally perform text polarity inversion and then background padding.

Text polarity is detected by examining the following three conditions:

- Is the ratio of number of white pixels along the boundary of segmented word image to the length of boundary greater than 0.5?
- Is the ratio of number of white pixels on the vertical sides of the segmented word image to the total length of the side walls greater than 0.5?
- Is the ratio of maximum widths of 'white' to 'black' connected components in the segmented word image greater than 1?

If two out of these three conditions are true, then polarity needs to be inverted, after which the text pixels will be white.

During thresholding stage, uneven illumination causes salt and pepper noise in the binarized image. Hence, we perform median filtering with a structuring element of size 5x5. The images rescaled using Rule 1 are excluded from median filtering, since they are of low resolution and filtering may degrade the binarized image. Post-processed binarized word image is then fed to the recognition engine.

To prevent the text connected components from touching the boundary of the image, we pad background pixels (zeros), vertically by half the number of rows and horizontally by half the number of columns. An example is shown in Figure 8. In Figure 8(i), the text connected components touch the word boundary. After padding background padding, we clearly observe the distinction between foreground and background pixels.

Figure 7. The background is broken, where the text connected components touch the boundary of the word image. This creates ambiguity in determining the text polarity.

(i) (ii)

Figure 8. Postprocessing the binarized image by padding background pixels to eliminate foreground-background ambiguity. (i) The text connected components touch the boundary. (ii) Foreground and background are clearly separated, after background padding.

4. EXPERIMENTAL RESULTS

ICDAR 2003 training dataset contains 1156 word images extracted from scene images. Testing dataset consists of 1110 word images. ICDAR 2011 dataset has 849 and 716 word images for training and testing, respectively. We have evaluated our algorithm on both the test datasets. We observe that there exists some correlation between the images used in both datasets. Apart from correlation, in some of the cropped images in ICDAR 2003 dataset, the boundaries are improper. The cropped images in ICDAR 2011 dataset are tight or closely bounded, which resulted in characters touching the boundary. We have calculated the recognition rate for the individual planes and also for the plane which had maximum discrimination among different planes. We have also varied the γ value, from 1.0 to 2.0, to observe any variation in the result.

Tables 1 and 2 show word recognition rates for ICDAR 2003 and ICDAR 2011 datasets with different variations proposed in NESP algorithm. From the Tables, we observe that the trend differs for different planes. For some, the performance increases with increasing γ and for others, it decreases. Even the peak recognition rate occurs at different gamma values for ICDAR 2003 and 2011 datasets, indicating the differences between the datasets.

Table 3 compares word and character recognition rates of our algorithm with those of methods in the literature on ICDAR 2003 dataset. Table 4 compares the total edit distance measure and word recognition rate of our algorithm with others on ICDAR 2011 dataset. Both Tables show that our algorithm performs favourably when compared to other methods. Edit distance measure was introduced in ICDAR 2011: Robust Reading Competition.[21] Hence, the edit distance measure is used in ICDAR 2011 datasets replacing the character recognition rate shwon in Table 3.

Table 1. Performance evaluation of NESP algorithm on the entire test images of ICDAR 2003 dataset for different planes, for different values of gamma .

Plane / γ	Red	Green	Blue	Intensity	Lightness	NESP
1.0	58.65	61.62	58.47	63.87	62.25	64.41
1.2	57.57	62.70	58.29	62.88	63.78	**66.22**
1.4	58.38	64.32	57.84	63.69	63.42	66.04
1.6	57.03	63.87	55.59	63.24	63.51	64.41
1.8	56.40	63.15	55.00	63.60	62.70	65.23
2.0	56.04	63.06	55.59	63.15	62.43	63.51

Table 2. Performance evaluation of NESP algorithm on the entire test images of ICDAR 2011 dataset for different planes, for different values of gamma.

Plane / γ	Red	Green	Blue	Intensity	Lightness	NESP
1.0	62.01	69.97	63.55	71.09	69.97	72.49
1.2	62.01	70.81	62.29	69.97	70.53	70.25
1.4	63.27	71.23	62.43	69.83	70.11	**72.77**
1.6	62.43	70.39	62.01	68.57	69.83	71.23
1.8	62.57	69.55	62.01	70.95	70.11	71.23
2.0	59.78	68.99	60.75	69.97	70.39	71.50

Table 3. Comparison of performances of NESP and other algorithms on ICDAR 2003 dataset.

Algorithm	Word recognition rate (%)	Character recognition rate (%)
Benchmark[26]	**83.9**	**92.1**
NESP algorithm	**66.2**	**80.2**
MAPS algorithm[27]	64.5	79.2
Mishra et.al[24]	61.1	—
Wang et.al[19]	53.8	—
Zeng et.al[23]	—	75.3
Otsu[1]	38.4	56.8
Kittler[2]	37.8	55.2
Sauvola[5]	22.6	39.1
Niblack[4]	18.7	38.0

4.1 Benchmarking

We also report the word recognition rate for an ideal image segmentation (ground-truth of word images) for both ICDAR 2003 and 2011 datasets.[26] We have used MAST-CH toolkit[28] developed in our MILE laboratory for segmentation of word images. Initially, this MAST toolkit was planned to annotate scene images at pixel level.[20] We have additionally included character segmentation from word images. The datasets used in the experiment do not provide pixel level ground truth for word images. We have manually annotated word images to create ground truth for these datasets. Manually annotated images were used to perform OCR and record word recognition rate.[26] The recognition rates for the ground-truth images are also tabulated in Tables 3 and 4.

5. DISCUSSION

Each of our processing steps plays an important role in improving the recognition rate on ICDAR datasets. From Tables 3 and 4, we can observe that the performance of NESP method far exceeds those of others on ICDAR 2011 dataset and is better than others in ICDAR 2003 dataset. Some of the algorithms in the literature did not include the low resolution and degraded images of ICDAR 2003 dataset in their performance evaluation. We have averaged the results of those algorithms over the entire dataset. We have used image statistics itself, while choosing the plane with maximum discrimination. Hence, this method resembles a bottom-up approach.

We can use the training images of the dataset to learn text labels, which provide information regarding position of text labels, which could be incorporated in the segmentation. This is a top-down view to improve the segmentation and word recognition. The stroke width information of characters may also improve the segmentation.

The algorithm successfully countered the adverse effects of low illumination, illumination variation and low resolution of word images. The plane selection criterion is improved by power-law transformation. However, it is not affected by image scaling. The Otsu results reported in Tables 3 and 4 are obtained on the gray images (i.e.

Table 4. Comparison of performances of NESP and other algorithms on ICDAR 2011 dataset.

Algorithm	Total edit distance	Word recognition rate (%)
Benchmark[26]	**60.1**	**86.7**
NESP algorithm	**186.8**	**72.8**
MAPS algorithm[27]	199.7	71.6
TH-OCR System	176.4	41.2
KAIST AIPR System	318.5	35.6
Neumann's Method	429.7	33.1
Otsu[1]	596.4	18.2
Kittler[2]	644.6	18.0
Sauvola[5]	763.5	15.9
Niblack[4]	1469.4	12.7

Table 5. Number of images chosen from different planes ny NESP on ICDAR 2003 and 2011 datasets.

Plane	ICDAR 2003 dataset	ICDAR 2011 dataset
Red	289	158
Green	309	196
Blue	153	120
Intensity	108	74
Lightness	251	168

Figure 9. Word images where proposed algorithm fails to recognize the words in them.

no plane selection), without scaling and text polarity inversion. Thus, we see that each of the proposed steps is novel and helps in improved segmentation and recognition of word images. Table 5 reports the number of images from ICDAR 2003 and 2011 datasets finally selected from different planes by the NESP algorithm using the maximum discrimination factor. We observe that the number of images selected from each of R, G and B planes is correlated with the word recognition rate reported in Tables 1 and 2.

Figure 9 shows two word images, where our algorithm fails to recognize words. Strong illumination and low contrast are the reasons for failure in those images. Artistic characters also create a problem, since even if binarization is proper, recognition result is poor.

6. CONCLUSION AND FUTURE WORK

We have proposed an algorithm for segmentation and recognition of words from camera captured word image datasets. The results in Table 4 show that it is not always a trivial task to recognize the word, even from the properly detected text bounding box. The performance of our method (72.8%) on ICDAR 2011 dataset far exceeds the performance of the competed techniques. This is achieved by breaking the whole segmentation and recognition chain into constituent parts. Due to artistic fonts and varying character widths, ideally we need to incorporate uniformity in character stroke width for improvement in recognition rates. However, in this work, we have not performed any direct operation to have uniformity of stroke width of characters; we have only normalized the image heights in the dataset.

REFERENCES

[1] Otsu, N.,"A Thresholding Selection Method from Gray-level Histogram," *IEEE Trans. SMC*, **9**, 62–66 (1979).

[2] Kittler, J., Illingworth, J. and Foglein, J., "Threshold selection based on a simple image statistic," *Computer Vision, Graphics, and Image Processing*, vol. 30, no. 2, 125–147, (1985).

[3] Canny, J., "A Computational Approach to Edge Detection," *IEEE Trans. PAMI*, **8**, 679–698 (1986).

[4] Niblack, W., [*An Introduction to Digital Image Processing*], Englewood Cliffs, N.J. Prentice Hall, 115–116 (1986).

[5] Sauvola, J.J. and Pietäikinen, M., "Adaptive document image binarization," *Pattern Recognition*, vol. 33, no. 2, 225–236, (2000).

[6] Gonzalez, R.C. and Woods, R.E.,[*Digital Image Processing*], Second Edition, Pearson Education (2002).

[7] Thode, Jr., H.C., [*Testing for Normality*], New York, Marcel Dekker, (2002).

[8] Matas, J., Chum, O., Urban, M. and Pajdla, T.,"Robust wide baseline stereo from maximally stable extremal regions," *BMVC*, 384–393 (2002).

[9] Lucas, S.M., et.al., "ICDAR 2003 Robust Reading Competitions: Entries, Results, and Future Directions," *International Journal on Document Analysis and Recognition*, **7**, 105–122 (2005).

[10] Lucas, S.M., "Text Locating Competition Results," in Proc. 8th *Intl. Conf. on Document Analysis and Recognition (ICDAR)*, 80–85 (2005).

[11] Liu, H. and Ding, X., "Handwritten Character Recognition Using Gradient Feature and Quadratic Classifier with Multiple Discrimination Schemes," in Proc. 8th *Intl. Conf. on Document Analysis and Recognition (ICDAR)*, 19–25 (2005).

[12] Thillou, C.M. and Gosselin, B., "Color binarization for complex camera-based images," in *Electronic Imaging Conference of the International Society for Optical Imaging*, (2005).

[13] Thillou, C.M. and Gosselin, B., "Color text extraction from camera-captured images: the impact of the choice of the clustering distance," Proc. 8th *Int. Conf. on Document Analysis And Recognition*, 312–316, (2005).

[14] Duda, R.O., Hart, P.E. and Stork, D.G., [*Pattern Classification*], Second Edition, Wiley (2006).

[15] Thillou, C.M. and Gosselin, B., "Color text extraction with selective metric-based clustering," *Computer, Vision and Image Understanding*, vol. 107, no. 2, 97–107, (2007).

[16] Kasar, T., Kumar, J. and Ramakrishnan, A.G., "Font and background color independent text binarization," Proc 2nd *Camera-based Document Analysis and Recognition (CBDAR)*, 3–9, (2007).

[17] Kasar, T. and Ramakrishnan, A.G., "COCOCLUST: Contour-based color clustering for robust binarization of colored text," Proc 3rd *Camera-based Document Analysis and Recognition (CBDAR)*, 11–17, (2009).

[18] Neumann, L. and Matas, J., "A Method for Text Localization and Recognition in Real-World Images," Proc. 10th *Asian Conf. on Computer Vision (ACCV)*, 770-783, (2010).

[19] Wang, K., Babenko, B. and Belongie, S., "End-to-End Scene Text Recognition," Proc. 13th *Intl. Conf. on Computer Vision (ICCV)*, 1457–1464, (2011).

[20] Kasar, T., Kumar, D., Anil Prasad, M.N., Girish, D. and Ramakrishnan, A.G., "MAST: Multi-script Annotation for Scene images Toolkit," Proc. *Joint workshop on Multilingual OCR and Analytics for Noisy and Unstructured Text Data*, pp. 1–8, Beijing, China, September 2011, http://mile.ee.iisc.ernet.in/mast/, (2011).

[21] Shahab, A., Shafait, F. and Dengel, A.,"ICDAR 2011 Robust Reading Competition - Challenge 2: Reading Text in Scene Images," Proc 11th *Intl. Conf. on Document Analysis and Recognition (ICDAR)*, 1491–1496, (2011).

[22] Mishra, A., Alahari, K., and Jawahar, C.V, "An MRF Model for Binarization of Natural Scene Text," Proc. 11th *Intl. Conf. of Document Analysis and Recognition*, 11–16, (2011).

[23] Zeng, C., Jia, W. and He, X., "An Algorithm for Colour-based Natural Scene Text Segmentation," Proc 4th *Camera-based Document Analysis and Recognition (CBDAR)*, 67–72, (2011).

[24] Mishra, A., Alahari, K. and Jawahar, C.V., "Top-Down and Bottom-Up Cues for Scene Text Recognition," Proc. *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, (2012).

[25] Kumar, D. and Ramakrishnan, A.G., "Power-law transformation for enhanced recognition of born-digital word images," Proc 9th *Intl. Conf. on Signal Processing and Communications (SPCOM) 2012*, (2012).

[26] Kumar, D., Anil Prasad, M.N. and Ramakrishnan, A.G., "Benchmarking recognition results on word image datasets," in *CoRR*, vol. abs/1208.6137, http://arxiv.org/abs/1208.6137 (2012).

[27] Kumar, D., Anil Prasad, M. N. and Ramakrishnan, A. G., "MAPS: Midline analysis and propagation of segmentation," Proc. 8th *Indian Conference on Vision, Graphics and Image Processing (ICVGIP 2012)*, 16-19 December 2012, IIT Bombay, Mumbai, India.

[28] Kumar, D., Anil Prasad, M. N. and Ramakrishnan, A. G., "Benchmarking recognition results on camera captured word images datasets, " Proc. *Workshop on Document Analysis and Recognition (DAR 2012)*, 16 December 2012, IIT Bombay, Mumbai, India.

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

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

[31] Inzisoft. http://www.inzisoft.com/english/.

[32] KAIST AIPR. http://ai.kaist.ac.kr/home/.

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