Color Components Clustering and Power Law Transform for the Effective Implementation of Character Recognition in Color Images

Ravichandran Giritharan and A.G. Ramakrishnan

Abstract The problem related to the recognition of words in the color images is discussed here. The challenges with the color images are the distortions due to the low illumination, that cause problems while binarizing the color images. A novel method of Clustering of Color components and Power Law Transform for binarizing the camera captured color images containing texts. It is reported that the binarization by the proposed technique is improved and hence the rate of recognition in the camera captured color images gets increased. The similar color components are clustered together, and hence, Canny edge detection technique is applied to every image. The union operation is performed on images of different color components. In order to differentiate the text from background, a separate rectangular box is applied to every edge and each edge is applied with Power Law Transform. The optimum value for gamma is fixed as 1.5, and the operations are performed. The experiment is exhaustively performed on the datasets, and it is reported that the proposed algorithm performs well. The best performing algorithm in ICDAR Robust Reading Challenge is 62.5% by TH-OCR after pre-processing and 64.3% by ABBYY Fine Reader after pre-processing. The proposed algorithm reports the recognition rate of 79.2% using ABBYY Fine Reader. This proposed method can also be applied for the recognition of Born Digital Images, for which the recognition rate is 72.5%, earlier in the literature which is reported as 63.4% using ABBYY Fine Reader.

Keywords Color component clustering · Canny edge detection · Power law transform · Binarization · Word recognition

R. Giritharan (✉)
E.G.S. Pillay Engineering College, Nagapattinam, Tamil Nadu, India
e-mail: rvenkkatprabu@gmail.com

A.G. Ramakrishnan
Medical Intelligence and Language Engineering Laboratory,
Indian Institute of Science, Bengaluru, Karnataka, India

1   Introduction

The recognition of texts in the image is an active research area. From the literature, it is reported that there are many different techniques proposed for the recognition of the text in the images. In those, only the grayscale images had been considered, and there involves two different processes. They are binarization and recognition. Binarization is an important process in the recognition, since optimum pre-processing is necessary for the proper recognition of the images. There are varieties of kind of OCRs available for Roman Scripts [1, 2]. But it is not necessary; the images submitted to those OCRs should produce good results. Basically, there are two different procedures involved in word recognition. The former one is detection or localization, and the latter one is recognition [3]. The best performing algorithm produced the result of recognition rate of 52% in ICDAR 2003 [4] and of 62.5% using TH-OCR and 64.3% using ABBYY Fine Reader in ICDAR 2011 [5]. As stated already, the recognition involves two stages: binarization followed by recognition. For recognition, there are training datasets or OCR Engines for Roman Scripts [7–10]. Hence, the latter part needs no more research expertise. It is only the earlier part; that is, the binarization process needs to be improved. Since the literature reported different algorithms for the binarization of the color images, none of them produced proper results. Hence, a novel algorithm of combing the method of color component clustering [6, 7] and Power Law Transform [8] is applied for the binarization of the images. Here the clustering of similar color components is a

Fig. 1 Some color image dataset considered

separate algorithm, and PLT is yet another different algorithm proposed separately for the binarization. But a novel algorithm is proposed by combining these two algorithms. Here we have considered certain scanned images containing different colors like pamphlets and posters are being considered (Fig. 1).

Here initially clustering of the color components is applied to the image, and the algorithm of PLT is applied. Since different color components are present in an image, the clustering algorithm gives the more efficient results. Since most of the characters seem to be linked with each other, Power Law Transform reduces the stroke width and hence separates the characters. By combining these two algorithms, the recognition rate would further be increased.

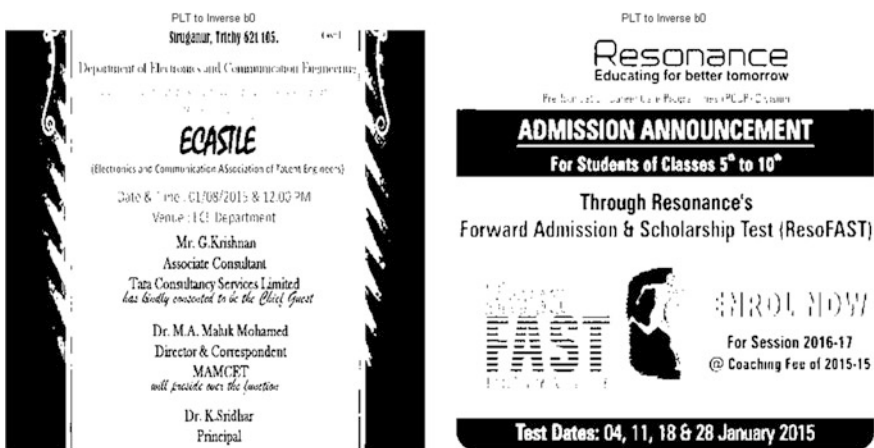**Fig. 2** Images in which PLT fails

Fig. 3 Binarized images by the PLT

2 Related Work

There are various algorithms proposed in the literature for the binarization of the color images. They include Global Threshold Algorithm [11], Adaptive threshold algorithm [8], Power Law Transform method [12], Clustering Algorithm [13–15] etc. All the algorithm have certain flaws, which reduce the recognition rate. Of these, power law transform is the most simple algorithm, with very good efficiency [8]. But in certain cases, the power law transform fails to binarize.

Figures 2 and 3 show the cases in which PLT algorithm fails to produce proper results.

3 Color Components Clustering

Text is one of the most important concentrations for the image which is taken into consideration. Since normally the colored images are multi-colored, the binarization becomes a challenging phenomenon. Hence, a novel approach of clustering different color components is considered. The major color planes of red, green, and blue color planes are considered. They are first extracted and the Canny edge detection [16–18] is performed on those extracted images independently and finally the union operation is performed between those three edge detected images. The edge map is obtained as shown below:
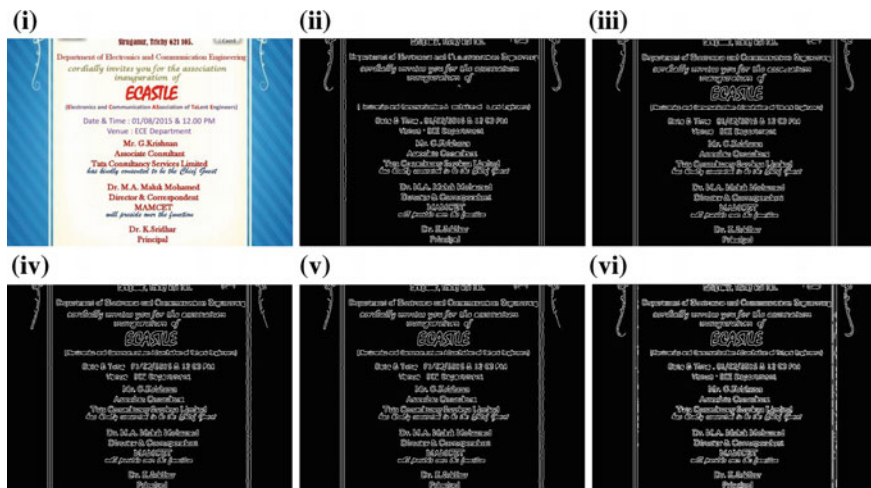

Fig. 4 Images considered and the subsequent images showing the edge detected *red*, *green*, and *blue* plane images and the final image representing the addition of all the three images (Color figure online)
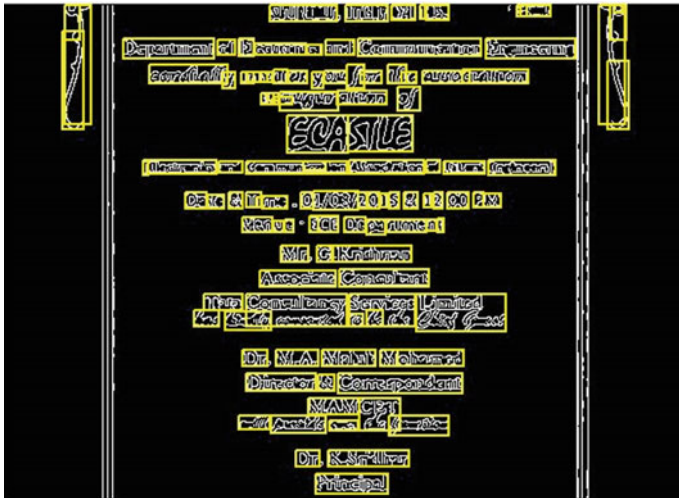
Fig. 5 Edge map image with edge box (EB)

$$E = E_R \vee E_G \vee E_B$$

Here E_R, E_G, and E_B are Canny edge detected image of red, green, and blue plane images, respectively, and v is the union indicating the OR operation. Consider the same image taken in which the PLT had failed (Fig. 4).

Once this is done, based on 8 connected component labeling independent rectangular box is applied. Each of these rectangular boxes is called as edge box (EB). There are certain constraints for these EBs. The aspect ratio is fixed within the range of 0.1–10. The aspect ratio is assigned for eliminating the elongated areas. The edge boxes' size should be more than 10 pixels and less than 1/5th of the total image size. Else they should be subjected to further pre-processing (Fig. 5).

Here there is a possibility of the edge boxes to be overlapped. But the algorithm is designed in such a way that only the internal edge boxes are taken into consideration and others are ignored.

4 Power Law Transform

After the clustering of the similar color components and creating independent edge boxes for every text character, the Power Law Transform is applied within every edge box so as to perform independent binarization within those edge boxes, the common thresholding technique of Global Thresholding, i.e., Otsu's method can be

applied after the Power Law Transformation [11]. Yet another problem is arriving the optimum threshold value. The histogram is split into two parts using the threshold value k. Hence, the optimum threshold value is arrived by maximizing the following objective function.

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

where,

$$\omega(k) = \sum_{i=1}^{k} p_i; \quad \mu(k) = \sum_{i=1}^{k} ip_i; \quad \mu_T = \sum_{i=1}^{k} ip_i$$

Here, the no. of gray levels is indicated by 'L' and 'p$_i$' is the normalized probability distribution of the image. One of the most important advantages of the Power Law Transform is that it increases the contrast of the image and creates well-distinguishable pixels for improved segmentation.
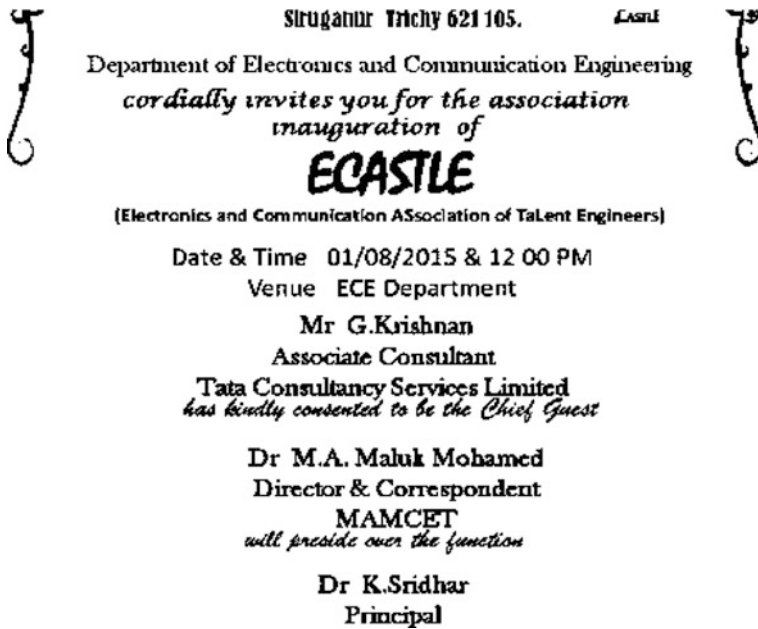
$$s = cr^\nu$$

Fig. 6 Independently binarized image using PLT

Here r is output intensity, s is input intensity, and c, υ are positive constants. Here the exponential component in Power Law is gamma. In experimentation, the gamma value is predefined as 1.5. To avoid another rescaling stage after PLT, the c is fixed as 1. With the increase in gamma value, the number of samples increases gradually, and for optimum segmentation, the gamma value for the PLT of images is fixed to 1.5 (Fig. 6).

5 Proposed Algorithm

As said above of the two important stages of recognition of text in images, binarization is an important task. And in case of color images, a novel method of combining the Clustering technique and PLT is utilized here, which shows more improved and universal results. Here the color components labeling is followed by the edge detection in separate clusters, edge map formation by union operation, and binarization using PLT method.

6 Recognition

Finally, the binarized image with text is given to OCR. Many OCRs are available such as OmniPage, Adobe Reader, ABBYY Fine Reader, and Tesseract. All the above-mentioned OCRs are relatively good, and hence, any of them can be used for the recognition of the binarized color images. Here, the train version of ABBYY Fine Reader is used. Performance evaluation is done with the ABBYY Fine Reader OCR's result.

7 Results

There are 2589 word images extracted from the color images as the training datasets. The testing dataset is of 781 word images. Table 1 compares the word recognition rate of proposed color clustering and PLT algorithms with those of other methods. Figure 7 shows certain images taken into consideration and their responses to the proposed algorithm.

Table 1 Performance evaluation of proposed color cluster and PLT algorithm on the color image dataset

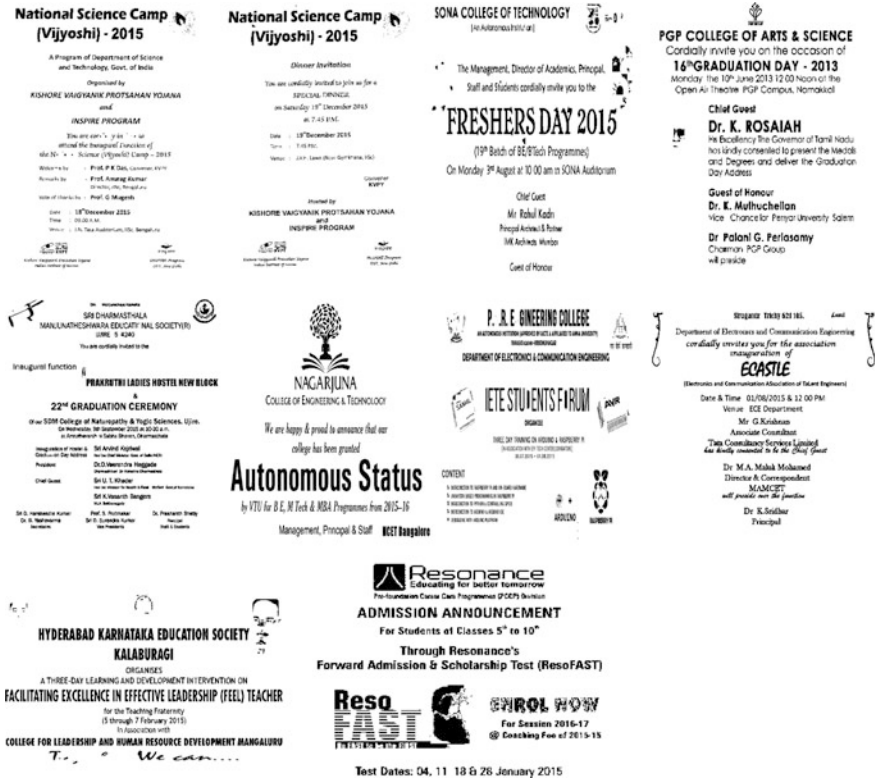| S. No	Algorithm	Word recognition rate (%)
1	Color cluster an PLT	79.20
2	TH-OCR	62.50
3	Baseline method	64.30

Fig. 7 Other examples showing the binary results obtained using the color clustering and PLT algorithm

8 Conclusion

A novel algorithm with the combination of color components clustering and PLT technique is proposed for improving the recognition of color word image dataset. From experimentation, it is obtained that the algorithms proposed in the literature are not sufficient for the word recognition in the color images. Also in case power law transform, one of the most simple and effective algorithm performs appreciably well on many image dataset. But it fails in certain cases as shown above. Hence, a novel method is proposed by utilizing the effective algorithm PLT along with the connected component algorithm. The results obtained show that the proposed algorithm performs even more well universally for all kinds of image datasets, even for the images which PLT failed to produce results, and consequently enhance the recognition rate.

References

1. Abbyy Fine reader. http://www.abbyy.com/.
2. Adobe Reader. http://www.adobe.com/products/acrobatpro/scanning-ocrto-pdf.html Document Analysis and Recognition, pp. 11–16, September 2011, 2011.
3. 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.
4. 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.
5. 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.
6. Thotreingam Kasar, Jayant Kumar and A.G. Ramakrishnan, "Font and Background Color Independent Text Binarization", Proc. II International Workshop on Camera-Based Document Analysis and Recognition (CBDAR 2007), Curitiba, Brazil, September 22, 2007, pp. 3–9.
7. T. Kasar and A. G. Ramakrishnan, "COCOCLUST: Contour-based Color Clustering for Robust Text Segmentation and Binarization," Proc. 3rd workshop on Camera-based Document Analysis and Recognition, pp. 11–17, 2009, Spain.
8. Deepak Kumar and A. G. Ramakrishnan, "Power-law transformation for enhanced recognition of born-digital word images," Proc. 9th International Conference on Signal Processing and Communications (SPCOM 2012), 22–25 July 2012, Bangalore, India.
9. J. N. Kapur, P. K. Sahoo, and A. Wong. A new method for gray-level picture thresholding using the entropy of the histogram. Computer Vision Graphics Image Process., 29:273–285, 1985.
10. C. Wolf, J. Jolion, and F. Chassaing. Text localization, enhancement and binarization in multimedia documents. ICPR, 4:1037–1040, 2002.
11. N. Otsu. A threshold selection method from gray-level histograms. IEEE Trans. Systems Man Cybernetics, 9(1):62–66, 1979.
12. Deepak Kumar, M. N. Anil Prasad and A. G. Ramakrishnan, "NESP: Nonlinear enhancement and selection of plane for optimal segmentation and recognition of scene word images," Proc. International Conference on Document Recognition and Retrieval(DRR) XX, 5–7 February 2013, San Francisco, CA USA.
13. Deepak Kumar, M. N. Anil Prasad and A. G. Ramakrishnan, "MAPS: Midline analysis and propagation of segmentation," Proc. 8th Indian Conference on Vision, Graphics and Image Processing (ICVGIP 2012), 16–19 December 2012, Mumbai, India.
14. 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.
15. 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.
16. J.Canny. Acomputationalapproachtoedgedetection. IEEE trans. PAMI, 8(6):679–698, 1986.
17. 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.
18. P. Clark and M. Mirmhedi. Rectifying perspective views of text in 3-d scenes using vanishing points. Pattern Recognition, 36:2673–2686, 2003.