A Framework for Detection and Extraction of Devanagari Text from Camera Based Images with 2-Stage Classification

Ganesh Kumar Sethi
Department of Computer Science, Multani Mal Modi College, Patiala, Punjab, India
ganeshsethi147@gmail.com

ABSTRACT: OCR is a field related to recognition of characters in scanned documents, PDF files or images captured by digital camera. With the advent of digital devices, a new branch of OCR has emerged that extract and recognizes the text from natural scene images captured by digital cameras. To extract text from these images a series of steps: pre-processing, binarization, segmentation, classification and recognition are required that differ from those applied in images scanned using a scanner. This paper proposes a 2-stage classification framework for detecting and extracting Devanagari text in camera based images.

KEYWORDS: Optical Character Recognition, Devanagari text, connected components, Text Information Extraction, Fine classification.

1. INTRODUCTION

OCR (Optical Character Recognition) is a field related to recognition of characters in scanned documents, PDF files or images captured by digital camera. Traditional OCRs use a flatbed scanner to input image with clean background, even illumination and minimal or no skew. With the advent of digital cameras, PDAs and mobile-cameras a new type of documents are originated. Documents are now not confined to the scanned images, but also contain images having text on real world objects taken from real scenes. Images captured with a digital camera like sign boards on roads, buildings, vehicle license plates, clothes contain text that need to be understood. This has emerged a sub-field of OCR for camera based digital images. A wide variety of images is taken with hand held devices like digital camera, PDAs, mobile cameras and hand held scanners. These documents differ from traditional scanned images due to complex background, uneven illumination, and variable font [1]-[2]. The traditional OCR technology that works for scanned images is not applicable here. This failure of traditional OCR system gave birth to new area of research- Camera Based Document Analysis and Recognition. The process carried out before passing image to OCR is Text Information Extraction (TIE) from images. Text information extraction (TIE) from digital images is a task performed on digital images to extract and recognize text content in them. A large number of fields have applications like web search [3-4], license plate recognition [5-6], helping visually Impaired [7], robot navigation [8],

***© 2018 Vidya Publications.* Authors are responsible for any plagiarism issues**

wearable computers [9-10], industrial automation [11], foreign language understanding [12] etc. linked to this extracted text.

(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Figure 1: Example Camera-images taken from Database exhibiting (a) non-uniform illumination (b)-(c) presence of multi-script content and Complex Background (d) Arbitrary

***© 2018 Vidya Publications.* Authors are responsible for any plagiarism issues**

text orientation and layout (e) Artistic fonts (f) Text like texture in background (g) Blurring (h) Arbitrary layout of text (i-j) No lightening (k-l) Images taken from distance by zooming.

2. RESEARCH CHALLENGES

State-of-the-art OCR systems have high accuracy for good quality, high resolution document images having a uniform and clean background. But most of the real life images captured with a digital camera do not satisfy these pre-requirements. Reading text in Camera captured images is a challenging task as the text can vary in font, color, size, orientation, and position

Further complexity may be due to the presence of non-uniform background and uneven illumination. Text can also be blurred due to motion or may be occluded by other objects [1] - [2], [13]. As different signs exist in three-dimensional space, text on these signs in scene images may be distorted by tilt, slant and shape of objects on which they are present. Text in addition to the horizontal left-to-right orientation may have other orientations including vertical, slanted, circularly wrapped around some object, and even mixed orientations within the same text area, such as text on a T-shirt or wrinkled sign. These factors can deteriorate the performance of any OCR system. Further camera captured images also face the problem of image acquisition, resolution, focus alignment and distance between camera and the object. Some challenging images exhibiting uneven illumination, perspective distortion, complex backgrounds, arbitrary layouts of text in image, textured appearance, color homogeneity, low contrast between text and background, multi-lingual images, limitation of Imaging device, problem in Image acquisition, resolution are shown in Figure 1.

3. PROPOSED FRAMEWORK FOR TEXT INFORMATION EXTRACTION

The proposed methodology for extracting Devanagari Text from camera based images is shown in Figure 2. As shown in figure, the process consists of a number of modules explained in subsequent sub-sections:

***© 2018 Vidya Publications.* Authors are responsible for any plagiarism issues**

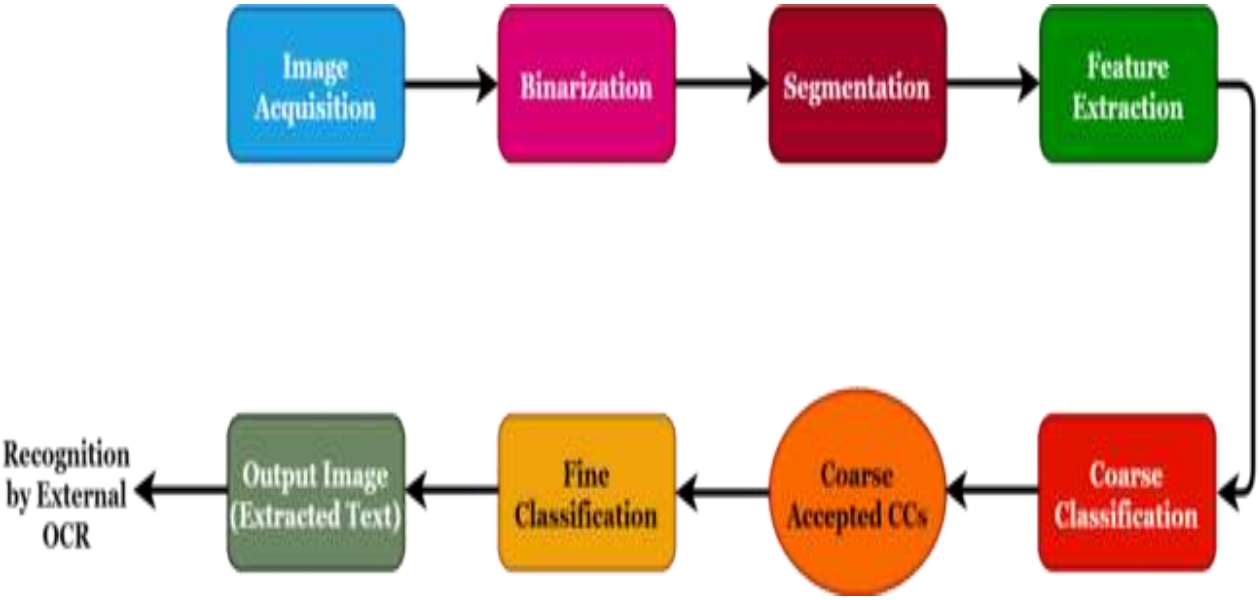


Figure 2: Proposed Methodology to extract Text from Camera Based Images

3.1 IMAGE ACQUISITION

Image acquisition refers to acquiring the image for the purpose of extracting textual content from it. The image can be acquired with the help of a digital camera, cell phone camera, PDA or any other hand held imaging device. Images are captured from natural scenes such as sign-boards (traffic signals, advertisement boards), containers, milestones, book covers etc. A number of datasets are available for the Roman [14]-[19], Chinese [20], Korean [21] and muti-scripts [20]-[21] that are being used for ranking different algorithms in various competitions. For Devanagari script, benchmark database of camera captured images is not available. So for the present work we have created our own database. Aks, the proposed database of images is created by taking images from different environments. Camera captured images has a useful application in understanding the natural scene texts. So most of the database is dedicated towards natural scene text images. The proposed dataset is divided into two categories: Devanagari Images and Multi-script images. In each category we further collected images in Natural light (Outdoor and indoor), Artificial light (Outdoor and indoor) and no light at all. Some images are also taken from internet. The images are captured from road-sides (sign boards), name plates (indoor and outdoor), hospitals, schools, historic monuments, book titles etc to cover different type of backgrounds. The captured images have different type of fonts with variable font size. The image may contain single word or multiple words arranged in different patterns of lines. The text appearing in the image may be machine-printed or handwritten. The images are captured with Nikon 14.1MP camera, Mobile camera (Sony 5MP and Motorola 16MP) set at different resolutions and modes. Sample

***© 2018 Vidya Publications.* Authors are responsible for any plagiarism issues**

images from our dataset are shown in figure 3 which consists of Outdoor images captured in Natural Light, indoor images captured in Natural Light, indoor images captured in Artificial Light, outdoor images captured in Artificial Light, Images captured in No Light (only flash of camera), and Images taken from internet.

3.2 Binarization of Camera-Based Images

The input image is a colored image and is converted to gray-scale image. A gray scale image has different intensity levels (0 to 255). For binarizing a gray scale image, intensities are to be classified into two categories that require Threshold intensity (T). All the pixels having value more than T are converted to 1 and having value less than T are converted to 0. If f(x, y) is input gray scale image and g(x, y) is binarized image then:

$$g(x,y) = \begin{cases} 0 & if\ f(x,y) < T \\ 1 & if\ f(x,y) \geq T \end{cases} \tag{3.1}$$

Several binarization techniques have been discussed in the literature. A chronological listing of reviewed techniques is given in Table 1.

(a) (b) (c)

(d) (e) (f)

***© 2018 Vidya Publications.* Authors are responsible for any plagiarism issues**

(g) (h) (i)

(j) (k) (l)

Figure 3: Camera-images containing Devanagari text taken from Aks Database (a)-(b) Outdoor images captured in Natural Light (c)-(d) indoor images captured in Natural Light (e)-(f) indoor images captured in Artificial Light (g)-(h) outdoor images captured in Artificial Light (i)-(j) Images captured in No Light (only flash of camera) (k)-(l) Images taken from internet.

Method	Year	Category	Approach
Otsu [23]	1979	Global	Histogram Analysis to minimize within class variance
Iterative [24]	1983	Global	Iterations till convergence
Minimum Error Thresholding [25]	1985	Global	Optimise Average Pixel Classification Error Rate
Niblacks [26]	1986	Local	Mean and Standard deviation of sub-image
Sauvola [27]	2000	Local	Mean and Standard deviation of sub-image with dynamic range of standard deviation
B. Gato et. al. [28]	2005	Local	Sauvola method with nearest neighbour interpolation
T.Kasar [29]	2007	Local	Edge Based Binarization
Extreme Value Based [30]	2010	Global	Gamma Transforms and Extreme Value Theory
Y.T. Pai et. al. [31]	2010	Hybrid	integration the global and local methods

***© 2018 Vidya Publications.* Authors are responsible for any plagiarism issues**

Z Zhou et. al. [32]	2010	Local	edge detection for video text images
B. Su, et. al. [33]	2013	Local	Adaptive Image Contrast
V. Morteza et. al. [34]	2013	Local	Adaptive water flow model for the binarization of degraded document images

Table 1: Chronological List of various Binarizing Techniques

3.2.1 Proposed Framework for Binarization

For the purpose of text detection and extraction from a camera captured image, an algorithm is required that can deal with uneven illumination, complex backgrounds, variable font size, a variety of font styles, low contrast between text and background etc. The proposed methodology for binarization [35] is based on edge detection and morphological dilation. The flowchart of the proposed methodology is presented in Figure 2. Corresponding intermediate results are shown in Figure 3. The results of binarization on our dataset and on bench-mark datasets are shown in Figures 4-7.

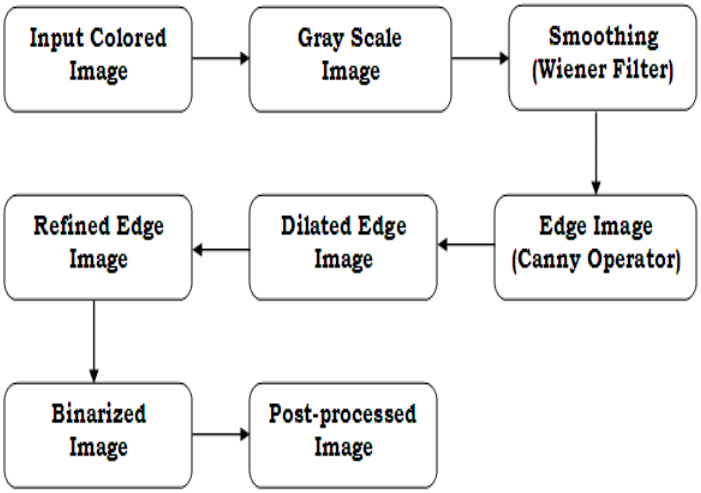


Figure 4: Proposed Method for Binarization of Images

(a) (b) (c)

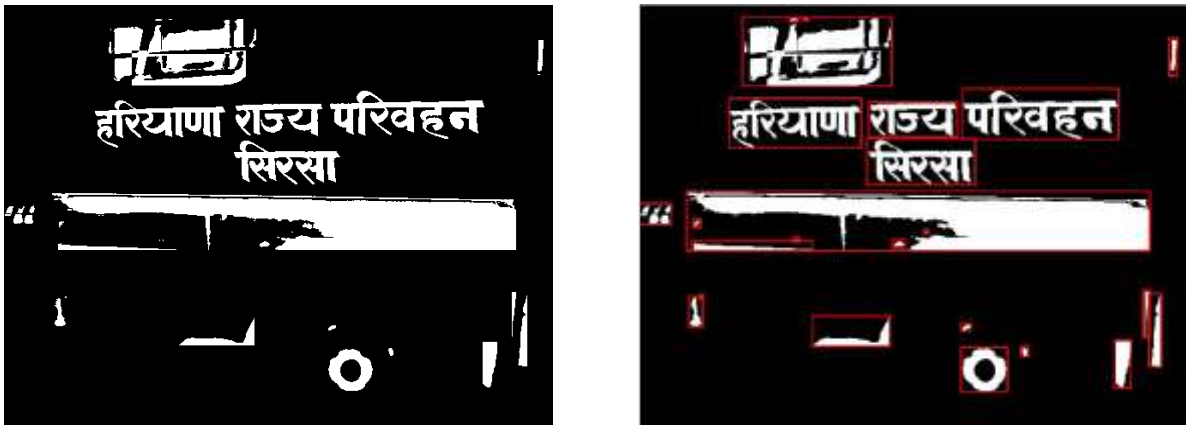
Figure 5: (a) Original Colored Image Captured by digital camera (b) Gray Scale Image after preprocessing (c) Image after performing Binarization and post processing

3.3 Segmentation

Segmentation is an operation that decomposes a binary image into sub-parts. The segmentation divides the binary image into constituent parts which are classified into text or non-text category. The

***© 2018 Vidya Publications.* Authors are responsible for any plagiarism issues**

connected components, obtained from the binarized image, are labeled with numbers 1,2,3…….. using the algorithm [36]. The extracted and labeled components are marked with bounding boxes as shown in Figures 6 and 7.



(a) (b)

Figure 6: (a) Input Binary Image (b) Extracted Connected Components shown in bounded boxes.

Figure 7: Some Extracted components from Binary Image

Feature-Class	Feature
Geometric Features	$AreaRatio = \frac{area(CC)}{area(\mathrm{Im}age)}$
	$AspectRatio = \min(\frac{Width(CC)}{Height(CC)}, \frac{Height(CC)}{Width(CC)})$
	$LengthRatioS = \min(\frac{Width(CC)}{Width(\mathrm{Im}age)}, \frac{Height(CC)}{Height(\mathrm{Im}age)})$
	$LengthRatioL = \max(\frac{Width(CC)}{Width(\mathrm{Im}age)}, \frac{Height(CC)}{Height(\mathrm{Im}age)})$

***© 2018 Vidya Publications.* Authors are responsible for any plagiarism issues**

Shape Features	$OccupyRatio = \frac{\lvert CC \rvert}{\lvert area(CC) \rvert}$
	$Compactness = \frac{\lvert area(CC) \rvert}{\lvert Contour(CC) \rvert^2}$
	$ContourRoughness = \frac{\lvert CC - imopen(CC, 2*2) \rvert}{\lvert (CC) \rvert}$
Stroke Features	$StrokeWidthMean = mean(Strokewidth(edge(CC)))$
	$StrokeWidthDev = \frac{dev(Strokewidth(edge(CC)))}{mean(Strokewidth(edge(CC)))}$
Edge Features	$EdgeContrast = \frac{\lvert Contour(CC) \cap edge(image) \rvert}{\lvert Contour(CC) \rvert}$
Texture Features	$CVF = \frac{dev(F(CC)))}{mean(F(CC))}$
	$CVB = \frac{dev(B(CC)))}{mean(B(CC))}$
Projection Profile	$HPP = \sum_{1 \le y \le n} f(x, y)$

Table 2: Extracted Features

3.4 FEATURE EXTRACTION

The term feature, in general, specifies some characteristics of an object. For classification tasks, feature is considered as a mathematical function of one or more measurements. This function computes some quantifiable properties from an object that represent significant characteristic of that object. Feature plays a significant role in the area of image processing. In the proposed work feature extraction is required to classify a component of image as text component or non-text component. Text components usually have distinct properties that can be quantified based on shape, stroke, geometry etc. Renjie Jiang [37] extracted 15 features based on geometry, shape regularity, edge, stroke and spatial coherence to classify roman characters in natural scene images as text or non-text. For the presented work, once the connected component labeling of binary image has been done, the features are extracted which are designed using the projection profile, geometry, boundary,

***© 2018 Vidya Publications.* Authors are responsible for any plagiarism issues**

stroke and texture properties from these components (Table 2).

3.5 CLASSIFICATION

The next step is to label the segmented components from binary image into text components and non-text components using these extracted features [37]. The non-text components are discarded whereas text components are retained and further fed to an external OCR. The classification architecture is divided into two stages: coarse classification using heuristic filtering and fine classification using machine learning approach.

3.5.1 HEURISTIC FILTERING OF NON-TEXT COMPONENTS

The text has distinguished features and certain thresholds can be experimentally determined to discard most of the non-text components. The components that pass the threshold limits are accepted as text candidates.

$$d_i(x) = \begin{cases} 1 & if\ THL_i \le f_i(x) \le THU_i \\ 0 & otherwise \end{cases} \quad (3.2)$$

Where x is the input connected component, THL_i and THU_i are the lower threshold limit and upper threshold limit for the ith feature. $d_i(x)$ is the decision parameter for the component x: 1 means the component is accepted, 0 means component is rejected.

As shown in Figure 8 at the beginning [38], all CCs extracted in the segmentation step are feed into the first weak classifier. It measures a particular feature on CCs one by one and categorizes the fed connected components into positive (candidate for text) or negative (non-text). The negative CCs are considered as non-text and rejected immediately. For positive CCs, similar procedure is repeated in following weak classifiers until the end of the cascading process. In the coarse result after this cascade, most of non-text CCs are discarded and almost all the text CCs are preserved. The preserved CCs are further classified into text or non-text by machine learning process.

3.5.2 FINE CLASSIFICATION

The cascading process as explained previously saves a huge computation on all the connected components as a component is immediately discarded whenever it fails a particular weak classifier. The extracted features from connected components are stored as an n-tuple feature vector with normalized value for each feature. The components that pass the heuristic filtering need to be classified in stage 2 of classification. For machine learning, the connected components are manually labeled as text or non-text to prepare training dataset. Some examples of positive samples are shown in Table 3 and negative samples are shown in Table 4.

***© 2018 Vidya Publications.* Authors are responsible for any plagiarism issues**

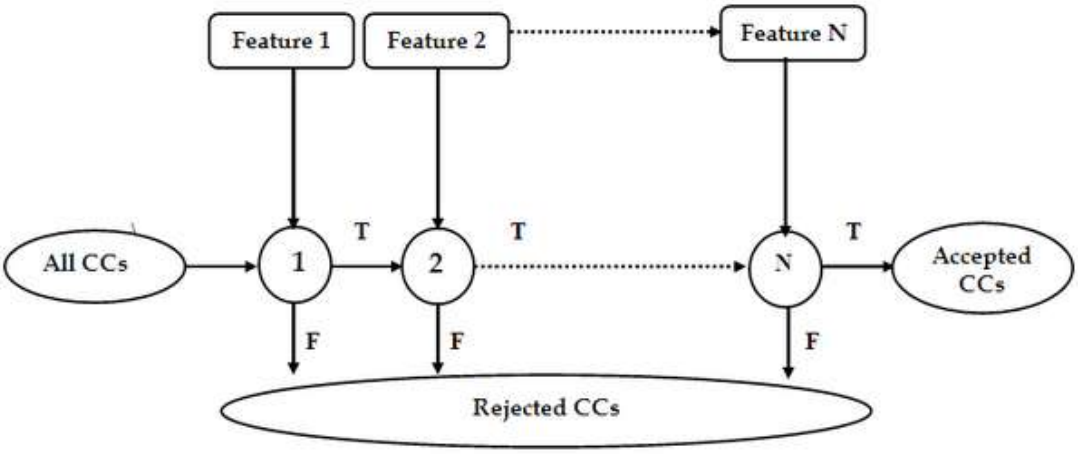


Figure 8: Non-text connected component filtering

सिहाग	नाहरगढ	निकास
क्षण	समय	प्रयास
दवाईयाँ	विश्वविद्यालय	सिटी
करना	फेस	मयूर
नोएडा	डेयरी	प्रसव

Table 3: Some Positive Samples from Binary Images

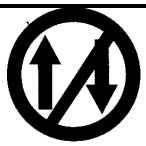
		

Table 4: Some Negative Samples from Binary Images

For machine learning, we have used SVM, nearest neighbour classifier and Naïve Baysian classifier. The dataset is divided into two parts-training set and testing set. The system is trained using components extracted and marked from training images and tested on components on testing images. The precision,

***© 2018 Vidya Publications.* Authors are responsible for any plagiarism issues**

recall, f-measure and accuracy of various categories are shown in table 5.

Image Class	No. of Images	Precision rate	Recall Rate	F-measure
Natural Outdoor	190	78.68	98.45	86.74
Natural Indoor	70	75.64	95.37	81.63
Artificial Outdoor	40	88.46	87.70	87.84
Artificial Indoor	70	72.41	87.52	77.59
No Light	25	90.92	89.18	89.90
Internet Images	60	66.65	96.70	75.51
Average	**455**	**78.05**	**95.66**	**84.26**

Table 5: Acquired Results

Figure 9: First column shows Original Image, Text is localized in column 2, Image after Text Extraction is shown in column 3 and in column 4 Ground Truth Image is given.

3.6 RECOGNITION

The extracted text is converted to plain text using OCR technology or it may be

***© 2018 Vidya Publications.* Authors are responsible for any plagiarism issues**

converted to speech using Text to Speech converters (TTS). For the proposed study extracted text is recognized using an external OCR of Hindi Language.

4. Conclusion and Future scope

In this paper a technique for detecting and extracting Devanagari Text from camera based natural scene images has been presented and the extracted text components have been recognized with the help of an external OCR. The results obtained are very promising. In future, the recognized text can be converted to speech which can be further translated to any language with the help of Text to Speech conversion and speech translation.

Figure 10: First column shows Original Image, OCR performance on original image is shown in column 2, Image after Text Extraction is shown in column 3 and in column 4 shows the OCR performance on Text extracted image.

***© 2018 Vidya Publications.* Authors are responsible for any plagiarism issues**

REFERENCES

[1] J. Liang, D. Doermann, H. Li, "Camera-based analysis of text and documents: a survey", *International Journal on Document Analysis and Recognition (IJDAR)*, Volume 7, Issue 2-3, pp. 84-104, 2005.

[2] K. Jung, K. I. Kim, A. K. Jain, "Text Information Extraction in Images and Video: a Survey", *Pattern Recognition*, Volume 37, pp. 977-997, 2004.

[3] J. Zhou, D. Lopresti, Z. Lei, "OCR for world wide web images" *Proceedings of SPIE on Document Recognition(IV)*, Volume 3027, pp. 58–66, 1997.

[4] D. Lopresti, J. Zhou, "Locating and recognizing text in WWW images", *Information Retrieval*, Volume 2, Issue 2-3, pp. 177-206, 2000.

[5] Y. Cui, Q. Huang, "Character extraction of license plates from video" *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 502–507, 1997.

[6] S.H. Park, K.I. Kim, K. Jung, H.J. Kim, "Locating car license plates using neural networks", *IEEE Electron*, Volume 35, Issue 17, pp. 1475–1477, 1999.

[7] N. Ezaki, M. Bulacu and L. Schomaker "Text Detection from Natural Scene Images: Towards a System for Visually Impaired Persons", *In Proceedings of 17th International Conference on Pattern Recognition (ICPR 2004)*, Volume 2, pp. 683-686, 2004.

[8] G. N. DeSouza, A. C. Kak, "Vision for mobile robot navigation: A survey", *IEEE Transactions PAMI*, Volume 24, Issue 2, pp. 237–267, 2002.

[9] Y. Watanabe, Y. Okada, Y.B. Kim, T. Takeda, "Translation camera", *In Proceedings of International Conference on Pattern Recognition*, Volume 1, pp. 613–617, 1998.

[10] I. Haritaoglu, "Scene text extraction and translation for handheld devices", *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Volume 2, pp. 408–413, 2001.

[11] Y. K. Ham, M. S. Kang, H. K. Chung, R. H. Park, and G. T. Park "Recognition of raised characters for automatic classification of rubber tires" *Optical Engineering*, Volume 34, Issue 1, pp. 102–109, 2005.

[12] J. Yang, X. Chen, J. Zhang, Y. Zhang, and A. Waibel, "Automatic detection and translation of text from natural scenes", *In Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Volume 2, pp. 2101-2104, 2002.

[13] C. Yao, X. Zhang, X. Bai, W. Liu, Y.Ma, and Z. Tu, "Rotation-invariant features for multi-oriented text detection in natural images", *PLoS One*, Volume 8, Issue 8, 2013.

[14] S.M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong, and, R. Young, "ICDAR 2003 robust reading competitions", *In Proceedings of International Conference on Document Analysis and Recognition*, pp. 682 – 687, 2003.

[15] A. Shahab, F. Shafait, and A. Dengel, "ICDAR 2011 Robust Reading Competition Challenge 2: Reading Text in Scene Images", *In Proceedings of the International Conference on Document Analysis and Recognition*, pp. 1491–1496, 2011.

[16] D. Karatzas, F. Shafait, S. Uchida, M. Iwamura, L. G. Bigorda, S. R. Mestre, J. Mas, D. F. Mota, J. A. Almazan, and D. L. Heras, "ICDAR 2013 Robust Reading Competition", *In Proceedings of the 12th International Conference on Document Analysis and Recognition*, pp. 1115–1124, 2013.

***© 2018 Vidya Publications.* Authors are responsible for any plagiarism issues**

[17] K. Wang and S. Belongie, "Word spotting in the wild", *In Proceedings 11th European Conference on Computer Vision*, pp. 591–604, 2010.

[18] R. Nagy, A. Dicker and K. Meyer-Wegener, "NEOCR: A Configurable Dataset for Natural Image Text Recognition", *In Proceedings of CBDAR Workshop at ICDAR*, pp. 53-58, 2011.

[19] R. Nagy, A. Dicker, and K. M. Wegener, "Definition and Evaluation of the NEOCR Dataset for Natural Image Text Recognition", *University of Erlangen, Dept. of Computer Science, Technical Reports*, 2011.

[20] C. Yao, X. Bai, W. Liu, Y. Ma and Z. Tu, "Detecting Texts of Arbitrary Orientations in Natural Images", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1083-1090, 2012.

[21] http://www.iapr-tc11.org/mediawiki/index.php/KAIST_Scene_Text_Database.

[22] T. Kasar, D. Kumar, M.N. Anil Prasad, D. Girish and A.G. Ramakrishnan, "MAST: Multi-Script Annotation toolkit for Scenic Text," *In Proceedings of Joint Workshop on Multilingual OCR and Analytics for Noisy Unstructured Text Data*, pp. 1-8, 2011.

[23] N. Otsu, "A threshold selection method from gray level histograms" *IEEE Transactions on Systems, Man, and Cybernetics (SMC-9)*, pp. 62-66, 1979.

[24] R. C. Gonzalez and R.E. Woods, "Digital Image Processing" Third Edition, pp. 763-764, 2009.

[25] J. Kittler, and J. Illingworth, "Minimum error thresholding", *Pattern Recognition*, Volume 19, Issue 1, pp. 41-47, 1986.

[26] W. Niblack, "An Introduction to Image Processing" Prentice-Hall, Englewood Cliffs, NJ, 1986.

[27] J. Sauvola and M. Pietikainen, "Adaptive document image binarization", *Pattern Recognition*, Volume 33, Issue 2, pp. 225-236, 2000.

[28] B. Gatos, I. Pratikakis and S.J. Perantonis, "Text Detection in Indoor/Outdoor Scene Images", *First International Workshop on Camera-based Document Analysis and Recognition (CBDAR'05)*, pp. 127-132, 2005.

[29] T. Kasar, J. Kumar and A.G. Ramakrishnan, "Font and Background Color Independent Text Binarisation", *Camera Based Document Analysis and Recognition*, pp. 3-8, 2007.

[30] B. Fernando and S. Karaoglu, "Extreme Value Theory Based Text Binarization In Documents and Natural Scenes", *The 3rd International Conference on Machine Vision (ICMV 2010)*, pp. 144-151, 2010.

[31] Y.T. Pai, Y.F. Chang, S.J. Ruan, "Adaptive thresholding algorithm: efficient computation technique based on intelligent block detection for degraded document images", *Pattern Recognition*, Volume 43, Issue 9, pp. 3177–3187, 2010.

[32] Z. Zhou, L. Li, C.L. Tan, "Edge based binarization for video text images", *In Proceedings of 20th International Conference on Pattern Recognition (ICPR)*, pp. 133–136, 2010.

[33] B. Su, S. Lu, C.L. Tan, "Robust document image binarization technique for degraded document images", *IEEE Transactions on Image Processing*, Volume 22, Issue 4, pp. 1408–1417 , 2013.

[34] V. Morteza, K. Ehsanollah, "An adaptive water flow model for binarization of degraded

***© 2018 Vidya Publications.* Authors are responsible for any plagiarism issues**

document images", *International Journal of Document Analysis and Recognition (IJDAR)*, Volume 16, Issue 2, pp. 165–176, 2013.

[35] R. K. Bawa, G.K. Sethi, "A Binarization Technique for Extraction of Devanagari Text from Camera Based Images", *Signal and Image Processing: An International Journal (SIPIJ)*, Volume 5, Issue 2, pp. 29-37, 2014.

[36] R.M. Haralick and L.G. Shapiro, "Computer and Robot Vision I", Addison-Wesley, pp. 28-48, 1992.

[37] R. Jiang, F. Qi, L. Xu, G. Wu, "Detecting and Segmenting Text from Natural Scenes with 2-Stage Classification", *Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications*, Volume 2, pp. 819-824, 2006.

[38] http://www.indsenz.com/int/index.php.

***© 2018 Vidya Publications.* Authors are responsible for any plagiarism issues**