Journal of Advanced Zoology

ISSN: 0253-7214
*Volume **44** Issue **S-6 Year 2023** Page **767:774***

Combined Contrast Enhanced and Wide-Baseline Technique for Kannada Text Detection in Images

Shahzia Siddiqua[1], C. Naveena, Sunilkumar S. Manvi[2], *Sr.*, *Member*[3], *IEEE*[4]

**Corresponding author's E-mail: Shahzia Siddiqua*

Article History

Received: 06 June 2023
Revised: 05 Sept 2023
Accepted: 30 Nov 2023

Abstract

Text characters contained in images are a valuable source of information for content-based indexing and retrieval applications. These text characters are difficult to identify and distinguish due to their various sizes, grayscale values, and intricate backgrounds. The paper presents a new method for identifying text contained in images of any grayscale value. The proposed scheme uses a combination of contrast-limited adaptive histogram equalization (CLAHE) algorithm, which enhances the local contrast and limits any noise in the image, and the wide baseline image matching technique which helps locate an object in the image. Applying a series of morphological operations and filtering at each stage, the resultant component is the detected text which is either a character, word or a line segment. MATLAB based simulation and evaluation on a self-curated Kannada, a popular south Indian language and other standard datasets proves that the proposed technique outperforms other methods consistently on precision, recall and F1-score. Importantly, on the Kannada dataset, it returns the highest recall of 98% since the system is specifically tuned for its linguistic features proving its robustness. Further, the proposed technique can be extended to image pre-processing tasks for deep learning models to improve their accuracy and for text recognition tasks.

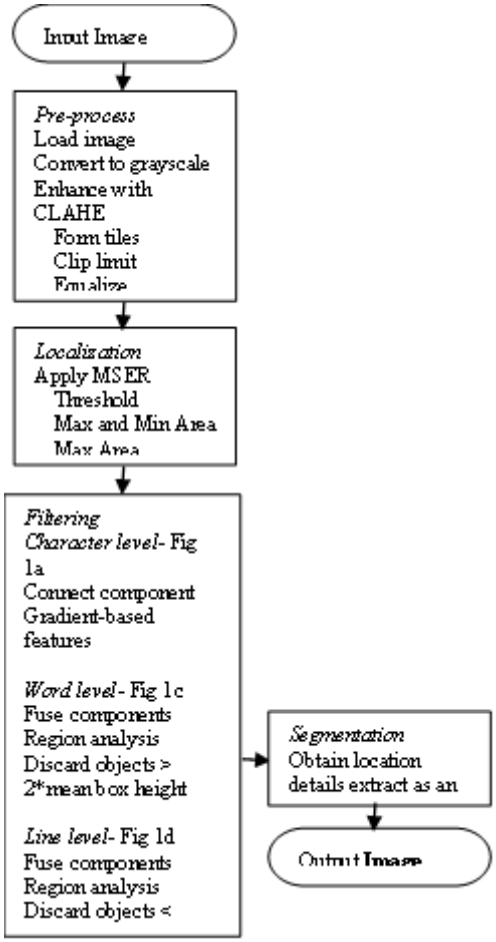
Keywords: *Contrast Limited Adaptive Histogram Equalization, Text Detection, Wide Baseline Image Matching*

1. Introduction

The fast expansion of camera-based apps that are easily available on smart phones and social media is posing a challenge to image processing businesses, necessitating the development of effective scene-text recognition algorithms. Extracting text from images is an extremely challenging proposition, as a single image may contain text of assorted fonts, colors, sizes and orientation. Also, images having poor effects like presence or absence of light, shadow and reflections make the task even more strenuous. Therefore, Date Submitted for review: 11th May, 2021. No sponsor or financial support received for this work. Shahzia Siddiqua is a Research Scholar at REVA University, Bengaluru, Karnataka – 560064 India with (e-mail: sshaz79@ gmail.com). Naveena C is Professor at SJBIT, Bengaluru, Karnataka – 560060 (e-mail: naveena.cse@gmail.com). Sunilkumar S. Manvi is Director, School of Computing and IT at REVA University, Bengaluru, Karnataka – 560064 India (e-mail: ssmanvi@reva.edu.in). there exists an inherent need for a robust text detection system to tackle this image deluge. Based on the methodology of candidate extraction, text detection techniques are broadly categorized into texture, region or hybrid methods. In texture-based method, the picture is skimmed categorizing the nearby pixels based on the textural attributes such as wavelets, rim-concentrations, discrepancy in intensities and gradients. The method uses classifiers like Support Vector Machine (SVM), neural networks, AdaBoost, etc. to detect texture, blur and noise related distortions but for horizontal text only. However, the disadvantages are high computational cost, time and they rarely distinguish color differences. Region-based methods include both connected component (CC)-based and rim-based techniques. The connected component method fragments the text image into several components on a particular geometrical attribute, unifying smaller connected components to form bigger one and splitting the textual connected components with the help of measured geometrical limits. Some classifiers used for this are Random Forest algorithm and classifier chains [1], Stroke Width Transform [2], Maximally Stable Extremal Regions (MSER) [3], Extended MSER and Extremal regions [4]. These methods are straightforward, effective and detect

multi-level text, but less robust for images with complex background and text-alike elements generating false positives. Edge based techniques benefit from the existence of elevated contrast amongst text and backdrop, making the border a significant character in an image. Here, rims of the script frontier are identified and clustered and non-text sections are sifted considering various heuristics. A few efforts using this technique are two-step catalogue and group [5], Projection profile [6] etc. They are quick and produce decent results for images having solid boundaries, text with assorted letterings, dimensions and colors. However, they cannot handle oriented text, intricate backdrops and similarity between background and the actual text, creating false positives. Hence, region and texture-based approaches possess both advantages as well as drawbacks. Combining both approaches, a few of their shortcomings have been overcome. These are known as Hybrid methods. Some examples of this are graph paradigm on MSER merged with area based and context related data to tag CCs [7], integrating CC and texture-based techniques [8] etc.

The authors of [27] proposed a novel arbitrary-shaped text identification technique by framing text detection as a visual relationship detection issue, which resulted in a substantial improvement in text detection accuracy. They used a GCN-based visual connection detection framework to successfully use context information to increase link prediction accuracy, allowing our text detector to identify even text instances with huge inter-character or extremely short inter-line spacing. The authors of [28] suggested a novel object-text detection and identification technique called DetReco to identify objects and texts that recognize the text contents using Deep learning. Object-text detection network and text recognition network make up the suggested approach. The object-text detection work is handled with YOLOv3, while the text recognition duty is handled with CRNN. Kannada is a Dravidian language spoken primarily by the people of Karnataka in India's southern region. Some related work for detecting Kannada text from images is seen in [9] using canny edge and CC filters. A Discreet wavelet and MSER [18], Gabor sifter and k-means grouping [10] with wavelet entropy are also proposed. From research, it is observed that mainstream.



work has been done for foreign vocabularies like Chinese, English, Korean, Arabic, etc. however very little amount of effort has been accomplished in Indian local languages like Kannada, Tamil, Malayalam, etc. In this paper, major focus is on Kannada script which is very intricate as compared to English and Chinese scriptures. The key contributions of this paper are:

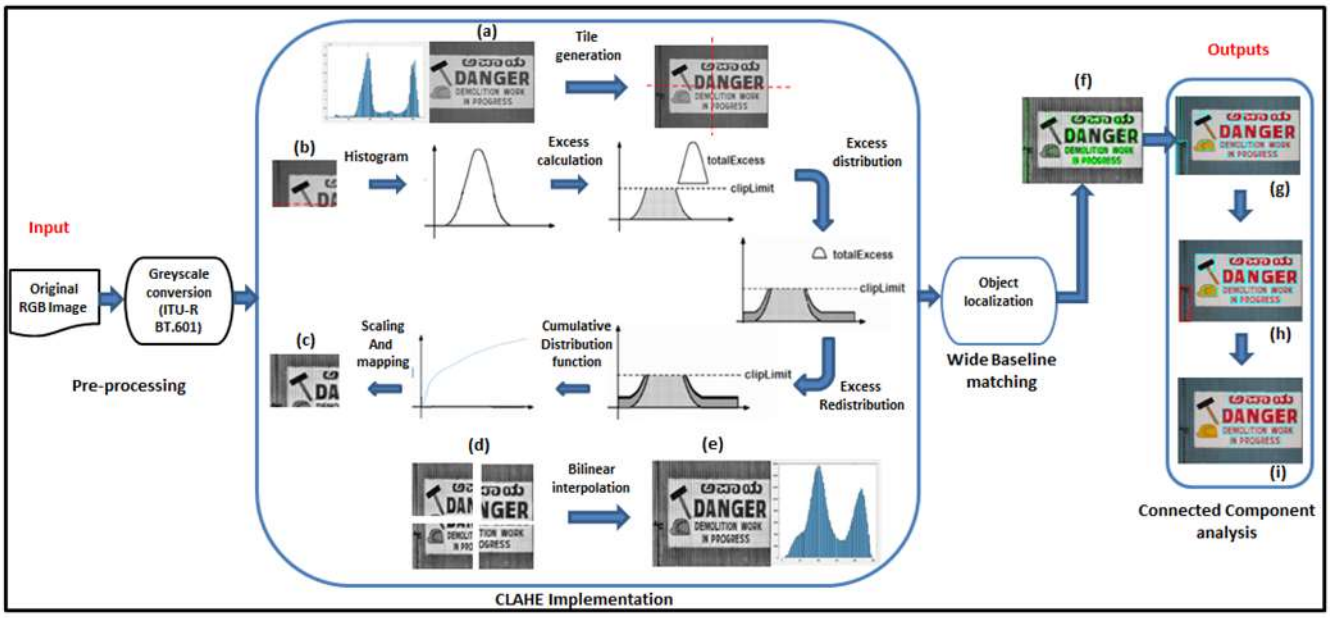
1) Curating a Kannada scenes dataset of 200 images and making it available to the research community.

2) Proposing a novel technique for detecting text in a scene image by returning a character, word, or line.

The following is a summary of the paper. The proposed technique is discussed in Section II. The results and analyses are presented in Section III. The conclusion and future study directions are covered in Section IV. The bibliography is included at the conclusion of the paper.

2. Materials And Methods

The proposed technique can be broadly divided into three major parts: Image Pre-processing and applying the contrast-limited adaptive histogram equalization (CLAHE) algorithm [11], detect stable object regions using wide-baseline method and finally.



use morphological operations like connected component analysis (CCA) to locate the text regions. Fig. 2 outlines the flowchart of the proposed method. Fig. 3 is a detailed schematic of the proposed methodology adopted which is explained below.

Fig 1: Image filtering, (a) character level connected components, (b) post filtering, (c) word level, (d) line level

A.Pre-processing of image

Here, the raw scene RGB image is loaded and the ITU-R BT.601 guidance is used to convert it into grayscale.

Fig. 2. Detailed flow-chart of the proposed model Next, we compute an inception I using (1) so as: (1)

to check for poor quality or blur in the image. We apply the CLAHE algorithm only if I<128 else, skip this step. This step activates smallest sections in the image, termed tiles. The histogram of every tile (Fig. 3 b) is calculated and a clip limit is obtained using equation (2). (2) After equalization (Fig.

3c), bilinear interpolation is done to combine neighboring tiles (Fig. 3d) and eliminate falsely generated boundaries. The histogram equalizer chooses the grayscale transformation T to minimize as in equation (3) Fig. 3. Detailed schematic of the proposed methodology. Steps (a)-(f) explained in

Dataset	Methods/Authors	Precision %	Recall %	F1-score %
South Indian language and Kannada dataset	Gabor Filter and k-means [10]	84	93	77
	Transforms and Gabor based [17]	84	98	73
	Discrete Wavelet and MSER [18]	82	64	79
	Proposed Method	**90**	**98**	**84**
MSRA-TD500	Yao et al. [19]	64	62	61
	Lu et al. [20]	29	63	40
	Yao et al. [21]	77	75	76
	Zhao et al. [22]	34	69	46
	Yin et al. [23]	81	63	71
	Kang et al. [24]	71	62	66
	Zhou et al. [26]	**87**	67	76
	Proposed Method	83	**90**	**85**
ICDAR 2015	EAST [25]	83	78	81
ICDAR 2013	Holistic [19]	72	59	65
ICDAR 2003	StradVision2 [26]	77	37	50
	NJU [26]	70	36	48
	CNN MSER [26]	35	34	35
	Proposed Method	**85**	**80**	**83**

section II. where h0 is the collective histogram and h1 is the collective addition of histogram for all concentrations s. The reduction is conditioned such that T needs to be monotonic and h1 (T(s)) cannot exceed h0(s) in more than half the space between the histogram counts. It is observed (Fig. 3e) that the dark objects appear darker and lighter areas appear brighter in the enhanced image than the original image. Also, the histogram comparison before and after the application of CLAHE technique shows this enhancement.

A. Object Localization

Here we binarize the image to a threshold value, set an area limit and maximum distance between two objects to detect object regions via wide-baseline matching approach [12]. On observing the presence of text in scenes, the threshold value is set to 3, minimum area is set to 40 pixels, maximum area to 8000 pixels and maximum space between the regions to 0.5, discarding regions greater than or lesser than this limit. Below equation shows the growth measure function $q(t_j)$ defined as the derivative of the region or area above the threshold values reaching a local minimum. $Q(t_j)$ is a binarized region where tj shows its threshold level and the difference $Q(t_j)-Q(t_j-1)$ describes the threshold increment Δtj. On applying equation (4) the regions are obtained as a centroid as shown in Figure 3(f).

$$q(t_j) = \frac{\| Q(t_j) - Q(t_j - 1) \|}{\| Q(t_j) \|} \quad (4)$$

B. Filtering and Segmentation

After identifying individual regions, we apply the connected component analysis to join nearby objects and Gradient Based Features [13] to remove smaller distant components. At this level single character segmentation is achieved as shown in Figure 1(g). Next we perform morphological operations to join adjacent characters. Region analysis for centroid and bounding-box properties to sieve uneven objects horizontally is carried out. For doing this, we extract vertical coordinates from the centroid of each component and compute the line spacing threshold using the mean of heights of all components in the image. Using these two values, histogram bin count is assigned as a line label for each component. Consequently, discard objects that are more than twice the mean height for each bin line. Figure 1(h) depicts this stage of word level detection with rejected components in red. To further fine-tune and connect words to form lines, the morphological operations are repeated. The bounding box and area properties of connected candidates are re-examined. Non-text components have a significantly lower area and width than text components. As a result, the residual components'

average area and mean width are calculated. The objects with area < 1/3rd and width < 1/4th fraction of their mean values are sieved out. The result at this stage is a single line component as shown in Figure 3(i). Further, the text can be segmented and saved as a single line image, using the properties of each resultant box. This extraction can be done at the character or even word level. The proposed model is simulated in MATLAB R2018a on a CORE-i5 4 GB RAM system.

3. Results and Discussion

A. Self-curated Kannada dataset

We have grounds-up self-curated an Kannada dataset which consists of 200 self-captured scene images with a few of them picked from the south Indian language dataset [10]. Different camera makes and models were used like the Canon EOS 500D, KENOX S760, Lumia 5MP, and 6.7 MP with focal.

Table 1: Comparison Between Various Methods

South Indian language and Kannada dataset	**Gabor Filter and k-means [10]**	**84**	**93**	**77**
	Transforms and Gabor based [17]	**84**	**98**	**73**
	Discrete Wavelet and MSER [18]	**82**	**64**	**79**
	Proposed Method	**90**	**98**	**84**
MSRA-TD500	**Yao et al. [19]**	**64**	**62**	**61**
	Lu et al. [20]	**29**	**63**	**40**
	Yao et al. [21]	**77**	**75**	**76**
	Zhao et al. [22]	**34**	**69**	**46**
	Yin et al. [23]	**81**	**63**	**71**
	Kang et al. [24]	**71**	**62**	**66**
	Zhou et al. [26]	**87**	**67**	**76**
	Proposed Method	**83**	**90**	**85**
ICDAR 2015 ICDAR 2013 ICDAR 2003	**EAST [25]**	**83**	**78**	**81**
	Holistic [19]	**72**	**59**	**65**
	StradVision2 [26]	**77**	**37**	**50**
	NJU [26]	**70**	**36**	**48**
	CNN MSER [26]	**35**	**34**	**35**
	Proposed Method	**85**	**80**	**83**

length of 3mm-40 mm. The image size ranges from 4KB-16MB, 24-bit depth, 80x40 pixels to 1700x3000 pixels dimensions. It is made available for the research at: Kannada Text in Scene Images

B. Analysis and comparison with existing models

Table 1 highlights the proposed model's performance in comparison to current systems. The findings listed are the best results obtained by those methods/authors on a variety of datasets, including the south Indian language and Kannada data set, the MSRA/TD dataset, and the ICDAR 2003, 2013, and 2015 datasets. On precision (5), recall (6), and F1-score (7), we evaluated the model's performance as follows:

Precision= True positive/True positive + False positive (5)

Recall= True positive/True positive + False negative (6)

F1-score= 2 x (Precision*Recall/Precision + Recall (7)

1) South Indian language and Kannada dataset: The suggested model is compared to three different techniques for word-level text identification. Because it is well-tuned for Kannada text characteristics, the proposed model outperforms all other approaches in accuracy (90 percent), recall (98 percent), and F1-score (84 percent). The suggested model's performance on example pictures from the Kannada dataset is shown in Figure 4.

Fig 4: Sample text detection from kannada images dataset

2)MSRA-TD dataset: The MSRA-TD500 dataset is made up of images with multi-oriented text in English and Chinese that were taken using a pocket camera. The proposed model outperforms in recall (90%) and F1-score (85%) whereas Zhou et al. [26] does better on precision (87% vs. 83% of the proposed model).

3)ICDAR datasets: The ICDAR datasets of 2003, 2013 and 2015 were used. The 2003 dataset consist of 501 images, captured with digital camera, 2013 dataset consists of around 460 scene images and the 2015 dataset includes 1500 images taken in angles using Google Glass, with low resolution and blur images. Though the proposed method addresses blur images, the high inclination of text reduces its performance. But it still outperforms other existing models. Fig. 5 and 6 show sample cases of successful text detection for scene images and born digital images accordingly.

Fig 5: Sample text detection from ICDAR 2013 scene images

Fig 6: Sample text detection from ICDAR 2013 born-digital images

4. Conclusion

Proposed is a novel method for detecting and locating text in natural scene images. The method is fine-tuned particularly for Kannada characteristics using a self-curated Kannada dataset. The algorithm detects every character in every text, and because it yields greater recall rates across many datasets, it may be inferred that the suggested approach is resilient and produces the most relevant

findings. The model allows these judgments to be made based on the text characteristics and the complexity of the content in the image since the criterion for rejecting objects at the filtering stage determines the detection progress. As a result, the suggested model can be concluded to be a reliable and promising tool for detecting text of any size, grayscale values, and complex backdrops. The suggested technique can be used to carry out image pre-processing for deep learning models as well as text recognition.

References:

[1] C. Yao, X. Bai, W. Liu, Y, Ma and Z. Tu, "Detecting texts of arbitrary orientations in natural images," IEEE Conf. on CVPR, vol. 8, pp. 1083–1090, 2012.

[2] B. Epshtein, E. Ofek and Y. Wexler, "Detecting text in natural scene with stroke width transform," in: IEEE Conf. on CVPR, pp. 2963–2970, 2010.

[3] L. Neumann and J. Matas, "A method for text localization and recognition in real-world images," in Kimmel R(ed) Computer Vision, ACCV. Springer, pp. 770–783, 2011.

[4] L. Neumann and J. Matas, "Text localization in real-world images using efficiently pruned exhaustive search," in Proc. 9th ICDAR, pp. 687–691, 2011.

[5] A. Zhu, G. Wang and Y. Dong, "Detecting natural scenes text via auto image partition, two-stage grouping and two-layer classification," Pattern Recognit Lett, vol. 67, pp.153–162, 2015.

[6] G. Aghajari, J. Shanbehzadeh and A. Sarrafzadeh, "A text localization algorithm in color image via new projection profile," in Proc. IMECS 2010 IAENG, vol. 2, pp. 1486-1489, 2010.

[7] C. Shi, C. Wang, B. Xiao, Y. Zhang and S. Gao, "Scene text detection using graph model built upon maximally stable extremal regions," Pattern Recognit Lett , vol. 34, no. 2, pp.107–116, 2013.

[8] R. Wang, N. Sang and C. Gao, "Text detection approach based on confidence map and context information," Neurocomputing, vol. 157, pp. 153–165, 2015.

[9] R. Aravindhan, G. N. Rathna and V. Gupta, "An Online Kannada Text Transliteration System for Mobile Phones," in Proc. International Conference on Communications and Mobile Computing IEEE, vol. 1, pp. 376-381, 2010.

[10] V. N. M. Aradhya, M. S. Pavithra and C. Naveena, "A robust multilingual text detection approach based on transforms and wavelet entropy," Procedia Technology, pp. 232-237, 2012.

[11] S. K. Swee, L. C. Chen and T. S. Ching, "Contrast Enhancement in Endoscopic Images Using Fusion Exposure Histogram Equalization," EL, vol. 28, no. 3, pp.715-723, 2020.

[12] L. I. Zhulin, H. U. JianXue and D. X. H. E. Jing Wang, "Wide Baseline Stereo Matching Algorithm Based on Advanced Extremal Regions," Proc. WCECS 2013 IAENG, vol. 1, pp. 531-535, 2013.

[13] S. Siddiqua, C. Naveena and S. Manvi, "A Combined Edge and Connected Component Based Approach for Kannada Text Detection in Images," ICRAECT -17, pp. 121-125, 2017.

[14] S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong and R. Young, "ICDAR 2003 robust reading competitions," in Proc. 7th ICDAR, pp. 682-687, 2003.

[15] D. Karatzas, F. Shafait, S. Uchida, M. Iwamura and L. Gomez, "ICDAR2013 robust reading competition," in Proc. 12th ICDAR, pp. 1484–149, 2013.

[16] P. Shivakumara, R. P. Sreedhar, T. Q. Phan, S. Lu and C. L. Tan, "Multi oriented video scene text detection through Bayesian classification and boundary growing," IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 8, pp. 1227–1235, 2012.

[17] V. N. M. Aradhya and M. S. Pavithra, "A comprehensive of transforms, Gabor filter and k-means clustering for text detection in images and video," Applied Computing and Informatics, vol 12, no. 2, pp. 109-116, 2016.

[18] B. N. Ajay and C. Naveena, "A mechanism for detection of text in images using DWT and MSER," in Integrated Intelligent Computing, Communication and Security, Springer, pp. 669-676, 2019.

[19] C. Yao, X. Bai, N. Sang, X. Zhou, S. Zhou and Z. Cao, "Scene text detection via holistic, multi-channel prediction," 2016. arXiv preprint arXiv: 1606.09002.

[20] C. Lu, C. Wang and R. Dai, "Text detection in images based on unsupervised classification of edge-based features," 8th ICDAR IEEE, pp. 610-614, 2005.

[21] C. Yao , X. Bai and W. Liu, "A Unified framework for multi-oriented text detection and recognition," IEEE Transactions on Image Processing, vol. 23, no. 11, pp. 4737-4749, 2014.

[22] X. Zhao , K. H. Lin , Y. Fu , Y. Hu , Y. Liu and T. S. Huang, "Text from corners: a novel approach to detect text and caption in videos." IEEE Trans. Image Process., vol. 20, no. 3, pp. 790-799, 2010.

[23] X. C. Yin ,W. Y. Pei, J. Zuang and H. W. Hao, "Multi-orientation scene text detection with adaptive clustering," IEEE Trans. Pattern Anal. Mach. Intell, vol. 37, no. 9, pp.1930-1937, 2015.

[24] L. Kang , Y. Li and D. Doermann, "Orientation robust text line detection in natural images," in Proc. IEEE Conf. on CVPR, pp. 4034-4041, 2014.

[25] X. Zhou, C. Yao, H. Wen, Y. Wang, S. Zhou, W. He and J. Liang, "East: an efficient and accurate scene text detector," in Proc. IEEE Conf. on CVPR, pp. 5551-5560, 2017.

[26] D. Karatzas, L. Gomez-Bigorda, A. Nicolaou, S. Ghosh, A. Bagdanov, M. Iwamura, J. Matas, L. Neumann, V. R. Chandrasekhar, S. Lu, F. Shafait, S. Uchida, and E. Valveny, in Proc. 13th ICDAR IEEE , pp. 1156-1160, 2015.

Shahzia Siddiqua was born in Bengaluru in 1979. She received a B. E. in computer science and engineering from Bangalore University, Bengaluru, Karnataka, 2001, M. Tech, in computer science and engineering from Visvesvaraya Technological University, Belagavi, Karnataka, India, 2010.

She is currently engaged in Ph. D. degree in computer science and engineering from REVA University, Bengaluru, Karnataka, India. She has a teaching involvement of 18 years in Computer Science and Engineering. Her interest are computer graphics, digital image processing, linear algebra, neural networks, pattern recognition, big data and data science.

Naveena C was born in Magadi (T) in 1984. He received his B.E. in computer science and engineering from Visvesvaraya Technological University, Belagavi, Karnataka, India, 2005. M.Tech in network engineering from Visvesvaraya Technological University, Belagavi, Karnataka, India, 2008 and Ph.D. degrees in computer science and engineering from Visvesvaraya Technological University, Belagavi, Karnataka, India,2014.

Presently, he is a Professor at CSE Department, SJBIT, Bengaluru. He has 12 years of teaching experience and 11 years of research exposure. He has published International Journals, Proceedings and Book Chapter in Image Processing and Pattern Recognition. His curiosities are document image analysis, OCR, Linear Algebra, Image Processing, and Pattern Recognition.

Sunilkumar S. Manvi was born in Bijapur in 1966. He received his B. E. Degree in electronics and engineering from Karnataka University Bijapur, Karnataka, India, 1987. M. E. degree in electronics from the University of Visweshwariah College of Engineering, Bengaluru, Karnataka, India, 1993 and a Ph. D. degree in electrical communication engineering, Indian Institute of Science, Bengaluru, Karnataka, India 2003.

Presently, he is working at REVA University Bengaluru as Director of School of Computing and IT. He has vast experience of more than 32 years in teaching in Electronics / Computer Science and Engineering. He has published around 240 papers in national and international conferences, 150 papers in national and international journals, and 15 publications/ books/ book Chapters. His research interests are: Software Agent based Network Management, Wireless Networks, Multimedia Networking, Underwater Networks, Wireless Network Security, Grid and Cloud computing, E-commerce, and Mobile computing.

Dr. Manvi is Fellow IETE (FIETE, India), Fellow IE (FIE, India), and senior member of IEEE (SMIEEE, USA). He received best research publication award from VGST Karnataka in 2014. He has been awarded with 'Sathish Dhawan Young Engineers State Award for Year 2015' for outstanding contribution in the field of Engineering Sciences by DST Karnataka. He is a reviewer and programme committee member for many journals (published by IEEE, Springer.