Neighborhood Structure-Based Model for Multilingual Arbitrarily-Oriented Text Localization in Images/Videos

H. T. Basavaraju[1]*, V.N. Manjunath Aradhya[1], D.S. Guru[2]

[1] Department of Computer Applications, JSS Science and Technology University (SJCE), Mysuru, Karnataka (India)
[2] Department of Studies in Computer Science, University of Mysore, Mysuru, Karnataka (India)

Received 14 April 2020 | Accepted 4 April 2021 | Published 4 May 2021

Abstract

The text matter in an image or a video provides more important clue and semantic information of the particular event in the actual situation. Text localization task stands an interesting and challenging research-oriented process in the zone of image processing due to irregular alignments, brightness, degradation, and complex-background. The multilingual textual information has different types of geometrical shapes and it makes further complex to locate the text information. In this work, an effective model is presented to locate the multilingual arbitrary oriented text. The proposed method developed a neighborhood structure model to locate the text region. Initially, the maxmin cluster is applied along with 3X3 sliding window to sharpen the text region. The neighborhood structure creates the boundary for every component using normal deviation calculated from the sharpened image. Finally, the double stroke structure model is employed to locate the accurate text region. The presented model is analyzed on five standard datasets such as NUS, arbitrarily oriented text, Hua's, MRRC and real-time video dataset with performance metrics such as recall, precision, and f-measure.

Keywords

Multilingual Text, Arbitrarily-Oriented, Maxmin Cluster, Neighborhood Structure, Double Stroke Structure.

DOI: 10.9781/ijimai.2021.05.003

I. Introduction

An image or a frame without textual information is tough to realize the precised situation in the real-life environment. Therefore, the text information present in a picture or a frame delivers the broad range of evidence of an incident. A person can easily realize the text information in a picture or a video. But, the computer could not understand the text as alike human. Hence, computer researchers follow the text extracting processes such as detection, localization, segmentation, and recognition. The text localization step plays a substantial role in the text understanding process for video indexing, video retrieval, video tracking, and video understanding. The main challenges encountered during the localization process are the complex background, shearing, low-resolution, illumination, embossed, night vision, alignment, variation in font size, color and style. This paper presents the neighborhood structure-based model to identify the text edges. The maxmin cluster is applied to discriminate the textual pixels from the unwanted pixels. Normal deviation helps to distinct the textual edges from other object edges in the neighborhood structure concept. Finally, the double stroke structure model is employed to identify the location of the real textual region.

II. Related Works

Ample of methods have been developed for text localization in horizontal directions, but a small quantity of approaches have been suggested to locate the multi-oriented textual information, and very less amount of models have been developed to locate the multilingual textual content [1]. Aradhya and Pavithra [2], [3], [4] worked on two-dimensional wavelet decomposition operation to identify the text information. Unar et al., [5] presented MSER, Sobel, and Canny edge detection algorithms to extract the text area. Dutta et al. [6] identify the actual scene text using entropy-based properties and histogram of oriented gradients. Jiang et al. [7] used an improved stroke feature transform, MSER and frequency tuned visual saliency to locate the textual information in natural scene images. Shivakumara et al. [8] worked on gradient and color properties to identify the location of the text region. He et al. [9] introduced a cascaded convolution text network (CCTN) to the coarse-to-fine text localization process. Gabor, wavelet, and k-means algorithms were employed by Pavithra and Aradhya [10], [11] to locate the actual textual area. Laplacian of Gaussian and full connected component concept is applied by Basavaraju et al. [12] to locate the textual content. Shekar et al. [13] proposed a text localization method using DWT and gradient difference model to obtain true text contents. Neumann and Matas [14] developed a text localization model using the sliding window concept, connected component, strokes and gradient concepts to locate the actual text blocks. A probabilistic model is introduced by Basavaraju et al. [15] to identify the actual text contents using hidden Markov random field, E-M algorithm. Xue et al. [16] presented a model to locate the textual contents in low-resolution pictures and videos using gradient values, low pass filter, and Bhattacharyya distance. Liu et al. [17] implemented a CTD concept to identify the curved text information using a transverse and longitudinal offset connection model. Li et

* Corresponding author.
E-mail address: basavaraju.com@gmail.com

al. [18] introduced a progressive scale expansion network (PSENet) to locate the text information. Xie et al. [19] introduced a supervised pyramid context network (SPCNET) to extract the region of the actual text space. Satwashil and Pawar [20] developed a hybrid technique to isolate the textual information from the scene pictures using character descriptor properties and SVM classifiers. Busta et al. [21] applied a trainable convolution neural network to locate the scene text region. Wu et al. [22] developed a strip-based text detection network (STDN) and a region proposal network to identify the location of the text region. Panda et al. [23] tuned the parameters of the MSER model for localizing the multilingual text present in scene images. The maximally stable extremal region (MSER) parameters are manually tuned to analyze the image dimension, text size and text region area. Villamizar et al. [24] developed a multi-scale fully convolutional and sequential network to segment the textual information. U shape network (UNet) identifies the text features from the several resolution images. Finally, the semantic text segmentation network helps to refine the semantic textual information. Zhang et al. [25] developed a fast dense residual network using fast residual dense blocks to recognize the character. With this literature knowledge, the present chapter discussed the segmentation process by applying a level set model with Gaussian mixture model and recognition process by implementing VGG-16 neural network. Ghoshal and Banerjee [26] implemented a model for segmenting and recognizing the character in scene images using canny edge technique, multi-layer perceptron and SVM.

Most of models were developed to identify the multilingual text based on certain combinations of features, but there is no a generalized model to identify the multilingual text. Hence, this paper introduces a neighborhood structure-based generalized model to locate the region of arbitrarily-oriented multilingual scene and graphical text present in images or videos.

III. Proposed Methodology

The presented method introduced the neighborhood pixel variance model to identify the location of the text candidates in images or videos. Primarily, the specified color image or video sequence (frame) is divided into R, G and B components. Later, the maxmin cluster concept is applied to enhance the textual space using three color values. Again the maxmin cluster is applied along with a 3X3 sliding window concept on the enhanced frame to sharpen the text edges. A flexible threshold is considered by calculating the normal deviation from the sharpened image. The neighborhood pixel variance model aids to produce the border lines for text instance of a picture or a frame based on normal deviation. Finally, the double stroke structure model is employed to determine the exact text space. Fig. 1 depicts the graphical representation of the presented algorithm.

A. Clustering of Color Information for Text Edge Enrichment

The maxmin cluster is applied to group the color pixel values for text region enhancement by suppressing the non-text region. The maxmin cluster extracts three color components such as R, G, and B bands from the specified color image or frame. For each RGB pixel, the maximum, minimum and middle intensity levels are separated. If a middle intensity level is close to a maximum intensity level then the respective pixel belongs to max-cluster. Similarly, if the middle intensity level is close to a minimum intensity level then that pixel is determined as min-cluster. Finally, the maxmin cluster model helps to enhance the input image or video frame. Further, the distance between the text space and non-textual space is increased by applying again the maxmin cluster concept along with the 3X3 sliding window on the enhanced frame. The maxmin cluster with the 3X3 sliding window concept results in the sharpened image is as shown in Fig. 3b.

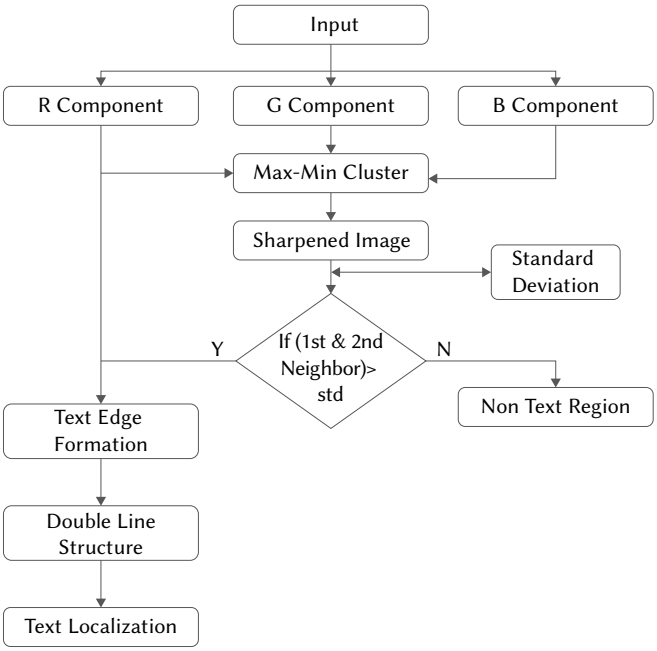

Fig. 1. Graphical representation of the proposed model.

B. Normal Deviation for Text Localization

Normal deviation (nd) is a geometric dimension and it can also be termed as standard deviation, which aids to evaluate the scattering of information based on mean value. The probable error is estimated by calculating the dissimilarity among each and every intensity levels. Therefore, the normal deviation is applied to compute the dissimilarity of every pixel in the sharpened frame. Primarily, the normal deviation calculates the mean score and then it computes the square root of deviation for all intensity levels from its mean value. The minimum normal deviation value denotes that entire intensity levels are closely spaced. The maximum normal deviation value denotes that entire intensity levels are distributed in a wide range. With this idea, the normal deviation(nd) is calculated from the sharpened image to identify the actual text edges. The calculation of normal deviation is represented in Equation (1).

$$nd = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \overline{X})}{N-1}} \quad (1)$$

Where:

nd: Normal deviation.

N: The Size of data points.

X_i: Represent each pixel value.

$\overline{X}$: Depicts the mean value of all pixel values.

C. Neighborhood Pixel Variance Method

The neighborhood structure model helps us to originate the edge information for the textual region in the image or a frame. The presented model emphasizes the borderlines and suppresses the constant section. In the picture or frame sequences, an intensity level delivers substantial information about the space. If the pixel value stands nearly constant in the environment, then respective pixel is considered as area pixels. If the pixel levels have sudden variation, then corresponding pixel is determined as edge pixels. In the actual life situation, maximum of the textual candidates are terminated with its individual borderline as relate to non-textual space. Therefore, the neighborhood structure model has presented to obtain these edge intensity levels in the image or video sequences.

The possible textual region is obtained by differencing the pivot pixel from the 1st and 2nd neighborhood intensity levels. Initially, flexible threshold is measured by computing the normal deviation for the sharpened image. Later, the R component of the given input is considered to extract the potential text edges. For each pivot pixel in the R component, the first neighbor consists of eight picture elements and the second neighbor involves sixteen picture elements. The pivot picture element is differenced with every eight and sixteen picture elements. If the subtracted value is bigger than the flexible threshold, then the respective picture element is determined as a prominent text picture element, else the respective picture element is determined as a non-text picture element. Finally, the neighborhood structure model produces the edge information for text components in the specified picture or frame. Equation 2 is the representation of the potential text pixel extraction process. In the equation 2, f(x,y) represents the pivot pixel, if the difference between 1st neighborhood pixels, 2nd neighborhood pixels, and pivot pixel is larger than nd (normal deviation), then respective f(x,y) intensity level is replaced by 1, otherwise it is allocated by 0. This process continues for each and every intensity level to yield the edges for text candidates. The 1st and 2nd neighborhoods of a pivot picture element are illustrated in Fig. 2.

1	2	3
4	0	5
6	7	8

(a)

1	2	3	4	5
6				7
8		0		9
10				11
12	13	14	15	16

(b)

Fig. 2. (a) First neighborhood, (b) Second neighborhood.

$$f(x,y) = \begin{cases} if\ [f(x,y) - \{\, I\&\ II\ Neighbors\,\} > \ nd], & 1 \\ if\ [f(x,y) - \{\, I\ \&\ II\ Neighbors\,\} < \ nd], & 0 \end{cases} \quad (2)$$

Where:

nd: Normal deviation.

x and y: Space related coordinates.

a. Input Image b. Gray Level c. Possible Text

Fig. 3. Extraction of possible text component process.

D. True Text Candidates

In real-life situation, the textual area seems with circular arbitrarily-oriented shape is depicted in Fig. 3c. According to the organization of textual components, it understands that the actual textual components can be obtained from the outcome of the neighborhood structure model. Therefore, the internal space of circular arbitrarily oriented region is filled by performing the morphological operation. In this specific direction, the specified RGB picture or frame is enhanced by employing the maxmin cluster concept and sharpened by the 3X3 sliding window. Consequently, the latent text information is identified by employing the innovative pixel variance model. Then true text components are obtained by applying the double-line structure approach [8]. If the initial and terminating points of the component are similar, then it is termed a double stroke structure model or arbitrary oriented circular shape. The neighborhood structure model draws the double stroke structure model of text parts. The internal holes of the arbitrary oriented shapes are occupied by applying morphological operation. Lastly, actual text components are recognized by taking the difference between the resultant of the neighborhood structure model represented in Fig. 3c and the filled region shown in Fig. 4a. Fig. 4 presents an extraction of the actual text components using the neighborhood structure model. Fig. 4c signifies the localized textual regions of the actual text components (i.e., Fig. 4b).

a. Holes Filled b. True Text Region c. Final Output

Fig. 4. Extraction of true text region and final output.

IV. Experimental Analysis

The presented model is verified on benchmark datasets like NUS dataset, Hua's dataset [27], the arbitrarily-directional dataset referred in [28], MRRC [29] and ICDAR 2013 video dataset. These datasets enfold all kinds of problems with multiple languages. The effectiveness of the presented model is calculated based on text lines. Precision (Eq. (4)), recall (Eq. (3)) and f-measure (Eq. (5)) evaluating factors are calculated to determine the efficiency of a presented method. The evaluating factors are computed by the following parameters. Real Textual Block (RTB) exhibits total amount of textual blocks in a picture or frame. Truly Located textual Block (TLB) depicts the textual space identified by the presented model. Fallaciously Located textual Block (FLB) refers non-textual space located by the presented model. The presented model employed the neighborhood pixel variance model using eight and sixteen neighbor pixels to generate the important textual spaces. The double-line structure helps us to identify the actual textual space from the outcome of the neighborhood pixel variance concept. The subsequent subsections depict the experimental analysis of five different datasets.

$$Recall(R) = \frac{TDB}{RTB} \quad (3)$$

$$Precision(P) = \frac{TDB}{TDB + FDB} \quad (4)$$

$$F - Measure(F) = \frac{2RP}{R + P} \quad (5)$$

A. Experimental Analysis on NUS Dataset

The NUS database contains 62 pictures. This database consists of straight textual line and curved textual line pictures with composite scenes and lighting effects. The presented model separates the text space from the non-text space. The outcome of the presented model on the NUS database is demonstrated in Fig. 5. Table I shows the qualified analysis of the presented model with a formerly existing method.

Fig. 5. Inputs and equivalent outcomes of NUS dataset.

TABLE I. Qualified Analysis of the Presented Model with an Earlier Existing Model on NUS Dataset

Methods	R	P	F
Shivakumara et al. [30]	85	84	82
Proposed method	91.77	81.32	85.14

B. Experimental Analysis on Hua's Dataset

Hua's database contains 45 pictures with straight textual lines. This database is gathered from sports and news programs. The proposed model has efficiently localized the straight line text region along with fewer false alarms in contrast variation. Fig. 6 represents the example outcomes of the presented model. Table II concludes that the presented model outperforms in recall and f-measure.

TABLE II. Qualified Analysis of the Presented Model with Earlier Existing Models on Hua's Data

Methods	R	P	F
Zhou et al. [31]	72	82	77
Wong and Chen [32]	51	75	61
Sharma et al. [33]	88	77	82
Fourier-RGB [34]	81	73	76
Lu et al. [35]	75	54	63
Bayesian [36]	87	85	85
Proposed method	91.85	80.17	85.64

Fig. 6. Inputs and equivalent outcomes of Hua's dataset.

C. Experimental Analysis on Arbitrariness-Oriented Dataset

An arbitrariness directional database contains 142 pictures with blended textual lines, low contrast, composite background, and lighting effects. The presented model effectively localizes the text space along with few false alarms in composite backgrounds. Fig. 7 depicts the example outcomes of the presented model. Table III represents the quantified outcome of the presented model. The extracted outcome concludes that the presented model outperforms in recall and f-measure.

Fig. 7. Inputs and equivalent outcomes of arbitrariness directional dataset.

TABLE III. Qualified Analysis of the Presented Model with Earlier Existing Models on an Arbitrariness Directional Dataset

Methods	R	P	F
Zhou et al. [31]	41	60	48
Wong and Chen [32]	34	90	49
Sharma et al. [33]	73	88	79
Fourier-RGB [34]	52	68	58
Lu et al. [35]	47	54	50
Bayesian [36]	59	52	55
Proposed method	85.25	81.39	80.21

D. Experimental Analysis on MRRC-334 Dataset

MRRC-334 expands that Multi-script Robust Reading Competition. This database contains 167 training frames and 167 testing frames. A diversity of textual localization issues are enfolded in this database. The main issues are blended, lighting effect, artistic, night visualization, obstruction, scratch, shiny and deepness with multiple languages. The presented model effectively locates the textual space with less false positives. Fig. 8 and 9 depict an outcome of the presented model. Table IV represents the testing outcomes of the presented model on the MRRC database. The extracted outcome overtakes all the parameters.

Fig. 8. Inputs and equivalent outcomes of MRRC training dataset.

TABLE IV. Qualified Analysis of the Presented Model with Earlier Existing Model on MRRC Dataset

Methods	R	P	F
Yin et al. [37]	64	42	51
Proposed method	75.37	74.85	71.93

Fig. 9. Inputs and equivalent outcomes of MRRC testing dataset.

E. Experimental Results on ICDAR 2013 Video Dataset

The presented model has also conducted experimentation on real-time videos. This video dataset is collected from ICDAR 2013 dataset. This dataset contains the scene text along with unwanted noise. With this dataset is difficult to locate the actual text blocks due to low-resolution, distortion, occlusion, illumination effect and complex background. The proposed approach is employed on all video frames. The proposed model conducts the experimentation on the subset of frame sequences to compute the efficiency. The presented model effectively and efficiently identifies the text blocks from video with recall 79.34, precision 71.63 and f-measure 75.28. Fig. 10 shows the example outcomes of the presented model on real-time videos.

The lower link shows the investigational results of video dataset:

https://drive.google.com/drive/folders/1Oa_BnddyLiyaiacZ_1gURbbraOIXXU5G

V. Conclusion

The proposed method developed an effective approach to locate the arbitrariness directional multilingual textual information in pictures or videos. This model is a broad-spectrum analyzed approach to identify the textual space. The maxmin cluster efficiently groups the color information to enhance the given frame. The sliding window concept increases the distance among the text space and the non-text space. The neighborhood pixel variance concept successfully locates the probable textual spaces and a double-line structure or closed arbitrary oriented circular shape effectively identifies the location of the actual textual space. The presented model conducts the evaluation on five standard datasets including video datasets by allowing all types of deviations. In future work, the current research work needs to be improved for the text segmentation process.

Fig. 10. Inputs and equivalent outcomes of video frames.

References

[1] V. N. Manjunath Aradhya, H. T. Basavaraju, and D. S. Guru, "Decade research on text detection in images/videos: a review," Evolutionary Intelligence, 2019, pp. 1-27, https://doi.org/10.1007/s12065-019-00248-z.

[2] V.N. M. Aradhya, and M. S. Pavithra, "An application of LBF energy in image/video frame text detection," 14th international conference on frontiers in handwriting recognition, 2014, pp.760–765.

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

[4] V. N. M. Aradhya, M. S. Pavithra, and S. K. Niranjan, "An exploration of wavelet transform and level set method for text detection in images and video frames," In Recent advances in intelligent informatics, 2014, pp. 419-426.

[5] S. Unar, A. H. Jalbani, M. M. Jawaid, M. Shaikh, and A. A. Chandio, "Artificial Urdu text detection and localization from individual video frames," Mehran University research journal of engineering and technology, vol. 37, no. 2, 2018, pp. 429–438.

[6] K. Dutta, N. Das, M. Kundu, and M. Nasipuri, "Text localization in natural scene images using extreme learning machine," In second international conference on advanced computational and communication paradigms (ICACCP), 2019, pp.1–6.

[7] M. Jiang, J. Cheng, M. Chen, and X. Ku, "An improved text localization method for natural scene images," In journal of Physics: conference series, Vol. 960, No. 1, 2018, pp. 012027.

[8] P. Shivakumara, D. S. Guru, and H. T. Basavaraju, "Color and gradient features for text segmentation from video frames," International conference on multimedia processing, communication and computing applications, 2013, pp.267–278.

[9] T. He, W. Huang, Y. Qiao, and J. Yao, "Accurate text localization in natural image with cascaded convolutional text network," arXiv preprint arXiv:1603.09423,2016.

[10] M. S. Pavithra, and V. N. M. Aradhya, "A comprehensive of transforms, Gabor filter and k-means clustering for text detection in images and video," Applied computing and informatics, 2014, pp. 1–15.

[11] V. N. M. Aradhya, and M. S. Pavithra, "An application of k-means clustering for improving video text detection," In intelligent informatics, 2013, pp.41-47.

[12] H. T. Basavaraju, V. N. M. Aradhya, and D. S. Guru, "A novel arbitrary-oriented multilingual text detection in images/video," In information and decision sciences, 2018, pp. 519–529.

[13] B. H. Shekar, M. L. Smitha, and P. Shivakumara, "Discrete wavelet transform and gradient difference based approach for text localization in videos," In fifth international conference on signal and image processing, 2014, pp. 280–284.

[14] L. Neumann, and J. Matas, "Scene text localization and recognition with oriented stroke detection," In Proceedings of the IEEE international conference on computer vision, 2013, pp. 97–104.

[15] H. T. Basavaraju, V. N. M. Aradhya, and D. S. Guru, "Text detection through hidden Markov random field and EM-algorithm," In information systems design and intelligent applications, 2019, pp.19–29.

[16] M. Xue, P. Shivakumara, C. Zhang, T. Lu, and U. Pal, "Curved text detection in blurred/non-blurred video/scene images," Multimedia tools and applications, 2019, pp. 1–25.

[17] Y. Liu, L. Jin, S. Zhang, C. Luo, and S. Zhang, "Curved scene text detection via transverse and longitudinal sequence connection," Pattern Recognition, Vol. 90, 2019, pp. 337–345.

[18] X. Li, W. Wang, W. Hou, R. Z. Liu, T. Lu, and J. Yang, "Shape robust text detection with progressive scale expansion network," arXiv preprint arXiv:1806.02559, 2018.

[19] E. Xie, Y. Zang, S. Shao, G. Yu, C. Yao, and G. Li, "Scene text detection with supervised pyramid context network," In proceedings of the AAAI conference on artificial intelligence, Vol. 33, 2019, pp. 9038–9045.

[20] K. S. Satwashil, and V. R. Pawar "English text localization and recognition from natural scene image," In international conference on intelligent computing and control systems (ICICCS), 2017, pp. 555–559.

[21] M. Busta, L. Neumann, and J. Matas, "Deep text spotter: An end-to-end trainable scene text localization and recognition framework," In proceedings of the IEEE international conference on computer vision, 2017, pp. 2204–2212.

[22] D. Wu, R. Wang, P. Dai, Y. Zhang, and X. Cao, "Deep strip-based network with cascade learning for scene text localization,". In 14th IAPR international conference on document analysis and recognition (ICDAR), Vol. 1, 2017, pp. 826–831.

[23] S. Panda, S. Ash, N. Chakraborty, A. F. Mollah, S. Basu, and R. Sarkar, "Parameter tuning in mser for text localization in multi-lingual camera-captured scene text images," In computational intelligence in pattern recognition. Springer, 2020, pp. 999–1009.

[24] M. Villamizar, O. Can´evet, and J. M. Odobez, "Multi-scale sequential network for semantic text segmentation and localization," in Pattern Recognition Letters, Vol. 129, Elsevier, 2020, pp. 63–69.

[25] Z. Zhang, Z. Tang, Y. Wang, J. Qin, H. Zhang, and S. Yan, "Fast dense residual network: Enhancing global dense feature flow for text recognition," in arXiv preprint arXiv:2001.09021.

[26] R. Ghoshal and A. Banerjee, "Svm and mlp based segmentation and recognition of text from scene images through an effective binarization scheme," In computational intelligence in pattern recognition. Springer, 2020, pp. 237–246.

[27] X. S. Hua, L. Wenyin, and H. J. Zhang, "An automatic performance evaluation protocol for video text detection algorithms," IEEE Trans CSVT, 2004, pp. 498–507.

[28] C. Lu, C. Wang, and R. Dai, "Text detection in images based on unsupervised classification of edge based features", In: Proceedings. ICDAR, 2005, pp. 610–614.

[29] Multi-script robust reading competition. http://mile.ee.iisc.ernet.in/mrrc/index.html.

[30] P. Shivakumara, H. T. Basavaraju, D. S. Guru, and C. L. Tan, "Detection of curved text in video: quadtree based method," In: 12th international conference on document analysis and recognition (ICDAR), 2013, pp.

594–598.
[31] J. Zhou, L. Xu, B. Xiao, and R. Dai, "A robust system for text extraction in video," In: Proceedings of ICMV, 2007, pp. 119–124.
[32] E. K. Wong, and M. Chen, "A new robust algorithm for video text extraction," Pattern Recognition, 2003, pp. 1397–1406.
[33] N Sharma, P. Shivakumara, U. Pal, M Blumenstein, and C. L. Tan, "New method for arbitrarily oriented text detection in video," In: Proceedings of DAS, 2012, pp. 74–78.
[34] P. Shivakumara, T. Q. Phan, and C. L. Tan, "New Fourier-statistical features in RGB space for video text detection," IEEE Transaction on CSVT, 2010, pp. 1520–1532.
[35] C. Lu, C. Wang, and R. Dai, "Text detection in images based on unsupervised classification of edge based features," In: Proceedings of ICDAR, 2005, pp. 610–614.
[36] 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. CSVT, 2012, pp. 1227–235.
[37] X. C. Yin, X. Yin, K. Huang, and H. W. Hao, "Robust text detection in natural scene images", IEEETrans. PAMI 36, 2014, pp. 970–983.

H T Basavaraju

H T Basavaraju received the B.Sc. and MCA degrees in Computer Science from the University of Mysore, Mysuru, Karnataka, India in 2009 and 2012 respectively. He worked as a research officer at AIISH from 2012 to 2013. He received the Ph.D. degree at the Department of Computer Applications, Sri Jayachamarajendra College of Engineering, Visveswaraya Technological University, Karnataka, India. His research interest includes Document image analysis, Computer vision, Speech processing, Machine learning, Deep Learning, Natural language processing and Artificial Intelligence.

V.N. Manjunath Aradhya

Dr. V.N. Manjunath Aradhya is currently working as an Associate Professor & Head in the Dept. of Computer Applications, JSS Science and Technology University, Mysuru. He received the M.S. and Ph.D. degrees in Computer Science from the University of Mysore, Mysuru, India, in 2004 and 2007 respectively. He is a recipient of "Young Indian Research Scientist" from the Italian Ministry of Education, University and Research, Italy during 2009-2010. An awardee of "Young Scientist" from the Department of Science and Technology (DST) in 2009 under FAST TRACK SCHEME. Recently awarded prestigious ARP (Award for Research Publications) by Vision Group on Science and Technology (VGST), Govt. of Karnataka for the year 2016-17. His professional recognition includes as a Technical Editor for Journal of Convergence Information Technology (JCIT), Editor Board in Journal of Intelligent Systems, reviewer for IEEE Trans on System, Man, and Cybernetics - PART B, Pattern Recognition (PR), and Pattern Recognition Letters (PRL). His research interest includes Pattern Recognition and Image Processing, Speech Processing, Document Image Analysis, Computer Vision, Machine Intelligence, Applications of Linear Algebra for the Solution of Engineering Problems, Biclustering of Gene Expression Data and Web Data Analysis and Understanding.

D. S. Guru

Prof. D. S. Guru received his B.Sc, M.Sc and Ph.D. degrees in Computer Science and Technology from the University of Mysore, Mysuru, India in 1991, 1993 and 2000 respectively. He is currently a Professor in the Department of Studies in Computer Science, University of Mysore, India. He was a fellow of BOYSCAST and a visiting research scientist at Michigan State University. He has authored 65 journals and 225 peer-reviewed conference papers at international and national levels. His area of research interest covers image retrieval, text mining, machine learning, object recognition, shape analysis, sign language recognition, biometrics, and symbolic data analysis.