



Saudi Computer Society, King Saud University

Applied Computing and Informatics

(<http://computer.org.sa>)  
[www.ksu.edu.sa](http://www.ksu.edu.sa)  
[www.sciencedirect.com](http://www.sciencedirect.com)



ORIGINAL ARTICLE

# A comprehensive of transforms, Gabor filter and $k$ -means clustering for text detection in images and video



V.N. Manjunath Aradhya <sup>a,\*</sup>, M.S. Pavithra <sup>b</sup>

<sup>a</sup> Dept. of MCA, SJCE, Mysore, India

<sup>b</sup> Dept. of MCA, DSCE, Bangalore, India

Received 8 December 2013; revised 3 August 2014; accepted 6 August 2014  
Available online 23 August 2014

## KEYWORDS

Wavelet transform;  
Multilingual text;  
Wavelet decomposition;  
Gabor filter;  
 $k$ -means clustering;  
Linked list approach;  
Wavelet entropy

**Abstract** The present paper presents one of the efficient approaches toward multilingual text detection for video indexing. In this paper, we propose a method for detecting text located in varying and complex background in images/video. The present approach comprises four stages: In the first stage, combination of wavelet transform and Gabor filter is applied. By applying single level 2D wavelet decomposition with Gabor Filter, the intrinsic features comprising sharpen edges and texture features of an input image are obtained. In the second stage, the resultant Gabor image is classified using  $k$ -means clustering algorithm. In the third stage, morphological operations are performed on clustered pixels. Then a concept of linked list approach is used to build a true textline sequence of connected components. In the final stage, wavelet entropy of an input image is measured by signifying the complexity of unsteady signals corresponding to the position of textline sequence of connected components in leading to determine the true text region of an input image. The performance of the approach is exhibited by presenting promising experimental results for 101 video images, standard ICDAR 2003 Scene Text Test dataset, ICDAR 2013 dataset and on our own collected South Indian Language dataset.

© 2014 King Saud University. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

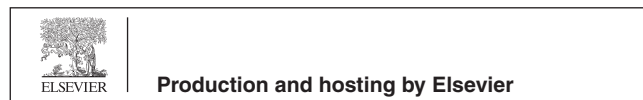
## 1. Introduction

With the current multimedia technology, the captured images and understanding these images through its contents have gained lots of attention from the computer vision community. Contents of an images and video help in clear understanding the information present within. A text is one of the images and video content which carries semantic information, and may help to provide the scene description of an image. Hence, the detection and extraction of either scene or graphics text has

\* Corresponding author.

E-mail addresses: [aradhya.mysore@gmail.com](mailto:aradhya.mysore@gmail.com) (V.N. Manjunath Aradhya), [ngspavithra@gmail.com](mailto:ngspavithra@gmail.com) (M.S. Pavithra).

Peer review under responsibility of King Saud University.



been widely used in content based image indexing and retrieval. The detection of text present in video is being used in video summary and in video sequence retrieval. Text detection is a pre-processing task for text recognition. Nowadays, digital images also carry useful information. Uploading these images to social networking sites is getting increased day by day. So detecting texts of digital images also plays an important and challenging role in image retrieval system. Born-digital images are generated by computer software and are saved as digital images. Born-digital images have complex foreground/background, low resolution and have severe soften edges [1]. So detecting text from such born-digital images is difficult. Texts in natural scene images also have to be robustly detected before being recognized and retrieved. Scene images contain texts such as the advertising boards, name plates, address boards of houses, and landmarks of streets, are captured naturally when the scene images are taken by the camera. Therefore scene text is embedded in the background as a part of the scene. Scene images are complex because the backgrounds are complex containing the text in different sizes, styles and alignments and the resolution of the image is low. Hence detection of text from natural scene images remains still challenging task. Text detection and recognition in images and video frames aim at integrating advanced optical character recognition (OCR) and text based searching technologies [2]. In this regard, the existing text detection methods use text features such as gradient [3], edge [4] and texture [5] information. With such efforts, detection of text in images and video remains a challenging task due to variations of text background, font and orientations.

Recent text detection in natural scene images, born-digital images and video text detection has also been surveyed. Gonzalez and Bergasa [6] described a method to read text in natural images, using geometric and gradient properties. Zeng et al. [7] presented a framework for detecting text from webpage and email images, based on maximum gradient difference values. Hence the recent literature study implies that detection of text from either natural scene or digital images and text detection in video are still in the pace of research. Our proposed algorithm implements  $k$ -means clustering algorithm in detecting true text regions. In this regard, earlier research works of text detection in video and images based on  $k$ -means clustering algorithm are surveyed herewith, Phan et al. [8] developed a text detection approach with the Laplacian operator. Then  $k$ -means is used to classify all the pixels into clusters. Wu et al. [5] described a text localization method based on texture segmentation by computing texture features.  $k$ -means algorithm is applied for classification. Shivakumara et al. [9] describe a method based on the Laplacian in the frequency domain. In this, the input image is filtered with Fourier–Laplacian. Then,  $k$ -means clustering is used to identify candidate text regions.

The above stated studies revealed that the text detection approaches are either region or texture based methods. Though the concept of  $k$ -means clustering algorithm and a connected component analysis have used, the detection accuracy of the text region can still be improved without missing text data and reducing most of the falsely detected blocks of an image. Detection of texts of south Indian language is still a challenging task. Words of such south Indian languages are framed with modifiers and compound bases. To detect such a texts in an images/video, we propose a method based

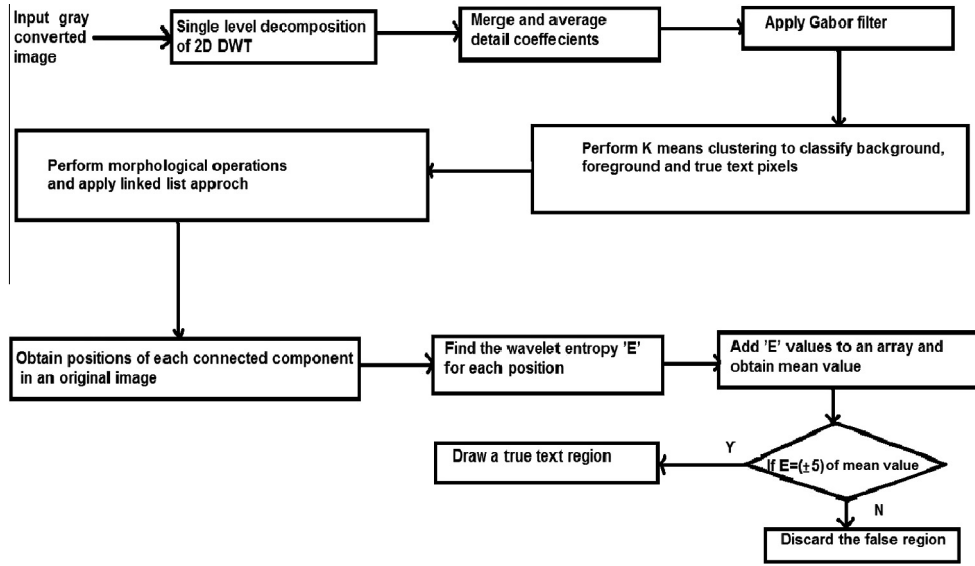
on Transforms, Gabor filter and  $k$ -means clustering. By sustaining the development of our system [10], which describes the text detection method in color and regular images. In the first stage of the system [10], wavelet transform and Gabor filter are applied to extract sharpened edges and textural features of a given input image. In the second stage of the method wavelet entropy is calculated to get an energy value of a resulted Gabor image in order to find the high frequency texture elements of a processed one to determine the true text region of an image. As a progress of our work [11], in the present paper, we propose a multilingual text detection system with the wavelet transform, Gabor filter and  $k$ -means clustering. The proposed method concentrates mainly on detection of English and south Indian language texts in images/video. The system yields better results for various background complexities and texts between other dominant non-text objects. Experiments are carried out on 101 video images dataset of [8], our own collected multilingual language dataset, ICDAR 2003 scene test dataset, ICDAR 2013 dataset and on video frames of Kannada. Comparative studies are reported in detail. The rest of this paper is organized as follows. Section 2 describes our proposed method. Experimental results are presented and performance evaluation on considered datasets is discussed in Section 3. Finally, conclusions are drawn.

## 2. Proposed methodology

The proposed method is a robust multilingual text detection approach based on the sequential adoption of Wavelet transform, Gabor filter,  $k$ -means clustering and a measure of wavelet entropy. First, by applying single-level discrete 2-D wavelet transform, a single-level 2-D wavelet decomposition is performed. As a result approximated and detailed co-efficients are obtained. Then detail co-efficients are merged and averaged to extract efficient texture feature information. Gabor filter is further applied in order to obtain edge information of an image. The resulted Gabor output image is grouped into three clusters by applying the  $k$ -means algorithm to classify the background, foreground and the true text pixels of an image. In the next stage, morphological operations are performed to obtain connected components, then a concept of linked list approach is in turn used to build a true text line sequence of connected components. In the final stage, wavelet entropy is measured in an each connected component sequence in order to determine the true text region of an input image. A complete text detection procedure of the proposed work is shown in Fig. 1 and explained in the following sub-sections.

### 2.1. Discrete wavelet transform for texture feature extraction

The Wavelet Transform is a method convolution of the wavelet function with the signal. The ability of the discrete wavelet transform to decompose a signal at different independent scales and to do it in a very flexible way [12]. The discrete wavelet transform (DWT) is an implementation of the wavelet transform using a discrete set of the wavelet scales and translations. In our research work wavelets are used as analytical tools for signal processing. A study has performed about the development of the discrete wavelet transform (DWT), as a series expansion of signals in terms of wavelets and scaling functions which are associated with low pass and high pass



**Figure 1** Work flow diagram of the proposed text detection system.

filters respectively. With DWT a signal is decomposed into a coarse approximation and detail information. One of the beneficial application of the DWT is that, it performs texture classification in order to distinguish micro and macro textures. The 2D Discrete Wavelet Transform works as follows: The rows of the 2D signal are passed through a HPF and a LPF, and down-sampled by two. This divides the image into two sub images. Second, all columns of each sub image are filtered using the HPF and LPF and down sampled by two. We experimented on Daubechies family's first order Daubechies wavelet function i.e., db1. The advantage of compact supportness property of the Daubechies wavelets is finite number of filter parameters and fast implementations. Accordingly single level 2-D wavelet decomposition is performed with respect to the db1 Daubechies wavelet function to extract texture feature and edge information of an input image. The obtained result is shown in Fig. 2(b) for an input image Fig. 2(a).

## 2.2. Gabor filter

Gabor filter mainly provides better spatial localization. The main intension of employing Gabor filter is for texture segmentation. In [13], Gabor filters are utilized via multiresolution structure, consisting of filters tuned to several different frequencies and orientation. The main purpose of using Gabor filter after employing wavelet transform are the following:

- The multiresolution structure relates the Gabor features to wavelets by utilizing several frequencies and orientations.
- Has supporting properties for feature extraction.
- To perform localized analysis and to extract a local phenomena.

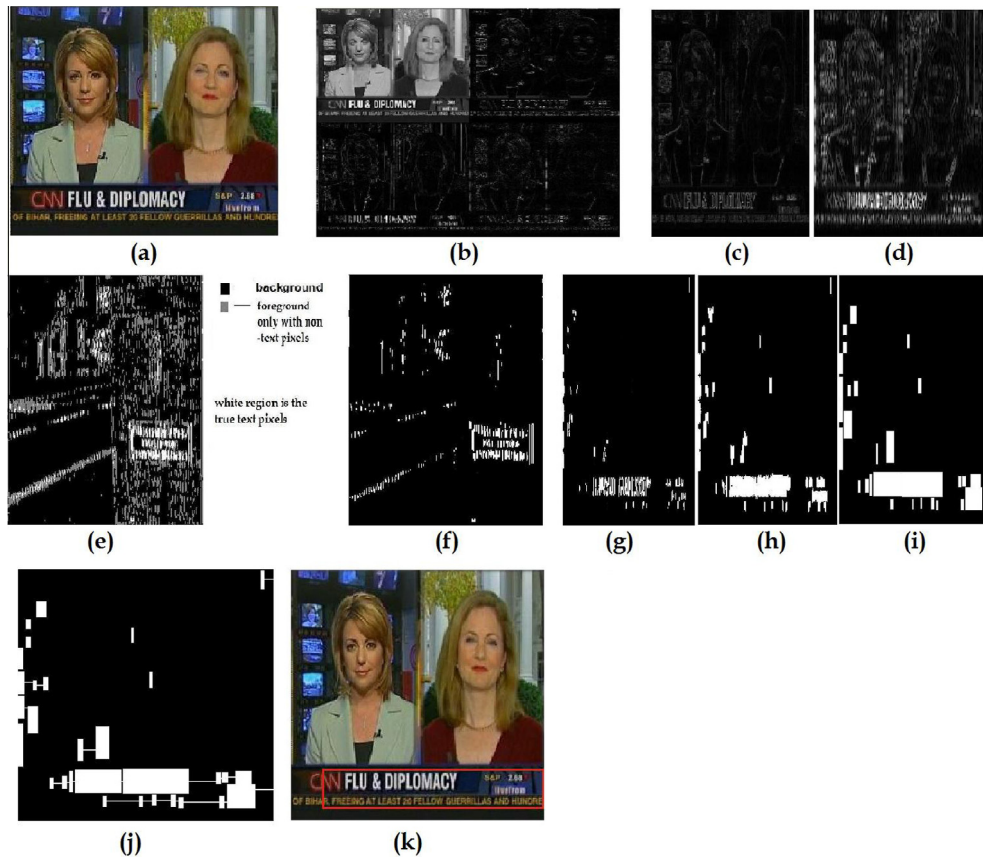
The vital part in the process of employing Gabor filter is filter parameter selection and filter construction. The Gabor filter is basically a Gaussian modulated by a complex sinusoid. The filter is represented as follows:

$$G(x, y, \theta, f) = \exp\left[-\frac{1}{2}\left(\frac{x'}{S_x}\right)^2 + \left(\frac{y'}{S_y}\right)^2\right] * \cos(2 * \pi * f * x') \quad (1)$$

where  $x' = x * \cos(\theta) + y * \sin(\theta)$  and  $y' = y * \cos(\theta) - x * \sin(\theta)$ . Where  $G(x, y, \theta, f)$  be the function defining a Gabor filter, ' $\theta$ ' is the orientation and its value set to  $\pi/2$ , ' $f$ ' is the spatial frequency its value set to  $1/\sqrt{2}$ , ' $S_x$  and  $S_y$ ' are variances along  $x$  and  $y$  axes respectively. Gabor filter representation is optimal and shows better performance for classifying text region. The resultant image of detail co-efficients obtained by performing single-level 2D decomposition is averaged and is shown in Fig. 2(c). Then, the Gabor filter function is applied on to the obtained average image of detail co-efficients, in order to extract local phenomena. The Gabor filter resulted image is shown in Fig. 2(d). The implementation of Gabor filter proves better performance in classifying text region.

## 2.3. Implementation of $k$ -means clustering algorithm

Clustering in [14] is a group of methods for finding and describing cohesive groups in data, such as compact clusters of entities in the feature space. To find and describe a cohesive groups in our data, we applied  $k$ -means clustering. The main objective of applying  $k$ -means is to minimize the total intra-cluster variance as well as to build a classification over empirical data. The procedure of  $k$ -means clustering begins with initiating  $k$  tentative centroids. Further, tasks such as collecting clusters around centroids and updating centroids with cluster means, until its convergence are repeatedly performed.  $k$ -means mainly depends on the initialization of the centroids. In this regard, we initialized  $k$  value as 2 and observed the detected true text region. When we initialize  $k$  value as 3, the obtained true text region was more accurate than when  $k = 2$ . So by initializing  $k$  value as 3, the background, foreground and the true text pixels of an image are well classified which is shown in Fig. 2(e). A cluster with true text pixels is shown in Fig. 2(f).



**Figure 2** (a) Input image of 101 video images dataset. (b) Resulted image of single level 2-D wavelet decomposition. (c) Resultant average image of single level 2-D wavelet decomposition of detail co-efficients. (d) Resulted Gabor image with feature extraction. (e) Resulted image of  $k$ -means clustering by setting  $K = 3$ . (f) An image of third cluster consisting of true text pixels. (g), (h) and (i) Resultant images of morphological operations. (j) A Resultant image obtained for linked list approach. (k) True text region detected image.

#### 2.4. Morphological operations for detection of true text pixels of an image

To the obtained  $k$ -means clustering resultant image, we applied morphological dilation operation with a rectangle structuring element of size  $5 \times 3$  to get connected components of true text pixels. The obtained resultant images are shown in Fig. 2(g), (h) and (i). A concept of linked list approach is then used to build a true text line sequence in order to get a sequence of connected components as in [15] to detect a sequence of true text regions of an input image. A linked list is a concept of linear data structure which accesses the elements sequentially. In a list rowwise elements are arranged from starting to the end of the list. In order to access each element in the list for data processing, a function has to be created which traverses the list. In the proposed system, a linked list concept has selected as the choice of data structure and the point to be noticed that an element of the list refers to a connected component obtained by applying morphological operations. A group of nearby elements of a list i.e., nearby connected components together represent a sequence. To build this sequences of connected components, center point of the right most part of a connected component is calculated using Eq. (3). By finding  $R_{mid}$  point, a horizontal line with respect to the zero degree up to the line reaches to the nearest

connected component is obtained. 'mid' is center point of the connected component and ' $R_{mid}$ ' is the center point of the right most part of the connected component. With this approach, the proposed system is able to maintain elements by position and allows for data manipulation of an each connected component. The result obtained for this stage is shown in Fig. 2(j).

$$\text{mid} = \frac{y_{\max} + y_{\min}}{2} \quad (2)$$

$$R_{\text{mid}} = (x_{\max}, \text{mid}) \quad (3)$$

#### 2.5. Calculation of wavelet entropy

The wavelet entropy carries information about the degree of order, disorder associated with a multi-frequency signal response. From the obtained sequence of connected components, we calculated the wavelet entropy using Eq. (4) of the corresponding region of a sequence of connected components in an input image, in order to obtain the true text region as well to eliminate falsy blocks of an image. Then we extracted an energy information from an input image of the regions specified. Average energy of all the regions specified in the input image is fixed as a threshold  $t$ . If the specified sequence of a text region is  $\pm 5$  of  $t$ , as a result, a region is considered as a text region else it is considered as a non-text region.



Fig. 2(k) shows the text region obtained from the calculation of wavelet entropy.

$$E(s) = \sum_i \log(s_i^2) \quad (4)$$

where  $s$  is the signal and  $(s_i)$  is the coefficients of  $s$  in an orthonormal basis.

### 3. Experimental results and discussion

The proposed method was experimented on intel CORE Duo 2.0 GHz machine with MATLAB R2008a. In our experiment, we used four challenging datasets. The first dataset contains 101 video images of [8], which are extracted from news programmes, sports videos and movie clips. The dataset includes both graphic text and scene text of different languages, e.g. English, Chinese and Korean. The second dataset is the South Indian Language dataset, in which an image comprising either a text of Kannada, Tamil, Telugu and Malayalam languages. The main purpose of considering south Indian languages texts was because, words were framed on the basis of compound bases, modifiers and extra modifiers. Detecting such a text is a challenging task. The third dataset is the mostly cited ICDAR 2003 Scene Trial Test dataset comprising 251 camera-based images and Fourth dataset is the ICDAR 2013 Born-Digital Images (Web and Email) and Scene Images dataset. The performance of the proposed system is evaluated at the block level. The blocks are determined as the categories described in [8].

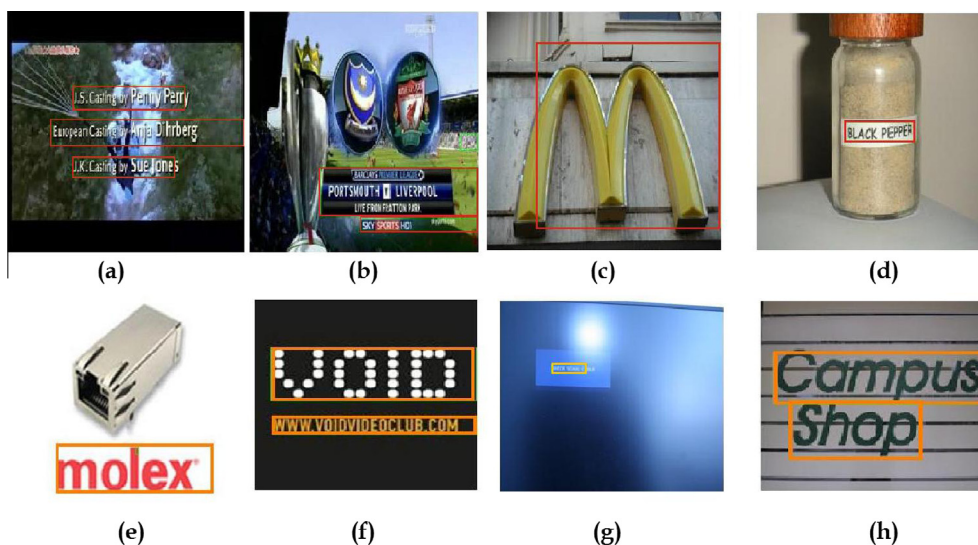
#### 3.1. Experiment on 101 video images dataset

The experiment was conducted on 101 video images provided by [8]. The dataset comprises video images with horizontal text lines with a different font color and varying background. The successful text detection images obtained for a dataset of 101 video images are shown in Fig. 3(a) and (b).

**Table 1** Results obtained for the dataset of 101 video images.

Method	ATB	TDB	FDB	MDB
Edge-based [4]	491	393	86	79
Laplacian [8]	491	458	39	55
Transforms and Gabor based [10]	491	481	78	53
<b>Proposed</b>	491	486	78	15

For each image in the dataset we manually count the Actual Text Blocks (ATB). We compared the proposed method with the existing text detection methods such as Edge-based [4], Laplacian [8] and Transforms and Gabor based [10] methods. In order to evaluate the performance of the proposed method on 101 video images dataset, we use the performance measures defined in [8]. Table 1 shows the comparative study of the proposed and existing methods. From the table, it is clear that the obtained TDBs are more i.e., the system detects more number of true text blocks, FDBs are sustained as in Transforms and Gabor based method [10] indicating that there exists few false alarms. MDBs are considerably reduced, which shows that the miss detection of text blocks is very few in number. We compared the proposed method with existing text detection methods such as Edge-based [4], Laplacian [8] and Transforms and Gabor based [10] methods. From Table 2, it is clear that the proposed method has higher DR and lesser MDR compared to existing methods and FPR is sustained as of Transforms and Gabor based method [10]. The main goal of the proposed system is to achieve highest DR by detecting true text blocks of an image, in this regard we reached DR = 98.9%. By our experiment, it is proved that the proposed method exhibits higher detection rate and considerably lesser miss detection rate than the existing methods. For this dataset, Average Processing Time is 2.9 s. The best execution time is 2.6 s of image size  $320 \times 240$  and the worst execution time is 3.5 s for  $720 \times 576$  image size.



**Figure 3** (a) and (b) Results of true text region detected images 101 video images dataset. (c) and (d) Results of true text detection of ICDAR 2003 Scene Trial Test dataset. (e) and (f) Results of true text detection of ICDAR 2013 Born-Digital Images dataset. (g) and (h) Results of true text detection of ICDAR 2013 Text in Scene Images dataset.

**Table 2** Performance results obtained for 101 video images dataset.

Method	DR	FPR	MDR
Edge-based [4]	80.0	18.0	20.1
Laplacian [8]	93.3	<b>7.9</b>	12.0
Transforms and Gabor based [10]	97.9	13.9	11.0
<b>Proposed</b>	<b>98.9</b>	13.7	<b>3.0</b>

3.2. Experiment on our own collected south Indian languages dataset

Our own collected multilingual dataset is challenging and texts present in varying complex background with different font color and size. In some images, texts are present in between different dominant non-text objects of an image. This particular kind of images leads in resulting false positives. The main applicability of the proposed system is that it detects text of South Indian Languages which may further help to develop south Indian languages OCR. To exhibit the performance of the proposed method with South Indian languages dataset (based on the formation of a word of considered four south Indian languages are described), we have conducted experiment on our own collected South Indian language dataset comprising 114 images with Kannada, Tamil, Telugu and Malayalam texts. Kannada images are collected by capturing scene text as well as front cover of the Kannada text books and novels. Remaining languages text images are collected from magazines and few are images of posters. In this dataset texts present in an image are with different font size, color and varying background. Since Kannada text is framed with modifiers and compound bases, a text is detected with missing of curving portion as well as modifiers and its curves. With all these reasons, performance measures are considerably less compared to English text images. Successful Text detection results of south Indian languages are shown in Fig. 4 (Table 3).

From the experimental results in Table 4, it is noticed that the texts of south Indian languages present in an image are completely detected along with its modifiers and compound bases as a whole word. Our previous method i.e. Transforms and Gabor based [10] detects text region with an existence of false alarms and few missing text data. For the same dataset, the present method detects the region of text with its modifiers and compound bases in an input image. Though the Detection Rate (DR) reduced to low and Miss Detection Rate (MDR) is

**Table 3** Results obtained for our own collected south Indian language dataset.

Method	ATB	TDB	FDB	MDB
Transforms and Gabor based [10]	314	309	116	19
<b>Proposed</b>	314	291	85	28

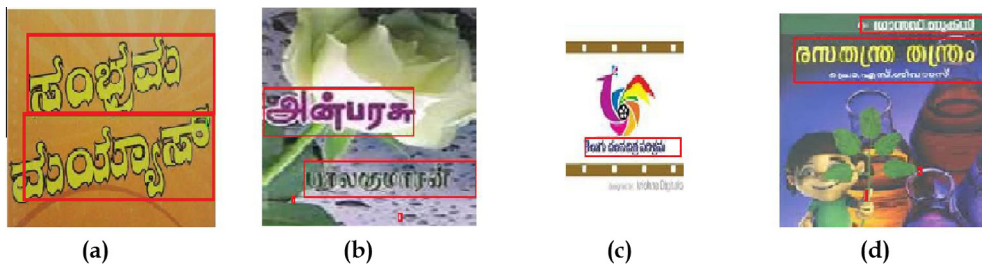
**Table 4** Performance results obtained on our own collected south Indian language dataset.

Method	DR	FPR	MDR	APT (s)
Transforms and Gabor based [10]	<b>98.40</b>	27.29	<b>6.14</b>	7.9
<b>Proposed</b>	92.67	<b>22.60</b>	9.62	<b>2.6</b>

increased compared to the method [10], but one of the noticeable measure i.e. False Positive Rate (FPR) gets reduced. From this, it is proven that the proposed system executes toward finding most of the text elements of an image. Average Processing Time (APT) is also measured, the proposed system yields result much faster in time compared to [10]. The best execution time is 2.6 s for image size 160 × 45 and the worst execution time is 3.0 s for image size 1024 × 500.

3.3. Experiment on ICDAR test datasets

Experiment is conducted on ICDAR 2003 Text Locating dataset and Scene Images dataset and ICDAR 2013 Born-Digital Images (Web and Email) and Scene Images dataset. ICDAR 2003 Scene Trial Test dataset contains 251 camera-based images. During the experiment, for each image in the dataset number of Actual Text Blocks(ATB), Truly detected block (TDB), Falsely Detected Block (FDB) and Text block with Missing Data (MDB) are manually counted. To evaluate the proposed method on ICDAR 2003 Scene Trial Test dataset, we compared the proposed method with three existing methods such as Gradient [3], Laplacian [9] and Bayesian Classification and Boundary Growing [16] methods that are experimented on ICDAR 2003 Scene Trial Test dataset. In addition to the existing categories, Average Processing Time (APT): Processing time per frame required for detecting text in the images defined in [16] is considered. The performance measures defined in [16] Recall (R) = TDB/ATB, Precision



**Figure 4** (a–d) Successful text detection results of south Indian languages dataset of Kannada, Tamil, Telugu and Malayalam texts respectively.

**Table 5** Experimental result on ICDAR 2003 Scene Trial Test dataset.

Method	Recall	Precision	$F$ -Measure	MDR	APT (s)
Gradient [3]	0.52	<b>0.83</b>	0.64	0.08	<b>1.0</b>
Laplacian-Fourier [9]	0.86	0.76	0.81	0.13	6.8
Bayesian Classification and Boundary Growing [16]	0.87	0.72	0.78	0.14	7.9
<b>Proposed</b>	<b>0.92</b>	0.73	<b>0.82</b>	<b>0.06</b>	3.9

**Table 6** Experimental result on ICDAR 2013 Born-Digital Images (%).

Method	Recall	Precision	$F$ -score
USTB-TexStar	<b>82.38</b>	<b>93.83</b>	<b>87.74</b>
TH-TextLoc	75.85	86.82	80.96
I2R-NUS-FUS	71.42	84.17	77.27
<b>Proposed</b> based on word level	75.45	80.11	77.71
<b>Proposed</b> based on textline level	89.00	90.00	89.00

**Table 7** Experimental result on ICDAR 2013 Reading Text in Scene Images dataset (%).

Method	Recall	Precision	$F$ -Score
USTB-TexStar	66.45	<b>88.47</b>	<b>75.89</b>
Text Spotter	64.84	87.51	74.49
CASIA-NLPR	68.24	78.89	73.18
<b>Proposed</b> based on word level	<b>71.09</b>	61.99	66.22
<b>Proposed</b> based on textline level	98.90	80.20	88.50

$(P) = TDB/(TDB + FDB)$ ,  $F$ -measure  $(F) = 2 \times P \times R / (P + R)$ , Misdetction Rate (MDR) =  $MDB/TDB$  are considered to evaluate our experiment. The successful text detection images obtained for ICDAR 2003 Scene Trial Test dataset are shown in Fig. 3(c) and (d). By observing the images with the detected true text blocks, we noticed that the proposed method estimates are most representative. Table 5 shows the experimental results of the proposed method with existing methods. The proposed method achieves the highest recall,  $F$ -measure and lowest MDR. Thus for the dataset ICDAR 2003 Scene Trial Test, the proposed method outperforms all the existing methods in most of the performance measures. Average Processing Time for the dataset is 3.9 s. The best execution time is 3.3 s for image size  $640 \times 480$  and the worst execution time is 4.1 s for image size  $1600 \times 1200$ .

To evaluate the performance of the proposed method for ICDAR 2013 Born-Digital Images (Web and Email) dataset and Reading Text in Scene Images dataset, the computation is performed based on detection of words as per the protocol

of ICDAR 2013 competition. To evaluate the detection of text based on word wise, linked list approach is not employed in the procedure of the proposed system. Performance evaluation metric is followed as in [17]. ICDAR 2013 Born-Digital Images (Web and Email) dataset comprises 141 images. Experiment results are compared with the methods submitted to the challenge: ICDAR 2013 Robust Reading Competition of born-digital images [17]. Experimental results are detailed in Table 6. The result proves that the proposed system is competitive and comparable to other existing methods. For reference purpose, we also provided the text detection results in terms of text line by employing the linked list approach. Successful text detection images of ICDAR 2013 Born-Digital images dataset are shown in Fig. 3(e) and (f). The average processing time for this dataset is 2.6 s, the best execution time is 2.1 s for an image size  $680 \times 106$  and the worst execution time is 2.9 s for an image size  $573 \times 174$ .

We also conducted an experiment on ICDAR 2013 Reading Text in Scene Images dataset. Experimental results are



Figure 5 Text detection results of Kannada news video.



compared with the few of the competitive methods submitted to the challenge: ICDAR 2013 Reading Text in scene images [17]. Successful text detection images results of ICDAR 2013 Text in Scene Images dataset are shown in Fig. 3(g) and (h). Experimental results are shown in Table 7. For reference purpose, in Table 7 we also provided the text detection results in terms of text line by employing the linked list approach. In this case, the average processing time for the dataset is 3.3 s, the best execution time is 2.6 s for an image size  $533 \times 263$  and the worst execution time is 7.0 s for an image size  $3888 \times 2582$ .

#### 3.4. Experiment on video

We also conducted an experiment on our own collected video sequences consisting of English and Kannada texts. The system successfully detects text present in the video. The system detects the true text region in a video with the same bit rate of the input video. This approach plays an important role in the recent progress of text detection technique for video. The resultant text detection in video is shown as video frames in Fig. 5.

#### 4. Conclusion

An efficient method for text detection in images and video is introduced. The combination of wavelet transforms, Gabor filter extracts texture features and sharpen edges of an input image. Grouping the data of an extracted information is well performed using  $k$ -means clustering by considering  $k$  value as 3, resultant of this is classification of background, foreground and the true text pixels. By employing the morphological operations and the concept of linked list approach, connected components of true text line sequences are obtained. Finally, wavelet entropy of an input image is measured corresponding to the positions of true text line sequence of connected components. The most beneficial feature of using Gabor filter with the combination of  $k$ -means clustering is that, the proposed system could able to extract most of the text information, sharpen edge information and able to classify background, foreground and true text pixels of an input image efficiently. On the highly cited ICDAR 2003 Scene Text Test dataset the method surpasses the existing methods in terms of most of the considered performance measures. Performance results on ICDAR 2013 Robust Reading Competition datasets of born-digital images and scene images prove competitive compared with the methods submitted to the Text Localization of ICDAR 2013 Robust Reading Competitions. The proposed system is better suitable for horizontal texts, texts framed with modifiers and compound bases such as texts of south Indian languages present in images/video. Thus, the proposed system can extensively be used to detect a true text region accurately in attaining high detection rate with a very few missing data in numbers. In this regard, further we propose a better text detection system to represent discontinuities along with the edges or curves in the images especially for South Indian language scenario in order to detect a close boundary around the texts.

#### References

- [1] J. Zhang, R. Cheng, K. Wang, H. Zhao, Research on the text detection and extraction from complex images, in: The Proceedings of Fourth International Conference on Emerging Intelligent data and Web Technologies, 2013, pp. 708–713.
- [2] D. Chen, J. Odobez, J. Thiran, A localization/verification scheme for finding text in images and video frames based on contrast independent features and machine learning methods, *Signal Process.: Image Commun.* 19 (2004) 205–217.
- [3] E.K. Wong, M. Chen, A new robust algorithm for video text extraction, pattern recognition, in: The Proceedings of First Asian Conference on Pattern Recognition (ACPR), vol. 36, 2003, pp. 1397–1406.
- [4] C. Liu, C. Wang, R. Dai, Text detection in images based on unsupervised classification of edge-based features, in: ICDAR, 2005, pp. 610–614.
- [5] V. Wu, R. Manmatha, E.M. Riseman, Finding text in images, in: The Proceedings of the ACM International Conference on Digital Libraries, 1997, pp. 23–26.
- [6] A. Gonzalez, L.M. Bergasa, Text reading algorithm for natural images, *Image Vis. Comput.* 31 (3) (2013) 255–274.
- [7] C. Zeng, W. Jia, X. He, Text detection in born-digital images using multiple layer images, in: ICASSP, IEEE, 2013, pp. 1947–1951.
- [8] T. Phan, P. Shivakumara, C. Tan, A Laplacian method for video text detection, in: The Proceedings of 10th International Conference on Document Analysis and Recognition, 2009, pp. 66–70.
- [9] P. Shivakumara, T. Phan, C. Tan, A Laplacian approach to multi-oriented text detection in video, *IEEE Trans. Pattern Anal. Mach. Intell.* 33 (2011) 412–419.
- [10] V.N. Manjunath Aradhya, M.S. Pavithra, C. Naveena, A robust multilingual text detection approach based on transforms and wavelet entropy, The Proceedings of 2nd International Conference on Computer, Communication, Control and Information Technology (C3IT-2012), vol. 4, Procedia Technology (Elsevier), 2012, pp. 232–237.
- [11] V.N. Manjunath Aradhya, M.S. Pavithra, An application of K-means clustering for improving video text detection, The Proceedings of the International Symposium on Intelligent Informatics (IS<sup>2</sup>I), vol. 182, Springer, Berlin, Heidelberg, 2013, pp. 41–47.
- [12] C.S. Burrus, R.A. Gopinath, H. Guo, Introduction to Wavelets and Wavelet Transforms, A Primer, Rice University Houston, Prentice-Hall, Inc., Texas, New Jersey, 1998.
- [13] J. Ilonen, J.K. Kamarainen, H. Kalviainen, Efficient computation of Gabor features, in: Research Report, Lappeenranta University of Technology, Lappeenranta, Finland.
- [14] B. Mirkin, *Core Concepts in Data Analysis: Summarization, Correlation and Visualization*, Springer-Verlag, London, 2012.
- [15] C. Naveena, V.N. Manjunath Aradhya, A linked list approach for handwritten textline segmentation, *J. Intell. Syst.* (2012) (in press).
- [16] P. Shivakumara, R.P. Sreedhar, T.Q. Phan, S. Lu, C.L. Tan, Multioriented video scene text detection through Bayesian classification and boundary growing, *IEEE Trans. Circuits Syst. Video Technol.* 22 (8) (2012) 1227–1235.
- [17] D. Karatzas, F. Shafait, S. Uchida, M. Iwamura, L. Gomez, S. Robles, J. Mas, D. Fernandez, J. Almazan, L.P. de las Heras, ICDAR 2013 robust reading competition, in: The Proceedings of 12th International Conference of Document Analysis and Recognition, IEEE CPS, 2013, pp. 1115–1124.