A New Log Gabor Approach for Text Detection from Video

Sudir Prakash, M. Ravishankar
Dayanand Sagar College of Engineering
India
{sudirhappy, ravishankarmcn}@gmail.com

ABSTRACT: *Text information embedded in video frames play an important role in content-based multimedia indexing and retrieval, as a result the automatic detection of texts from videos has gained wide attention in recent years. However,large variations in size, color, orientation and background complexity makes video text detection and extraction a challenging problem. In this paper an effective method for text detection and localization approach based on Log Gabor filter and Block Eigen map analysis is proposed. Log Gabor filter response is used to identify candidate text regions and Block wise Eigen map analysis is used to classify pixels as true text pixels. Finally,a heuristic based on gray scale pixel co-occurrence to eliminate false positives from the frame is used. Experimental results show the promising overall performance of the proposed method against the best features reported in the literature for the standard databases.*

Keywords: Video Frame, Multimedia Indexing, Gabor Filter, Video Text Detection

Received: 19 January 2014, Revised 2 March 2014, Accepted 9 March 2014

1. Introduction

With growing advancements in Data communication, video generation, storage and transmission have received tremendous attention in research community. Semantic video analysis and management, which include video indexing and retrieval procedures, are necessary for effective utilization of video repositories. Textual information in videos proves to be an important source of high-level semantics for content based video analysis and hence text information is utilized for automatic annotation, indexing, summarization and searching in videos [1], [2], [3]. There exist mainly two kinds of text occurrences in videos, first being the artificial text [4] in which text is artificially added in order to describe the content of the video or to give additional information related to it, which makes it highly useful for building keyword indexes. Second is the scene text that is captured by a camera as part of a scene such as text on T-shirts, Road signs, Warnings such as FIRE and EXIT, Name Plates [5] finds its importance in Navigation, Video Surveillance, Sports Video Indexing etc. In case of graphics text or artificial text, we can expect uniform color pixels, high contrast compared to its background and horizontal or vertical direction while scene text can have multiple color pixels, different fonts, contrast and orientation etc.

Basically there exist three approaches for video text detection as described below: The first approach consists of Connected Component (CC)-based methods [4], [6], [7] which assume uniform color text regions and uniform size, shape, and spatial alignments. The extracted text is thus of monotonous colors which follow certain size constraints and horizontal alignment constraints. Because of these constraints, these methods are not effective for the video images having complex background and too low contrast text. The second approach is the Edgebased methods [8], [9], [10], [11] which require text to have a considerably high contrast to the background in order to detect the edges. This approach is faster than CC-based methods but produces many false alarms because of strong contrast background.The texture based approach such as [4], [12], [13], [14], [15] considers the appearance of text pattern as a special texture property overcoming the problems of the CC and Edge based methods. Gabor filters, FFT, Wavelets are mainly employed to exploit the textural properties. Celina Mancas et al. [16] proposed a system for Character Segmentationby- Recognition using Log-Gabor Filters to take advantage simultaneously of gray-level variation and Spatial location. Guru et al. [15] proposed a Eigen Analysis based text detection scheme on gradient images.This method basically fails in fixing proper threshold value to find potential text candidate blocks and hence gives a high false positive rate and low precision. Overall texture based methods work well for complex background with less false positives but at the cost of increased computations due to use of expensive classifier and samples for training. Apart from the superimposed video text detection methods,there are methods available in the literature for text detection from natural scene images [16], [17], [18]. Though these methods work well for complex scene images, we cannot use the same methods directly for Artificial video text detection due to the presence of both low and high contrast text.We also noted that there are methods for multi-oriented text detection in video which detect both horizontal and non-horizontal text. However, these methods use expensive transform and theory to handle the problem of orientations [19], [20], [21], [22]. Achieving good detection rate for the video images having both graphics and scene text is challenging and interesting. Most of the existing methods focus on either graphics text or scene text but not both to achieve better accuracy. Therefore, there is room for developing a method for detecting both graphics text and scene text in video.Restricting the scope to horizontal and vertical text we propose a new method based on Log Gabor filter and Eigen value analysis to detect text of both graphics and scene text in video. The Log Gabor helps in extracting prominent text pixels in different orientations and the Eigen value analysis over Log Gabor feature helps greatly in distinguishing between text and non-text pixels. Proposed method is the first work that considers the combination and concept of Log Gabor and Eigen value analysis for Video Text Detection and Extraction. This is the main contribution of our work.

2. Proposed Approach

Drawing inspiration from texture based methods which show good accuracy for images having complex background a Log Gabor based method is proposed to achieve better accuracy without using any classifier. Observing that strong responses in vertical, horizontal and diagonal directions of text pixel give strong clue about the presence of the text the Log Gabor filter detects text in different directions. As it is stated in [15] the Eigen value analysis highlights text pixel when it performed on gradient of the input image. This analysis motivated us to use Eigen value analysis on Log Gabor filtered image. The output of Eigen value analysis gives sufficient difference between text and non-text pixels. Finally,the Gray Level Co-Occurrence Matrix (GLCM) of the text region is used for false positive elimination. The process of algorithm is outlined with the help of block diagram as shown in Figure 1.

2.1 Identification of Text Regions

Gabor Filters [23], [24], [25] are orientation sensitive local frequency detectors. Gabors work synthesizes the studies of Nyquist in Communication Theory in 1924 and Heisenberg in Quantum Mechanics in 1927, by which he proposes the Gaussian shape as an optimal envelope for time-frequency representation turning the uncertainty principle from inequality to equality. Gabor Filters provide spatial and frequency information for text localization in images but since the maximum bandwidth captured by a Gabor filter cannot exceed more than one octave it limits the feature size that can be captured for text detection. To overcome this problem, we propose Log- Gabor function [26] based text detection method. Log-Gabor filters make it possible to vary the bandwidth from one to three octaves. The features used become more informative, effective and reliable. Hence the focus is on the Logarithmic frequency scale instead of linear frequency scale. Log-Gabor Filters have null DC component and can be constructed with an arbitrary bandwidth which can be optimized to produce a filter with minimal spatial extent.On the linear frequency scale the Log- Gabor function has a transfer function of the form

$$g(\omega) = exp\left(\frac{-\ log\ (\omega/\omega_0)^2}{2log\ (k/\omega)^2}\right) \tag{1}$$

where ω_0 is the centre frequency of the filter and the bandwidth is determined by the k/ω_0 term. To obtain constant shape ratio filters the term k/ω_0 must be held constant for varying ω_0. More Specifically Log-Gabor filters in frequency domain for a nxn image can be defined by the polar coordinates by

$$G_{(s,t)}(\rho, \theta) = exp\left(\frac{-1}{2}\left(\frac{\rho - \rho_s}{\sigma_p}\right)^2\right) exp\left(\frac{-1}{2}\left(\frac{\theta - \theta_{(s,t)}}{\sigma_\theta}\right)^2\right) \tag{2}$$

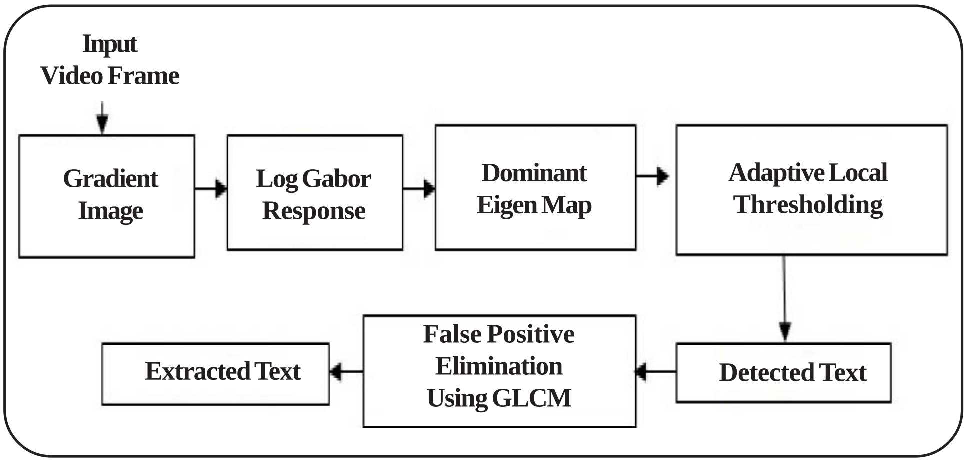


Figure 1. Block Diagram of the proposed approach

where

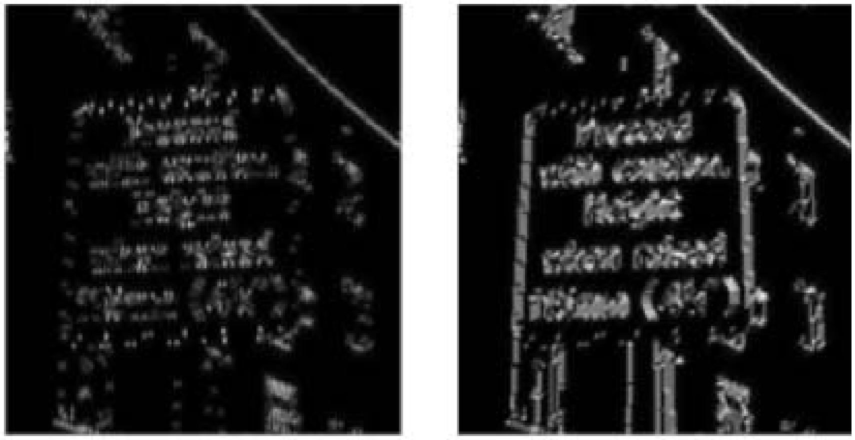
$$\theta_{(s,t)} = \begin{cases} \dfrac{\pi}{n_t}\, t & \text{if } s \text{ is even} \\ \dfrac{\pi\,(t+0.5)}{n_t} & \text{if } s \text{ is odd} \end{cases} \quad (3)$$

(a)

(b)

(c)

(d)



(e) (f)

Figure 2. (a) Original Image (b) Gradient Image (c) Filter Response in 0^o (d) Filter Response in 45^o (e) Filter Response in 90^o (f) Filter Response in 270^o

$$(\sigma_\rho, \sigma_\theta) = 0.996 \left(\sqrt{\frac{2}{3}}, \frac{1}{\sqrt{2}} \frac{\pi}{n_t} \right) \tag{4}$$

where (ρ, θ) are the Log-polar coordinates, n_s is the scale of the multiresolution scheme, n_t is the number of orientations. $s \in (1,.., n_s)$ and $t \in (1,.., n_t)$ indexes the scale and the orientation of the filter respectively. (ρ_s, θ (s, t)) are the coordinates of the center of the filter and (σ_ρ, θ) are the bandwidths in and θ common for all filters. The proposed approach chooses wavelet scale $n_s = 1$ sufficient to describe an image and filter orientations is choosen as $n_t = 4$ or 8.

Illustration for responses given by Log Gabor is shown in Figure 2 where for the input image shown in Figure 2 (a), Gradient Image output is shown in Figure 2(b) and the responses in different directions are shown respectively in Figure 2 (c-f). It is observed that text pixels show strong responses in the respective directions. It is also observed that Log Gabor filter also enhances high contrast background information which looks like text. To solve this problem, we propose Eigen value analysis on Log Gabor filter images, which is capable of preserving the magnitude of text region of maximum variance. The potential candidate text blocks are found by generating Dominant Eigen map.

2.2 Dominant Eigen Map generation

Each filters response is subdivided into $k \times k$ blocks of size $m \times m$ where m is power of two. For each k block, the covariance matrix is computed given by

$$\Psi = \frac{1}{m} \sum_{j=1}^{m} ((k_j - \bar{k})^2 \; (k_j - \bar{k})) \tag{5}$$

where k_j is the intensity value in every j^{th} column and $\bar{k}$ is the mean of all elements in k^{th} block.The Eigen values λ_{ij} of the covariance matrix Ψ of the block are computed. Maximum Eigen value of each block is stored in a $k \times k$ matrix and is called an Eigen map. Figure 3 shows the Eigen maps for all the 4 filter responses and is denoted by EM_Ω where Ω = 0, 45, 90, 135. The four Eigen maps generated as shown in Figure 3 (a)-(d) where we can see that the text pixels are sharpened and then the Eigen Maps are combined to form one image called as Dominant Eigen Map (DEM) as shown in Figure 3 (e) which is generated by using the following rule

$$DEM = \prod_{\Omega = 0,\, 45,\, 90,\, 135}^{4} EM\Omega \tag{6}$$

From Figure 3 (e), it can be seen that, non-text responses are suppressed to a maximum extent by generating the Dominant Eigen Map. Next An adaptive thresholding scheme is proposed to ensure that text regions only are highlighted and non-text regions are suppressed. The result is shown in Figure 3 (f) where one can see clear text region and non-text region.

2.3 Adaptive Local Thresholding

The Global thresholding technique is not satisfactory for video text detection as complex background video frames exhibit local variance whereas adaptive thresholding changes the threshold dynamically over the image. It is noted from the literature [27], [28], [29] that Adaptive local thresholding can accommodate changing lighting conditions in the image, for example, those occurring as a result of a strong illumination gradient or shadows. Adaptive local thresholding dramatically

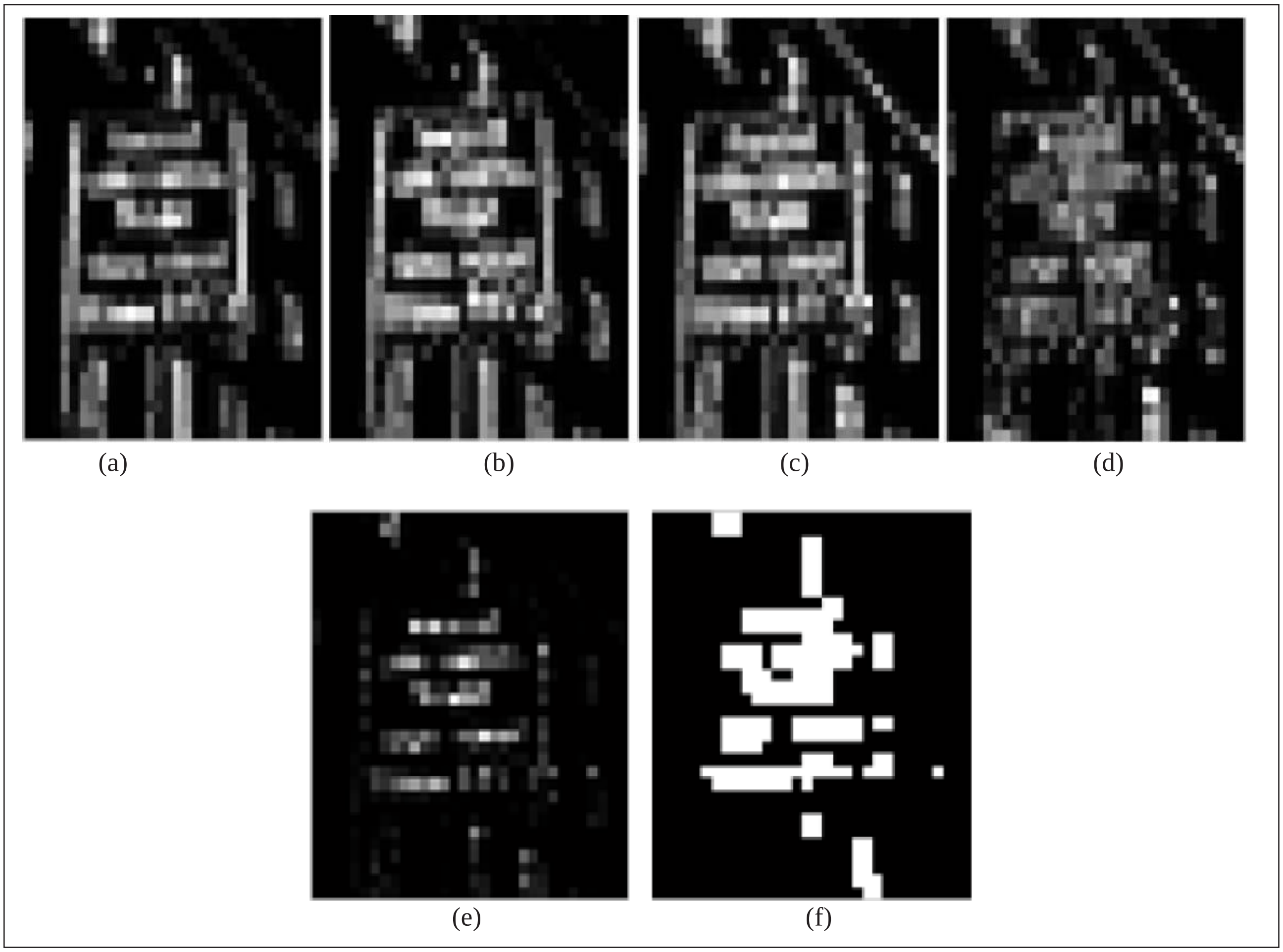


Figure 3. (a)-(d) Eigen maps e) Dominant Eigen map f) Adaptive Threshold Image

improves the quality of segmentation results for small characters. A window size varying between 2-10 is chosen based on experimental study on different dataset. The same values are used for all experiments on different database. The values of the parameters are studied based on geometrical properties of the text blocks. Figure 3 (f) shows the result after Adaptive local thresholding choosing 88 neighborhood. The adaptive threshold image ensures that probable text blocks have value 1 and non text Blocks have value 0. However, adaptive thresholding technique may not prevent false positives completely. Therefore, we propose new criteria based on gray level co-occurrence matrix to eliminate false positives.

2.4 False Positive Elimination

A statistical method of examining texture that considers the spatial relationship of pixels is the Gray-Level Co-occurrence Matrix (GLCM) [30], also known as the gray-level spatial dependence matrix. The GLCM functions characterize the texture of an image by calculating how often pairs of pixels with specific values and in a specified spatial relationship occur in an image, creating a GLCM, and then extracting statistical measures from this matrix. Given an image I, of size $N \times N$, the co-occurrence, matrix P can be defined as

$$p(i, j) = \sum_{x=1}^{N} \sum_{y=1}^{N} \begin{cases} 1 & I(x+\Delta x, y+\Delta y) = j \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Here, the offset (Δx, Δy), specifies the distance between the pixel-of-interest and its neighbor. GLCM analyzes pairs of horizontally adjacent pixels in a quantized version of region within Bounding Box. In this case, there are $8 \times 8 = 64$ possible ordered combinations of values for each pixel pair. GLCM accumulates the total occurrence of each such combination, producing an 8-by-8 output array, GLCM. The row and column subscripts in GLCM correspond respectively to the first and second pixel-pair values. It is found that the GLCM of non text regions is dominated by value 0 because of least variations in gray levels whereas text regions GLCM has least zeros (because the Intensity level are well distributed). Hence number of zeros in a GLCM can clearly differentiate the text regions from non text regions.

Let AVG_{NCZ} be the average value of Zeros among Bounding Boxes. Let NCZ_{σ} be the Number of Zeros in GLCM of each Bounding Box. Let Max_{NCZ} be the Maximum Value of number of Zeros among regions. Let Min_{NCZ} be the Minimum Value of number of Zeros among σ regions.

$$AVG_{NCZ} = \frac{Max_{NCZ} + Min_{NCZ}}{2} \tag{8}$$

Discard Bounding Box if $NCZ_{\sigma} > AVG_{NCZ}$
Bounding Box = Text region σ if $NCZ_{\sigma} > AVG_{NCZ}$

We arrive at this threshold formula based on experimental study on different datasets. This approach outperforms traditional false positive elimination approach based on Size, Color, Aspect ratio, Edge density, Corner Density etc. as evident in Table 1, 2, 3 and the sample results are shown in Figure 4.

Figure 4. (a) Detected Text (b) Result after False Positive Elimination (c) Extracted Text

3. Performance Evaluation Metrics

The Performance of Text detection is compared with existing methods using metrics Recall, Precision, Fmeasure, Miss-Detection Rate (MDR), Average Processing Time (APT) [31]. Precision defines how many percents of the detected video text regions are correct. Recall evaluates how much percent of all ground-truth video text regions are truly detected. F-measure (false alarm rate) measures how much percent of the detected video text regions are wrong. The Detected block is represented by bounding box.

In this experiment, we treat all bounding boxes within the same horizontal text line as one text block to evaluate the performance of the methods. Due to the challenges of scene text and the arbitrary orientation of the text lines, it is difficult for a method toalways enclose a full text line in a block. Sometimes it misses some characters of very low contrast and detects only parts of a line. Since the goal of this paper is text detection (how well a method locates potential text blocks), partial detection (even to the extent of just a few true characters) is still acceptable because it shows that a method is able to locate a block (albeit partially).

The following measures for each detected block by text detection method are defined.

• **Truly Detected Block (TDB):** A detected block that contains at least one true character. We define that a detected text region is true on the condition that the intersection of the detected text region and ground-truth text region exceed 80%.

• **False Detected Block (FDB):** A Detected block that do not contain text.

• **Text Block with Missing Data (MDB):** A Detected block that misses more than 20% of the characters of a text line. We choose 20% based on instructions given in the papers [15], [19], [20], [32] for calculating misdetection rate.

• **Average Processing Time (APT):** Processing time per frame required for detecting text in images (in seconds).

For each image in the dataset, The Actual Text Blocks (ATB) are counted manually, i.e. ground truth data. The performance measures are defined as follows.

(i) Recall $(R) = \dfrac{TDB}{ATB}$

(ii) Precision $(P) = \dfrac{TDB}{TDB + FDB}$

(iii) F-measure $(F) = \dfrac{2 \times P \times R}{P + R}$

(iv) Misdetection Rate $(MDR) = \dfrac{MDB}{TDB}$

4. Datasets

Due to non-availability of standard video database, a database of 100 MPEG-I Video sequences with 320×240 resolutions from various news channels (Tv9, NDTV, Times Now, BBC) was created to evaluate the validity of our proposed method. There were a total of 3000 frames with 2750 overlay text events consisting of variety of texts in different font-size, font-color and languages, text in textured background, poor quality text etc. Video sequences were captured at 30 frames per second. Apart from own database, the proposed algorithm is tested on publicly available datasets such as Hua0s (Microsoft Asia Database) dataset [31]which contains 45 images (33 for graphics text and 12 for scene text). Though it is a small dataset, it provides an objective study for the proposed method. We consider this publicly available video data as a standard and benchmark data to evaluate the proposed method performance. Though our focus is on text detection in video, we test the proposed method on camera based scene data which is available publicly to show the effectiveness and advantage of the proposed method. ICDAR 2003/05 Dataset [5] which contains 251 Camera images and ICDAR 2011 Robust Reading Competition Dataset [32] which consists of 255 images are considered . The text characters in these images include English and numeral characters. The proposed approach is implemented using MATLAB running on a PC with Pentium IV 1 GHz processor. The proposed method is processed in frames. Since the images in the ICDAR dataset are of high resolution, they are resized to a standard width and height of 256×256 to save computational costs.

5. Experiments Results

5.1 Performance Analysis on Our Dataset

Proposed Work is compared with the method in [15] because it uses gradient and Eigen value analysis similar to the proposed method and it is a recent method. To show the effectiveness of the different directions of Log Gabor filter, we conduct experiments for four directions and eight directions as shows in Figure 5 where (a), (b) are the input video images, Figure 5 (c), (d) are the results of 4 directions of Log Gabor filter and Figure 5 (e), (f) are the results of 8 directions of Log Gabor filter. It is noticed from the Figure 5 (c), (d) and Figure 5 (e), (f) that when we increase the directions from 4 to 8 directions the text detection results improves as it eliminates false positives and fixes proper bounding boxes. As a result, it is concluded that 8 directions of Log Gabor filter is better than 4 directions. Hence, we use 8 directions Log Gabor filter to compare with the existing method for other dataset. Figure 5 (g), (h) shows the results of the existing method [15] which produces lots of false positives compared to the proposed methods results shown in Figure 5 (e), (f). The reason for getting more false positives of the existing method is that the Eigen method fixes a constant threshold to classify text and non-text candidates.As per the experiment results listed in Table 1, one can observe that the proposed method outperforms the existing method [15] in terms of recall, precision F-measure, MDR and Average Processing Time (APT).

5.2 Performance Analysis on Hua’s Dataset

Though Hua’s dataset is small, it gives a fair comparison with the existing methods. Sample results of the proposed and the existing method are shown in Figure 6 where (a)-(c) are the output of the proposed method and (d)-(f) are the results of the existing method [15]. It is noticed from Figure 6 that the proposed method detects almost all text lines in the input images while the existing method fails to detect text in all the images. The results reported in Table 2 show that the proposed method is better than the existing methods in terms of Recall, Precision, F-measure and Average Processing Time.

Methods	R	P	F	MDR	APT(secs)
Proposed	0.87	0.86	0.86	0.08	0.49
Guru [15]	0.76	0.85	0.8	0.17	0.57

Table 1. Comparison of the Proposed Method With the Existing Method on our Database

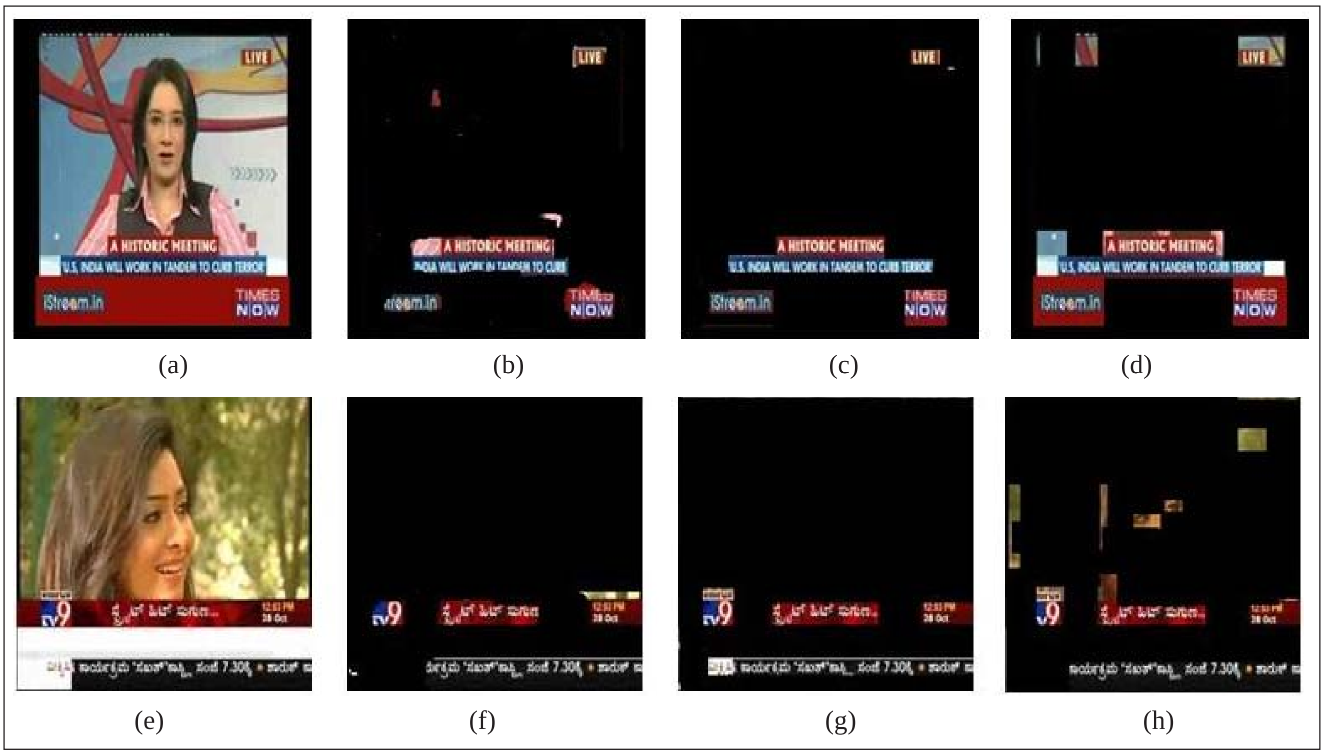


Figure 5. Text extraction result for video frames: (a, b) video frames (c, d) text extraction results for 4 orientations (e, f) text extraction results for 8 orientations (g, h) text extraction results for Eigen based method

Methods	R	P	F	MDR	APT (secs)
Proposed	0.95	0.82	0.88	0.19	0.56
Guru [15]	0.85	0.82	0.83	3.83	0.76
Liu [33]	0.75	0.54	0.63	0.16	22
Cai [9]	0.69	0.43	0.53	0.13	1.1
Laplacian [31]	0.93	0.81	0.87	0.07	7.8

Table 2. Comparison of the Proposed Method With the Existing Method on Hua's Database

5.3 Performance Analysis on ICDAR 2003/05 and 2011 Dataset

Experiments conducted shows that the proposed method is capable of detecting text in natural scene images which are usually high resolution and complex background images compared to low resolution and complex background video images. However, we calculate Recall, Precision and f-measure at line level but not word level suggested by the ICDAR 2003 competition

(a) (b) (c)

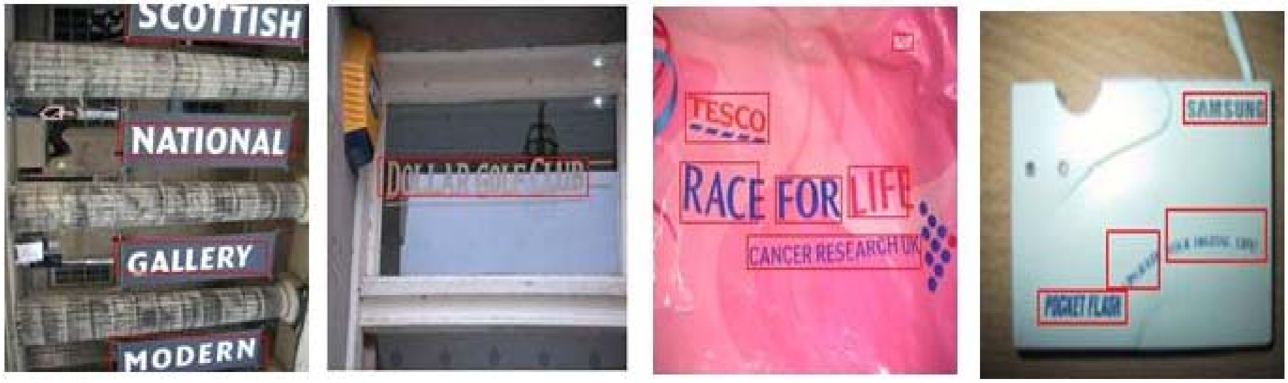
(d) (e) (f)

Figure 6. (a), (b), (c) Text Detection Result of the proposed method
(d), (e), (f) Text Detection Result of the Eigen based method

and ICDAR 2011 competition because generally, according to literature on video text detection, the methods calculate measures at line level for video data due to lots of disconnections and loss of information. This is common practice for video text detection methods [15], [20], [31], [34]. Hence, to maintain consistency, we use the same criteria that used for video to calculate measures for ICDAR data. The sample results of the proposed method for ICDAR 2003/05 and ICDAR 2011 are shown in Figure 7 where one can notice that the proposed method works well for even ICDAR data with few false positives and misdetections. The proposed method has the Highest Recall rate, Good Precision rate and Highest F-Measure (almost the same as that of the Nabins method). This shows the advantage of the proposed method because it achieves good results while making fewer assumptions about text. The existing method can remove false positives more easily by using GLCM based approach irrespective of size and Aspect ratio of the CCs. In addition, the computational time is also quite low as compared to other methods which use many more and more costly classifiers like SVM and NN to classify text and non-text

Dataset	**Methods**	**R**	**P**	**F**	**MDR**	**APT(secs)**
ICDAR 2003/05	Proposed	0.88	0.72	0.79	0.14	0.65
	Eigen[15]	0.832	0.57	0.67	0.38	0.75
	Laplacian[31]	0.87	0.72	0.78	0.14	7.9
	Wavelet[21]	0.54	0.836	0.65	0.65	2
	Nabin[32]	0.81	0.78	0.79	0.23	5.39
ICDAR 2011	Proposed	0.629	0.74	0.68	0.22	0.6

Table 3. Comparison of the Proposed Method with the Existing Method on ICDAR Database



(a) (b) (c) (d)

Figure 7. Text Detection Result of the proposed method

regions with a large number of samples to train the classifiers. Till recent time there are nomethods that calculate Precision, Recall and F-measure at line level for ICDAR 2011 database and proposed work is the first of its kind hence comparison is not possible. Overall the proposed method is effective since it works well for video data and ICDAR data as well when compared to existing methods.

6. Conclusion

This paper presents a novel method based on Log Gabor filter and Eigen value analysis for text detection in video. We have explored the combination of Log Gabor Eigen values analysis to achieve better accuracy in text detection. It is shown that the proposed method is good enough to handle both graphics and scene text in video because of the advantages of 8 directions of Log Gabor filter and Eigen values analysis. It is also shown that the proposed method works well for both video data and camera based scene images. To improve the precision, the method explores gray level co-occurrence matrix to eliminate false positives efficiently. As a result, the proposed method outperforms the existing methods in terms of recall, precision, F-measure and average processing time for all dataset. However, the method has not been tested on multi-oriented and arbitrary oriented text lines in video. It is considered as our future work.

References

[1] Kim, K. I., Jung, K., Jain, A. K. (2004). Text information extraction in images and video: A survey, *Pattern Recognition*.

[2] Zang, J., Kasturi, R. (2008). Extraction of text objects in video documents: Recent progress, DAS.

[3] Pal, U., Sharma, N., Blumenstein, M. (2012). Recent advances in video based document processing: A review, DAS.

[4] Kia, O., Li, H., Doermann, D. (2000). Automatic text detection and tracking in digital video, IEEE Trans. Image Process.

[5] Lucas.Icdar, S. M. (2005). Text locating. *In*: Proc. ICDAR.

[6] Huang Y. Cui. (1997). Character extraction of license plates from video. *In*: Proceedings of International Conference on Pattern Recognition.

[7] Mariano, V. Y., Kasturi, R. (2000). Locating uniform-colored text in video frames, *In*: Proceedings of International Conference on Pattern Recognition.

[8] Hua, X. S., Xi, J. (2001). A video text detection and recognition system, *IEEE International Conference on Multimedia and Expo*.

[9] Lyu, M. R., Cai, M., Song, J. (2002). A new approach for video text detection, *IEEE International Conference on Image Processing*.

[10] Cai, M., Lyu, M.R., Song. J. (2005). A comprehensive method for multilingual video text detection, localization, and extraction, *IEEE Transactions on Circuit and Systems for Video Technology*.

[11] Li, H., Doermann, D. (2000). Automatic text detection and tracking in digital video, *IEEE Transactions on Image Processing*.

[12] Wernike, A., Lienhart, R. (2000). On the segmentation of text in videos. IEEE International Conference on Multimedia and Expo (ICME).

[13] Park, S. H., Kim, K. L., Jung, J. (2001). Support vector machine-based text detection in digital video, *Pattern Recognition*.

[14] Jung, K., Kim, K. L., Kim, S. H. (2003). Texture-based approach for text detection in images using support vector machines and continuous adaptive mean shift algorithm, *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[15] Shivakumara, P., Guru, D. S., Manjunath, S., Tan, C. L. (2010). An eigen value based approach for text detection in video, *In*: Proc. DAS.

[16] Thillou, C. M, Gosselin, B. (2006). Character segmentation-by-recognition using log-gabor filters, *In*: Proc. ICPR.

[17] Wexler, Y., Epshtien, B., Ofek, E. (2010). Detecting text in natural scenes with stroke width transform, CVPR.

[18] Prevost, L., Muhammad Hanif, S. (2007). Detection in natural scene images using spatial histograms. CBDAR.

[19] Phan, T. Q., Shivakumara, P., Tan, C. L. (2011). A Laplacian approach to multioriented text detection in video, *IEEE Trans. PAMI*.

[20]Phan, T. Q., Shivakumara, P., Tan, C. L. (2009). A Robust wavelet transform based technique for video text detection, *In*: Proc. of ICDAR.

[21] Phan, TQ., Lu, S., Shivakumara, P., Sreedhar, RP, Tan, C. L. A Robust wavelet transform based technique for video text detection. IEEE Trans CSVT.

[22] Liu W Ma Yand Tu Z Yao C, Bai X. Detecting texts of arbitrary orientations in natural images, *In*: Proc CVPR.

[23] Gabor, D. (1946). Theory of communications. *Journal of Institute of Electrical Eng*.

[24] Kamarainen, J. K., Kyrki, V., Klviinen, H. (2003). Simple gabor feature space for invariant object recognition. *Pattern Recognit. Lett*.

[25] Jain, A. K., Farrokhnia, F. (1991). Unsupervised texture segmentation using gabor filters, *Pattern Recognition*.

[26] Field, D. J. (1987). Relation between the statistics of natural images and the response properties of cortical cells, *J. Opt. Soc. Am*.

[27] Trier, D., Taxt, T. (1995). Evaluation of binarization methods for document images, *IEEE Trans. PAMI*.

[28] Trier, O. D., Jain, A. K. (1995). Goal directed evaluation of binarization methods, *IEEE Trans. Pattern Anal. Mach. Intell. PAMI*-17, 1191– 1201.

[29] Otsu, N. (1979). A threshold selection method from gray level histograms, *IEEE Trans. Systems, Man, and Cybernetics*.

[30] Baraldi, A., Parmiggiani, F. (1995). An investigation of the textural characteristics associated with gray level co-occurrence matrix statistical parameters., *IEEE Trans.On Geoscience and Remote Sensing*.

[31] Shivakumara, P., Phan, T. Q., Tan, C. L. (2009). A Laplacian method for video text detection. *In*: Proc.

[32] Umapada Pal, Michael Blumenstein Nabin Sharma, Palaiahnakote Shivakumara, Chew Lim Tan. (2012). A new method for arbitrarily-oriented text detection in video,10th IAPR International Workshop on Document Analysis Systems.

[33] Wang, C., Liu, C., Dai, R. (2005). Text detection in images based on unsupervised classification of edge-based features, *In*: Proc. ICDAR5.

[34] Trung Quy Phan, Shijian Lu Palaiahnakote Shivakumara, Rushi Padhuman Sreedhar, Chew Lim Tan. (2012). Multi-oriented video scene text detection through bayesian classification and boundary growing, *IEEE Transactions on Circuits and Systems for Video Technology*.