IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-8, Aug 2018 E-ISSN: 2347-2693

Gujarati Text Localization, Extraction and Binarization from Images

J. M. Patel[1*], A. A. Desai[2]

[1]M. K. Institute of Computer Studies, Bharuch, Gujarat, India
[2]Department of Computer Science, Veer Narmad South Gujarat University, Surat, Gujarat, India

[*]*Corresponding Author: jagin_2k@yahoo.com*

Available online at: www.ijcseonline.org

Accepted: 13/Aug/2018, Published: 31/Aug/2018

***Abstract*—** In this paper, first ever attempt to automatic detection and extraction of Gujarati text from image has been presented. This method is based on Haar discrete wavelet transform (HDWT), edge detection and connected component analysis. The distinct features of Gujarati script make it hard to directly apply the existing text detection methods designed for English, Chinese and other languages. In proposed work, first high-frequency wavelet coefficients are extracted using HDWT and then sobel filter is applied on to detect candidate text edges. Then connected component analysis is performed using area geometric feature to get rid of non-text area. The proposed method is tested on a variety of images such as images of complex background and images of different fonts, colour, and size of text. The experiment on over 878 images show that the proposed method can detect regions and isolation of text perfectly including modifier. The proposed framework has achieved a precision of 0.91, recall of 0.97 and f measure of 0.94.

***Keywords*—**Gujarati text detection, Text extraction, Text binarization, HDWT, Connected Componenet

I. INTRODUCTION

Text is a fundamental medium for visual communication. Various Images like – Charts, Logos, Graphs, Maps, Headers, Footers and Equations – may contain vital data in form of texts embedded in images. Current tools and search engines can detect text written in text document but are unable to extract any text from such type of images [1].

Automatic text detection from image is useful to access and utilize textual information. Factors contributing to the complexness of the text detection problem include: non-uniform background, complexity of image backgrounds, mixture of font sizes and styles, orientations, alignment, reflections, shadows, low-quality of image list to few.

Content-based image/video search, analysis and understanding have made text extraction a crucial task. So, In an Image processing research, it is necessary to make systems that allow us to understand image content, meaning and searching.

Over the past number of decades, Text extraction from images/video have many useful applications and several researchers have contributed to the progress in the field: multimedia retrieval [2, 3, 4], visual input and access [5, 6, 7], industrial automation [8, 9, 10], detection of vehicle number plate [11, 12], keyword based image search, and serving to visually impaired individuals [13, 14, 15, 16].

A. Characteristics of Gujarati script

Gujarati is an Indo-Aryan language, one of the official languages of India and used as linguistic communication to the Indian state of Gujarat. It has twelve vowels (known as Swar) and thirty six consonants (referred to as Vyanjan) known as basic symbols, and ten digits (known as numerals). In addition to basic symbols, diacritics (dependent vowel modifier) are used to express the linking of vowels with the core consonants. The diacritics can appear before, after, above, or below the basic symbol as shown in Fig. 1.

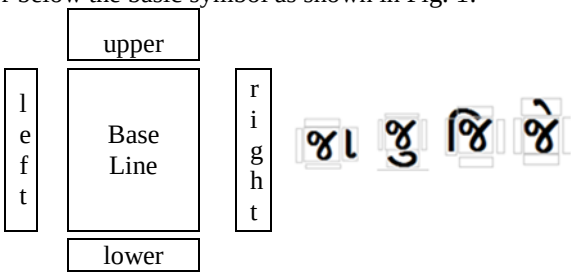


Figure 1. Position of base character and modifier

Gujarat script can be partitioned into three logical zones: Upper, Middle and Lower [17, 18] as shown in Fig. 2.

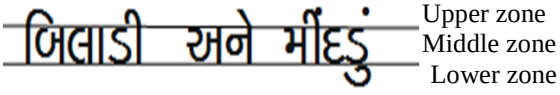


Figure 2. Zone in Gujarati script

Normally, Modifier binds to Upper/lower zone, whereas the consonants and vowel occupied in the middle zone. Middle zone may contain sub-parts of modifier i.e. some modifier appears in more than one zone.

One more characteristics of Gujarati script not found in Devanagari language and other Indian script like Bangle and Hindi is that it doesn't have shirorekha or Header lines.

Spotting Gujarati text in complex background images becomes a difficult problem, because of the distinctive features of the Gujarati text compared to the other most often used languages. The distinct features make it hard to directly apply the existing text detection methods designed for English, Chinese and other languages. Compared to other languages, Gujarati has enormous diversity in text writing like:

(i) Absence of shirorekha: The presence of shirorekha in different Indian languages helps to identify text lines and simplifies extraction of upper modifiers. In Gujarat script absence of shirorekha make it difficult for the extraction of upper modifiers, segmentation of text lines, and words extraction from Gujarati text

(ii) Some characters like ગા,ઘા,શા,ઘ are made of two or more distinct curve/shape and have space within the subpart of the character, which cause untrue determination of character boundaries and decode a single character into more than one characters.

(iii) The modifiers like ા (kano) associated with the basic characters.

(iv) Space between middle zone and upper/lower zone, i.e. upper part or lower part may not be connected with middle zone, cause issue in line extraction.

(v) Due to its shape and size, it is difficult to decide whether or not upper modifier anusvar (ં) is element of text. It will be simply mistaken as noise.

(vi) English characters presence solely within the middle zone, therefore, minimum height is fixed according to the height of the characters; however this cannot apply for Gujarati script because of the appearance of characters in all the three zones.

(vii) Diversity of text: Text in document usually appears with regular font, single colour, consistent size and uniform arrangement. While Gujarati texts in book cover page, books for kids, story books etc. may carry entirely different fonts, colours, scales even within the same image.

The rest of this paper is organized as follows. Section two briefly reviews the related works. In Section three, the complete description of the used technique for Gujarati text detection and localization is given. Text segmentation and binarization process is described in section four. The dataset and analysis protocols are elaborated in section five. Sixth section provides experimental results on various kinds of images. Finally, we give conclusion in Section seven

II. RELATED WORK

In recent years researchers from various communities like computer vision, pattern recognition and document analysis have projected many methods for text detection and recognition in natural images. Text detection and localization from image is worldwide problem in image processing and information retrieval. Lots of work has been done and several methods have been proposed in these direction. All these methods have pros and cons of their own, so there is always a possibility for improvement.

Text localization methods are classified into three classes: (1) texture based methods (2) Region based method (3) Hybrid.

1) **Texture based methods**: In this method texts are considered as a special type of texture and textural-related properties, such as local intensities, filter responses, wavelet coefficients, gradients, FFT etc, are used to make a distinction between text and non text areas in the images. It has ability in processing complex background. In back side, these methods suffer from high complexity as all locations and scales should be scanned. In addition, these methods mostly handle horizontal texts.

(2) **Region based method**: This approach may divide in two classes based on: (i) connected component (CC), and (ii) edge detection.

(i) Connected Component based methods: These methods first detect potential components through a kind of ways (e.g., clustering). Then it removes non-text components using manually designed rules, geometrical constrain, or various automatically trained classifiers. Normally, these methods are efficient, simpler to implement, and computationally inexpensive, because the number of components to be deal with is quite small. The main drawback of CC method is not very robust for locating text regions in images which have complex background. Component based text localization has been widely explored in [19, 20, 21, 22, 23, 24, 25].

(ii) edge-based method: References [26, 27, 28, 29, 30] used edge based methods, which depends on contrast between text and background. Key of this method is the ability to process low contrast text image and variant text size.

(3)**Hybrid method**: Both region based and texture based method have their advantages and drawbacks. Some researchers use Hybrid methods, which makes use of the benefits of each of these methods. To improve robustness on various text categories, hybrid features have been applied in text detection /localization [23, 31, 32, 33, 34, 35, 36, 37, 38].

Some researchers also used classifiers like AdaBoost[15, 39,40,41], CRF[34, 36, 42], neural network [43], SWT[44, 45, 46, 47, 48] ,SVM [49] and MSER [41,50, 51, 52, 53, 54, 55].

The ICDAR Robust text Reading Competition has been held for no. of times [56,57,58,59,60,61], organized around challenges to cover a wide range of real-world situations and text images contain mostly English languages.

In literature, Work is found for text detection and localization for languages like English, Chinese [44,62,63,64,65,66], Japanese[67], Farsi/Arabic [68], Arabic[69,70,71], Kanji [72], Korean [73,74], Urdu [75], Uyghur [76,77].
In addition to multi-oriented [78, 79] text detection, Researchers also worked on Multilingual/ Language independent approaches [80, 81, 82, 83, 84, 85, 86] for text detection.

Work is also discovered for Indian languages (script) such as, Telugu [87,88], Devanagari [89], Gurmukhi [90,91] and Bangla [70,92], Kannad [93,94,95]. Few researches put their attention on Gujarati OCR [17, 96, 97, 98, 99] as well.
In the direction of text extraction for Gujarati language, perhaps, first and only work for Gujarati language is found in 2014[1], which has concentrated on segmentation and recognition of Gujarati printed numeral from image with success of more than 95% recognition. However, to the best of our knowledge, there is no work reported for detecting and extracting Gujarati text from image.

Patel and Desai [1] presented testimonial for segment and recognize Gujarati numeral from image. In this work, researcher used edge detection, dilation and connected component analysis to detect Gujarati numeral from of images. This method successfully worked on variety of images like textured, map, magazine cover pages, and noisy images.

Dholakia et. al. [18] developed algorithm for three zone separation of Gujarati printed text by slop of line. In this work a smearing algorithm is used for joining modifiers (matras) with the text, horizontal and vertical profiles are used for separating words and lines respectively.

In 2016, ali and hasim [26] proposed technique for scene text detection and localization from image based on discrete wavelet transform, edge detection, connected component and morphological operation. Method give good result for law contrast image, but it fails to produce clear background for text extraction. Manjunath et. al. [83] used combination of wavelet transform and Gabor filter to extracts texture features and sharpens edges of an input image. First single level 2-D wavelet decomposition is performed to obtain approximated and detailed co-efficient. Then detail co-efficient are merged and averaged to extract efficient texture feature information. In next stages, Gabor filter is applied to detect edges, k-means algorithm to crate cluster, morphological operations are performed to obtain connected components, and wavelet entropy is used to determine the true text region of an input image.

Desai [96] presented OCR for handwritten Gujarati numeral recognition using a novel hybrid features extraction technique. In year 2010, Desai [97] described how various parameters like paper, writing style, pen etc. affect Gujarati numeral recognition. He also presented an OCR system for handwritten Gujarati numbers, in which, smoothing thinning, skew detection, correction and normalization is performed. A multi-layered feed forward back propagation neural network is used for Gujarati numeral classification. This system achieved success rate of 71.82% for standard fonts, 91% for handwritten training sets and 81.5% for testing sets.

Desai [98] presented OCR for handwritten Gujarati numeral recognition using hybrid features and Knn classifier. Here, 16 features are calculated by dividing image into 16 subimage and then find total number of pixels in each subimage. Aspect ratio of image is also considered as a feature. Thus 17 features are used for numeral identification. Patel and Desai [17, 99] addressed problem of word segmentation for Gujarati handwritten script. In this work they used Radon transform for skew detection and correction. More work for zone identification from Gujarati text is reported by Patel and Desai [99], which addressed problem of zone separation for handwritten Gujarati script, find three zones by its profile analyses and achieve accuracy of 75.2, 75.2, and 86.6 % for separating upper, middle, and lower zones respectively.

Gllavata et al. [100] automatically detect horizontally aligned text using wavelet transform. The distribution of high-frequency wavelet coefficients is used to characterize text and non-text regions. Then k-means algorithm is used to classify text regions and projection analysis is used for text localization. Work presented in [101] is for fast and robust text lines detection in complex images and video frames by using multiscale wavelet features. In this work all possible text pixels are merged using a density-based region growing method, candidate text lines are separated by structural information and an SVM classifier is used to identify true text from the candidates based on the selected features.

III. PROPOSED METHODOLOGY

This research uses hybrid approach for Gujarati text detection and localization from images, which is a combination of texture based and region based methods. Taking advantages of these two types of methods, the proposed work is suggested in three phases. In first phase potential text regions are detected using Haar Discrete Wavelet Transform (HDWT) and edge detection followed by connected component analysis. In second phase text is located and isolated using connected component bounding

box information. In third phase binarization is performed and text is clearly separated from background.
Here is brief algorithm for Gujarati text localization, extraction and binarization from images.

Step-1: Take colour image.
Step-2: Perform preprocessing.
Step-3: Obtain detailed features of an image using HDWT.
Step-4: Extract Gujarati candidate text edges by applying Sobel operator, Weighted OR and dilation operation on detailed components. To locate text region for candidate text, CC analysis and bounding box information is used.
Step-5: Detected text is extracted from original image and segmented. Segmented text is binarized in such way that text is white and background is black regardless of polarity.

Following sections described algorithm in detail.

A. Preprocessing

Here, during this step, Y element, which represents the luminance information, is extracted from colored image by forming weighted sum of the R, G and B elements using (1).

$$[Y] = [0.299 \quad 0.587 \quad 0.114] * \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{1}$$

This research uses luminance elements (Y) for next processing step and the whole Gujarati text detection algorithm afterward.
By applying median filtering the noise of image is reduced and simultaneously edges are also preserved.

B. HDWT

The Discrete Wavelet Transform (DWT) is a useful tool for signal analysis and image processing. Wavelet transform is capable of providing information on frequency and location information concurrently. The transform of a signal is just another way of representing the signal. Actually, it does not change the information content present in the signal.

In DWT, Signals are analyzed by passing it through an analysis filter bank, namely low-pass and high-pass filters. The low pass filter makes smooth variation between gray level pixels while the high pass filter makes high variation between gray level pixels. So, wavelet has ability to detect edges during high-pass filtering and provides details in term of textual content during low-pass filtering.

DWT decomposes signal into a coarse approximation and detail information. 2-D DWT hierarchically split an input image into four components. Actually, the 2D DWT is basically a one dimensional analysis of a 2D signal. It only operates on single dimension at a time, by analyzing the rows and columns of an image separately to produce four different components, one average component (LL) and three detail components (LH, HL, HH) delineated as:

1. LL components: It contains coarse approximation of the image and removes all high frequency information. I.e. in this sub-band average components are detected.
2. LH components: It represents horizontally low-frequent and vertically high-frequent components. This sub-band emphasizes information along the rows and preserves horizontal features, which results in detection of the horizontal edges.
3. HL components: It corresponds to horizontally high-frequent and vertically low-frequent components. This sub band emphasizes data along the columns and preserves vertical features, which results in detection of the vertical edges.
4. HH components: This sub-band holds the horizontally high-frequent and also the vertically high-frequent components. It tends to isolate localized high-frequency point features, and focuses on the diagonal edges
5.

This research makes use of the Haar wavelet to obtain the detailed features of an image for which a Gujarati text pixel is most likely to be located. Haar wavelet is chosen as a result of its effectiveness in text feature characterization. In addition to this, it is efficient in computation than other wavelet with filter coefficient as either 1 or -1. Haar works on data by pairing up input values, calculating the summation and differences of adjacent elements (or pixels), recursively as shown in Fig. 3.

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \rightarrow \begin{bmatrix} A+B & A-B \\ C+D & C-D \end{bmatrix} \rightarrow \begin{bmatrix} (A+B)+(C+D) & (A-B)+(C-D) \\ (A+B)-(C+D) & (A-B)-(C-D) \end{bmatrix}$$

Original image (matrix) | After row operation | After column operation

Figure 3. Single level decomposition of DWT

Fig. 4 shows Wavelet decomposition componenets.

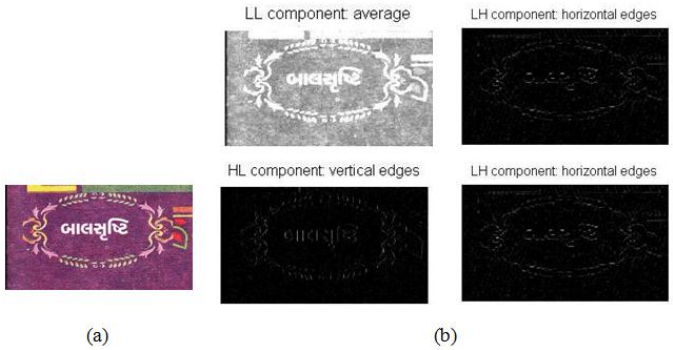


Figure 4. Single level decomposition of HDWT (a) original text image (b) Wavelet decomposition components

C. Extraction of candidate text edges

Gujarati Text edges are relatively short and composed of Horizontal, Vertical, and/or Diagonal edges. The Gujarati text edges could be visible in either one or more of the three edges. So, we can determine the text regions to be the regions where those three kinds of edges are intermixed. Another fact is that text pixels have high variation around its surrounding pixels; therefore a technique based on edge detection is appropriate.

Keeping this features in mind, first, strong edges viz. Horizontal (E^H), Vertical (E^V), and Diagonal (E^D) are extracted from each detail components LH, HL and HH respectively by applying Sobel edge operator [102]. Then, three edges E^H, E^V, and E^D are clubbed together so no text edges get missed. The resulting edge image 'E' is given by (2).

$$E = E^H \mid E^V \mid E^D \qquad (2)$$

where " |" is "logical OR" operator.

The resulting edge image "E" contains edges equivalent to text area within the input image as shown in Fig. 5.

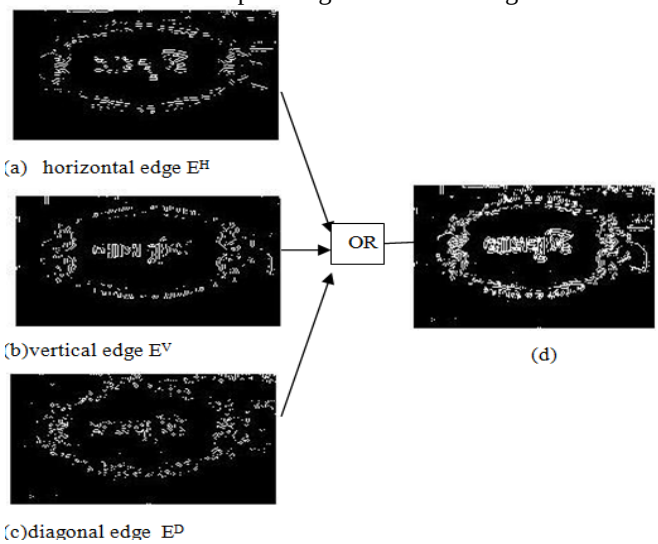


Figure 5. (a-c)After applying sobel operator on detail components and (d) edge image after fusion of E^H, E^V and E^D

D. Morphological operation

In next step, morphological dilation with square structuring element is applied on edge image E. In this way, small instructions into boundaries of a text region are filled in and it is attainable to recover the text areas that fail to be detected. The size of the structuring element for dilation process has been predetermined based on experimental evaluation.

E. Text Region Detection:

This stage localizes text regions through connected component analysis and boundary boxes generation.

1. Connected component (CC) analysis

The CCs-based routines assists text detection by dividing problem into sub tasks: (1) extract CCs from the image (2) produce text areas from the obtained CCs by investigating shape features, and (3) find candidate text area to obtain final text detection results.

To cluster along pixels that belong to the similar text area, the 8-connected component labeling is performed using procedure delineated in Haralick et al. [103]. Then components (CC) are extracted in conjunction with their bounding box (BB) information for analysis, where Bounding Box referred to smallest rectangle that precisely fits component. Several components extracted are not elements of Gujarati- texts. Thus it is necessary to filter-out non-Gujarati-text regions.

To achieve this, first, following geometric properties of each extracted components are derived:

1. Height (H)
2. Width (W)
3. Size(S)= H *W
4. Area (A): Total no. of foreground (ON) pixels
5. Aspect ratio: It is the ratio of a Width to Height.

In our approach, solely horizontal text area unit are thought of. To develop the criterion for filtering out non-Gujarati-text regions, the following geometric features of horizontal aligned Gujarati text are observed:

i. Gujarati Text always contains edges.
ii. Gujarati Text areas have widths larger than their heights.
iii. Gujarati Text is delimited in size.
iv. The pixels of a single Gujarati character have nearly uniform intensity values or colors.
v. There are no assumptions concerning what the image may contain, as an example it may have multiple lines of Gujarati text that require to be detected and isolated.

Above observation becomes robust basis for formulating rules R1 to R5 that may distinguish between obvious spurious text regions and true text regions:

R1. If the quantitative relation of size(S) to Area (A) is bigger than threshold Th1 as given in (3) then it is

thought of as non-text region. It also discards complex contour shape region.

$$S/A > Th1 \quad (3)$$

R2. If Aspect ratio is bigger than Th2, then it is discarded. This rule will result in removing most of the abnormal area size contained in image.

R3. If the height (H) of region is smaller amount than threshold Th2, then take away that region from further consideration.
This guideline is employed to get rid of the long vertical lines which may be detected in the HL components of the wavelet transform.

R4. The size of the region is another parameter that influences the results. To prune this kind of area, size of region is compared with threshold Th4 and if size of region is a smaller amount than threshold, then it is removed from further processing.

R5. If region's width is a smaller amount than Th5 then, that region are going to be discarded.

In this research, values of thresholds Th1, Th2, Th3, Th4 and TH5 are set empirically which is 2.2, 0.5,25,310 and 40 respectively.

The system applies filtration to all CCs and only those complying with all the selection rules succeed in becoming the final candidate text region. Subsequently, the original image that is associated with the BB of candidate text region is detected, on which next step will be performed.

IV. TEXT SEGMENTATION AND BINARIZATION

After detecting text, it needs to be segmented from the background. Segmentation means to separate text area from the non-text, i.e. background. In general documents, the letters are black and the background is white. However, for the images used in this research, there is no prior information. Therefore, it is necessary to isolate the Gujarati text and background in a Gujarati text area.

Using BB information, detected text is extracted from original image and is used as input data for the text binarization. Next two steps are repeated for each detected text region:

Step-1: Distinctive located text region image is transformed into a gray image. Then Otsu [104] formula is applied on gray image to calculate an optimum threshold value (T). The calculated threshold T is used to binarize the gray image using (4)

$$\text{img\_binary}(x, y) = \begin{cases} 1, \text{ if img\_gray}(x, y) \geq T \\ 0, \text{if img\_gray}(x, y) < T \end{cases} \quad (4)$$

Where img_binary(x, y) is binary image correspond to gray image img_gray(x, y).

Step-2: In this paper there is no assumption concerning text polarity (whether the text is bright on a dark background or dark on a bright background). The binarized text region obtained by step-1 represents the text as either 1 (or "White") or 0 (or "Black"), i.e. text polarity is either positive (black text on a white background) or negative (white text on a black background).

To eliminate the foreground-background ambiguity in addressing with such images, the entire area of the background and foreground is detected, and then polarity of the text is inverted, as required. If total area of the background (BG) is smaller than total area of the foreground (FG), then the binary image obtained in the process still remains; otherwise, the binarized image is inverted as in (5)

$$\text{img\_binary}(x, y) = \begin{cases} \text{img\_binary}^{-1}(x, y) & \text{if BG} \geq FG \\ \text{img\_binary}(x, y) & \text{if BG} < FG \end{cases} \quad (5)$$

In this way, the image img_binary(x, y) is made, which set intensity values of the text and background pixels, in order that text is white and background is black.

V. DATASET AND EVALUATION PROTOCOL

A. Dataset

There are numerous public datasets, which have been proved as an invaluable resource and are widely used for performance assessment and comparative study of various techniques within the domain of text detection and recognition. ICDAR robust reading competitions are held on regularly in the direction of text detection and recognition. Some of available benchmark datasets are Oriented Scene Text Database (OSTD) [20], MSRA-TD500 [44], KAIST [46], ICDAR [58, 59, 60]. Besides these, CHAR74K [95], Street View Text (SVT) dataset, NEOCR- Multilingual image dataset, SVHN (Street View House Number), and IIIT 5K-word are various additional available datasets. Most of the images of well-established datasets contain text in English and Chinese language and there is no dataset for Gujarati text.

As no standard dataset of images for Guajarati text detection is available for this work, we have build a dataset which is challenging and consist of a set of images which are collected from a variety of sources, such as book pages, magazines, newspapers, advertisements, cover pages, logo, story books of children, slogan etc. The texts may be in several fonts, sizes and colors. While making this dataset we have considered images with texture, varying background and foreground text of various fonts, sizes and colors, so that the resulting set reflects robustness of the proposed model. The dimensions of each image vary and quality of each image may differ, depending on the quality of the paper on which text appears or printed etc. These images mainly focus on horizontal text. Dataset contains total 878 images.

B. Evaluation protocol

We evaluate the performance at the block level. The detected text blocks area represented by their bounding boxes. The following parameters are outlined for each detected block:

1. False Negatives (FN): Those regions within the image which contain Gujarati text however omitted, i.e. text exists however not detected. In text, even single modifier/character missing is considered as false negative.
2. False positive (FP): A detected block that does not contain text.
3. True Positive (TP): A detected block that contains a minimum of one true Gujarati character, modifier or half character. Thus, a TP may or may not fully enclose a text line or word. Based on the dimensions of the characters within the image and spacing between the words the detected block may be composed of a single character or a complete word or a part of the word or a line. Detected block may also contain words from totally different lines if the words in several lines are connected by some background object.

The performance measures units are outlined as follows:

1. Recall (R)= TP / (TP + FN)
2. Precision = TP /(TP+FP)
3. F-measure: F-measure is weighted harmonic mean, which is a single measure of algorithm performance, combine P and R and defined by (6)

$$F = \frac{1}{\alpha \frac{1}{P} + (1-\alpha)\frac{1}{R}} \qquad (6)$$

Where α represents the relative weight between the P and R. In our analysis, the parameter α is set to 0.5, giving equal weights to precision and recall in the combined measure *f*.

VI. EXPERIMENTAL RESULT

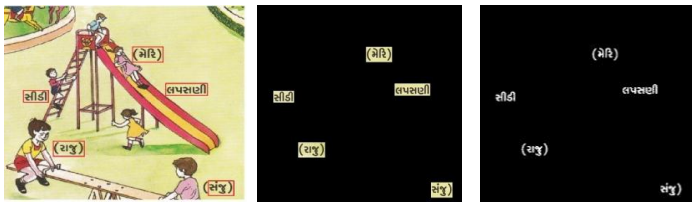
We conducted experiments on 878 images collected by us to evaluate the proposed method. The algorithm can handled several types of challenging scenario, e.g. texture, variations in text font, color, size. The final output is a binary image, where the text pixels appear in white on a black background Fig. 6 shows various steps of the proposed model.

Figure 6. CC analysis (a) connected component (b) candidate text area after filter out non-text regions (c) Detected text region marked by red rectangle(d) isolated text region (e) Binary of detected text

Fig. 7 show result before and after polarity detection step.

Figure 7. Image has text of both polarity (a) detected text (b) text binarization before text polarity step (c) text binarization after text polarity step

Some outputs of the proposed system are presented in Fig. 8.



(a) (b) (c)

Figure 8. (a) Detected text region marked with red BB (b) Isolated text region (c) Binary of text

There is total 573 upper modifier anuswar (.). Out of this 528 are correctly detected and extracted. So accuracy of anuswar detection is 92%.

The proposed system has achieved a precision of 0.92, recall of 0.97 and f-measure of 0.94 for text detection, which is pretty appreciable as this system is the first ever work on Gujarat Text Detection from image.

VII. CONCLUSION

To the best of our knowledge, this study present first ever work for Gujarati printed text detection and extraction from images. The hybrid approach is employed to effectively detect and isolate foreground text, including upper modifier unuswar (.) which is normally mistaken as noise, from complex colour images. Text is clearly extracted from its background, so output of the projected work can be used as input to Gujarati OCR or text recognition program for text recognition. This work is able to detect texts of larges variation in images. This system can be improved by applying text verification methods.

REFERENCES

[1] J. M. Patel, A. A. Desai, “Segmentation and Recognition of printed Gujarati numeral from Image”, International Journal of Engineering Research & Technology, Vol. 3, Issue. 2, pp. 1012-1020, February 2014

[2] SH Kim, KJ An, SW Jang, GY Kim, "Texture feature-based text region segmentation in social multimedia data", Multimedia Tools and Applications 75, No. 20, pp. 12815-12829, 2016

[3] Y. Zhong, H. J. Zhang, and A. K. Jain, “Automatic caption localization in compressed video”, IEEE transactions on pattern analysis and machine intelligence, Vol. 22, No. 4, pp. 385–392, Apr. 2000.

[4] L. Neumann , J. Matas, “Scene text localization and recognition with oriented stroke detection”, In Proceedings of the IEEE International Conference on Computer Vision, pp. 97–104, 2013

[5] I. Haritaoglu, “Scene text extraction and translation for handheld devices”, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. 408–413, 2001

[6] J. Liang, D. Doermann, H. Li, “Camera-based analysis of text and documents: A survey”, International Journal of Document Analysis and Recognition, Vol. 7, Issue. 2-3, pp. 84–104, 2005.

[7] L. Lin , C. L. Tan, “Text extraction from name cards using neural network”, Proceedings. IEEE International Joint Conference on, Vol. 3, pp. 1818–1823, 2005.

[8] P Sermanet, S Chintala, Y LeCun, "Convolutional neural networks applied to house numbers digit classification", In the Proceedings of international conference on pattern recognition ICPR, pp 10–13, 2012

[9] Lee, Chung-Mong, A Kankanhalli. "Automatic extraction of characters in complex scene images", International Journal of Pattern Recognition and Artificial Intelligence, Vol. 9, No. 01, pp. 67-82, 1995

[10] X Liu, J Samarabandu, "An edge-based text region extraction algorithm for indoor mobile robot navigation", In Mechatronics and Automation, IEEE International Conference, Vol. 2, No. 7, pp. 2008–2015, 2005

[11] G. S. Hsu, J. C. Chen , Y. Z. Chung, "Application-Oriented License Plate Recognition", in IEEE Transactions on Vehicular Technology, Vol. 62, No. 2, pp. 552-561, Feb. 2013

[12] P. Prabhakar , P. Anupama, "A novel design for vehicle license plate detection and recognition", Second International Conference on Current Trends In Engineering and Technology - ICCTET, pp. 7-12, 2014

[13] Liu, Xiaoqian, Ke Lu, ,W. Wang, "Effectively localize text in natural scene images", In Pattern Recognition (ICPR), 21st International Conference on, pp. 1197-1200, 2012.

[14] N. Ezaki, M. Bulacu, L. Schomaker, "Text detection from natural scene images: towards a system for visually impaired persons", Proceedings of the 17th International Conference on Pattern Recognition, ICPR, Cambridge, Vol. 2, pp. 683-686, 2004

[15] X. Chen , A. L. Yuille, “Detecting and reading text in natural scenes”, Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. 366–373, 2004

[16] B. T. Nassu, R. Minetto, L. E. S. d. Oliveira, “Text line detection in document images: Towards a support system for the blind”, 12th International Conference on Document Analysis and Recognition, pp. 638–642, 2013

[17] C. N. Patel, A.A. Desai, "Segmentation of text lines into words for Gujarati handwritten text", In Proceedings of international conference on signal and image processing (ICSIP’10), IEEEXplore, 15–17, 2010

[18] J Dholakia, A Negi, M. S. Rama, "Zone identification in the printed Gujarati text", In the Proceedings of the eight international conference on document analysis and recognition (ICDAR’05), 2005.

[19] L Neumann, J Matas, "Real-time scene text localization and recognition", In the Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 3538–3545, 2012.

[20] C. Yi, Y L. Tian , "Text string detection from natural scenes by structurebased partition and grouping", IEEE Transactions on Image Processing, Vol. 20, No. 9, pp. 2594–2605, 2011.

[21] X. Yin, X. Yin, K. Huang , H. Hao, "Robust Text Detection in Natural Scene Images", in IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 36, No. 5, pp. 970-983, May 2014.

[22] G. K. Myers, R. C. Bolles, Q. Luong, J. A. Herson, H. B. Aradhye, "Rectification and recognition of text in 3-d scenes", International Journal on Document Analysis and Recognition Vol. 7, No. 2, pp. 147-158, 2005.

[23] C. Yi , Y. Tian, ”Localizing text in scene images by boundary clustering, stroke segmentation and string fragment classification”, IEEE Transaction on Image Process., Vol. 21, No. 9, pp. 4256-4268, Sep. 2012.

[24] Huang, Weilin, Z. Lin, J. Yang, , J. Wang. "Text localization in natural images using stroke feature transform and text covariance

descriptors", In Proceedings of the IEEE International Conference on Computer Vision, pp. 1241-1248, 2013.
[25] H. Li , D. Doermann, "Text enhancement in digital video using multiple frame integration", In Proceedings of the seventh ACM international conference on Multimedia (Part 1), pp. 19–22, 1999.
[26] Ali, Suhad A., Ashwaq T. Hashim. "Wavelet transform based technique for text image localization", Karbala International Journal of Modern Science Vol. 2, No. B, pp. 138-144., 2016.
[27] M. Kumar, G. Lee, "Automatic text location from complex natural scene images", International conference on computer and automation engineering (ICCAE), Vol. 3, pp. 594–597, 2010.
[28] S. Lu, T. Chen, S. Tian, Lim J-H, Tan C-L, "Scene text extraction based on edges and support vector regression", International Journal on Document Analysis and Recognition (IJDAR) Vol. 18, Issue. 2, pp 125–135, 2015.
[29] A. Zhu, G. Wang, Y. Dong, "Detecting natural scenes text via auto image partition, two-stage grouping and two-layer classification", Pattern Recognition Letters, Vol. 67, pp. 153–162 ,2015.
[30] C Yu, Y Song, Y Zhang, "Scene text localization using edge analysis and feature pool", Neurocomputing Vol. 175, pp. 652–661, 2016.
[31] W Huang, Y Qiao, X Tang, "Robust scene text detection with convolution neural network induced Mser trees", In Proceedings of European Conference on Computer Vision, Springer, pp. 497–511, 2014
[32] Y. F. Pan, X. Hou, C. L. Liu, "A hybrid approach to detect and localize texts in natural scene images", IEEE Transaction on Image Processing, Vol. 20, No. 3, pp. 800–813, Mar. 2011.
[33] Y Liu, S Goto, T Ikenaga, "A contour-based robust algorithm for text detection in color", IEICE TRANSACTIONS on Information and Systems, Vol. 89, No. 3, pp. 1221–1230, 2006.
[34] Y. Pan, X. Hou , C. Liu, "Text Localization in Natural Scene Images Based on Conditional Random Field", 10th International Conference on Document Analysis and Recognition, Barcelona, pp. 6-10, 2009.
[35] M. Petter, V. Fragoso, M. Turk, C. Baur, "Automatic text detection for mobile augmented reality translation", IEEE international conference on computer vision workshops (ICCV), pp 48–55, 2011.
[36] Y. Pan, X. Hou , C. Liu, "A Hybrid Approach to Detect and Localize Texts in Natural Scene Images", in IEEE Transactions on Image Processing, Vol. 20, No. 3, pp. 800-813, March 2011.
[37] C. Shi, B. Xiao, C. Wang and Y. Zhang, "Graph-Based Background Suppression for Scene Text Detection," 10th IAPR International Workshop on Document Analysis Systems, pp. 210-214, 2012.
[38] P. Shivakumara, W. Huang, T. Phan, and C. Tan, "Accurate video text detection through classification of low and high contrast images", Pattern Recognition, Vol. 43, No. 6, pp. 2165–2185, 2010.
[39] J. Lee, P. Lee, S. Lee, A. Yuille, and C. Koch, "AdaBoost for text detection in natural scene", in Proceeding of IEEE International Conference on Document Analysis and Recognition, pp. 429–434, 2011.
[40] X. Chen , A. L. Yuille, "Detecting and reading text in natural scenes", Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 366–373., 2004.
[41] X. Yin, X.-C. Yin, H.-W. Hao, and K. Iqbal, "Effective text localization in natural scene images with MSER, geometry-based grouping and AdaBoost", Proceedings of the 21st International Conference on Pattern Recognition, pp. 725–728, Nov. 2012.
[42] C. Liu, C. Yang, X. Q. Ding, and K. Wang, "An improved scene text extraction method using conditional random field and optical character recognition", International Conference on Document Analysis and Recognition, pp. 708–712, 2011.
[43] Shi, Baoguang, Xiang Bai, and Cong Yao. "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition", IEEE transactions on pattern analysis and machine intelligence, Vol. 39, No. 11, pp. 2298-2304, 2017.
[44] C. Yao, X. Zhang, X. Bai, W. Liu, Y. Ma, and Z. Tu, "Detecting texts of arbitrary orientations in natural images", IEEE Conference on Computer Vision and Pattern Recognition, pp. 1083–1090, 2012.
[45] A. R. Chowdhury, U. Bhattacharya, and S. K. Parui, "Scene text detection using sparse stroke information and MLP", in Proceeding of IEEE International Conference on Pattern Recognition., pp. 294–297, 2012.
[46] S. Lee, M. Cho, K. Jung, J. Kim, "Scene text extraction with edge constraint and text collinearity", International Conference on Pattern Recognition, IEEE, pp. 3983–3986, 2010.
[47] K. Esther Amulya, P. Sanoop Kumar, "Scene Text Extraction using Stroke Width Transform", International Journal of Computer Sciences and Engineering, Vol. 6, Issue. 6, pp.375-379, 2018.
[48] C. Yan, H. Xie, S. Liu, J. Yin, Y. Zhang and Q. Dai, "Effective Uyghur Language Text Detection in Complex Background Images for Traffic Prompt Identification", in IEEE Transactions on Intelligent Transportation Systems, Vol. 19, No. 1, pp. 220-229, Jan. 2018.
[49] Dey, Soma, Rajat Subhra Goswami. "A morphological segmentation and curve-let features extraction on text region classification using SVM", In Computational Intelligence and Computing Research (ICCIC), IEEE International Conference on, pp. 1-6, 2015.
[50] C Shi, C Wang, B Xiao, Y Zhang, S Gao,"Scene text detection using graph model built upon maximally stable extremal regions", Pattern Recognition Letter Vol. 34, Issue. 2, pp.107–116, 2013.
[51] C. Shi, C. Wang, B. Xiao, Y. Zhang, S. Gao, and Z. Zhang, "Scene text recognition using part-based tree-structured character detections", in Proceeding IEEE Conference on Computer Vision and Pattern Recognition, pp. 2961–2968, 2013.
[52] H. Koo, D. H. Kim, "Scene text detection via connected component clustering and non-text filtering", IEEE Transaction Image Processing, Vol. 22, No. 6, pp. 2296–2305, Jun. 2013.
[53] L. Neumann , J. Matas,"Real-time scene text location and recognition", IEEE Conference on Computer Vision and Pattern Recognition, pp. 3538–3545, 2012.
[54] L. Neumann, J. Matas," Text localization in real-world images using efficiently pruned exhaustive search", In Proceedings of the international conference on document analysis and recognition (ICDAR), pp 687–691, 2011.
[55] H. Chen, S. S. Tsai, G. Schroth, D. M. Chen, R. Grzeszczuk and B. Girod, "Robust text detection in natural images with edge-enhanced Maximally Stable Extremal Regions", 18th IEEE International Conference on Image Processing, pp. 2609–2612, 2011.
[56] D. Karatzas, S. R. Mestre, J. Mas, F. Nourbakhsh and P. P. Roy, "ICDAR 2011 Robust Reading Competition - Challenge 1: Reading Text in Born-Digital Images (Web and Email)", International Conference on Document Analysis and Recognition, Beijing, pp. 1485-1490. IEEE, 2011.
[57] A. Shahab, F. Shafait , A. Dengel, "ICDAR 2011 Robust Reading Competition Challenge 2: Reading Text in Scene Images", International Conference on Document Analysis and Recognition, pp 1491–1496, 2011.
[58] D. Karatzas et al., "ICDAR 2015 competition on Robust Reading", 13th International Conference on Document Analysis and Recognition (ICDAR), Tunis, pp. 1156-1160, 2015.

[59] D. Karatzas et. al., "ICDAR 2013 robust reading competition", In Document Analysis and Recognition (ICDAR), 12th International Conference on, pp. 1484-1493. IEEE, 2013.
[60] Lucas et al., "ICDAR 2003 robust reading competitions", In Document Analysis and Recognition, Proceedings. Seventh International Conference on, pp. 682-687. IEEE, 2003.
[61] Kumar, Deepak, M. N. Prasad, A. G. Ramakrishnan. "Multi-script robust reading competition in icdar 2013", In Proceedings of the 4th International Workshop on Multilingual OCR, p. 14. ACM, 2013.
[62] X. Jiang, J. Lian, Z. Xia, X. Feng, A. Hadid, "Fast Chinese character detection from complex scenes", Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA), Oulu, pp. 1-4, 2016.
[63] X. Ren, K. Chen, X. Yang, Y. Zhou, J. He and J. Sun, "A new unsupervised convolutional neural network model for Chinese scene text detection", IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP), Chengdu, pp. 428-432, 2015.
[64] Y Wang, L Peng, S Wang, "A Multi-stage Method for Chinese Text Detection in News Videos", Procedia Computer Science, Vol. 96, pp. 1409-1417, 2016.
[65] O. Zayene, S. M. Touj, J. Hennebert, R. Ingold and N. E. Ben Amara, "Data, protocol and algorithms for performance evaluation of text detection in Arabic news video", In Advanced Technologies for Signal and Image Processing (ATSIP), 2nd International Conference on, pp. 258-263. IEEE, 2016.
[66] X. Chen, J. Yang, J. Zhang, A. Waibel, "Automatic detection and recognition of signs from natural scenes", IEEE Trans. Image Process., Vol. 13, No. 1, pp. 87–99, Jan. 2004.
[67] M. Iwamura, T. Kobayashi, and K. Kise, "Recognition of multiple characters in a scene image using arrangement of local features", in Proc. IEEE International Conference on Document Analysis and Recognition, pp. 1409–1413, 2011.
[68] M. Moradi, S. Mozaffari and A. A. Orouji, "Farsi/Arabic text extraction from video images by corner detection", 6th Iranian Conference on Machine Vision and Image Processing, Isfahan, pp. 1–6, 2010.
[69] Ghoshal, Ranjit, A. Roy, S. Parui. "Recognition of Bangla text from scene images through perspective correction", In Image Information Processing (ICIIP), 2011 International Conference on, pp. 1-6. IEEE, 2011.
[70] H Gaddour, S Kanoun, N Vincent, "A New Method for Arabic Text Detection in Natural Scene Image Based on the Color Homogeneity", In International Conference on Image and Signal Processing, Springer International Publishing pp. 127-136., 2016.
[71] O. Zayene, S. M. Touj, J. Hennebert, R. Ingold and N. E. Ben Amara, "Data, protocol and algorithms for performance evaluation of text detection in Arabic news video", 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Monastir, pp. 258-263, 2016.
[72] L. Xu, H. Nagayoshi, H. Sako, "Kanji character detection from complex real scene images based on character properties", In Document Analysis Systems, 2008. DAS'08. The Eighth IAPR International Workshop on, IEEE, pp. 278–285, 2008.
[73] J. Park, G. Lee, E. Kim, J. Lim, S. Kim, H. Yang, M. Lee, and S. Hwang, "Automatic detection and recognition of Korean text in outdoor signboard images", Pattern Recognition Letter., Vol. 31, No. 12, pp. 1728–1739, 2010.
[74] SH Lee, MS Cho, K Jung, JH Kim, "Scene text extraction with edge constraint and text collinearity link", In Proceedings of International Conference on Pattern Recognition. , pp. 3983–3986, 2010.
[75] A. Jamil, I. Siddiqi, F. Arif, A. Raza, "Edge-based features for localization of artificial Urdu text in video images", in Proceeding IEEE International Conference In Document Analysis and Recognition, pp. 1120–1124, 2011.
[76] Y. Song et. al., "Robust and parallel Uyghur text localization in complex background images", Machine Vision and Applications, Vol. 28, Issue. 7, pp. 755-769, 2017.
[77] S Fang, H Xie, Z Chen, S Zhu, X Gu, X Gao, "Detecting Uyghur text in complex background images with convolutional neural network", Multimedia Tools and Applications Vol.76, No. 13 ,pp. 15083-15103, 2017.
[78] C. Yao, X. Bai, W. Liu, "A Unified Framework for Multioriented Text Detection and Recognition", in IEEE Transactions on Image Processing, Vol. 23, No. 11, pp. 4737-4749, Nov. 2014.
[79] P. Shivakumara, T. Q. Phan, C. L. Tan, "A Laplacian approach to multi-oriented text detection in video", IEEE transactions on pattern analysis and machine intelligence, Vol. 33, No. 2, pp. 412–419, Feb. 2011.
[80] L Li, S Yu, L Zhong, X Li, "Multilingual text detection with nonlinear neural network", Mathematical Problems in Engineering 2015
[81] M. R. Lyu, J. Song, M. Cai, "A comprehensive method for multilingual video text detection, localization, and extraction", IEEE transactions on circuits and systems for video technology, Vol. 15, No. 2, pp. 243– 255, Feb. 2005.
[82] G. Zhou, Y. Liu, Q. Meng, and Y. Zhang, "Detecting multilingual text in natural scene", 1st International Symposium on Access Spaces (ISAS), Yokohama,, pp. 116–120, 2011.
[83] V.N. Manjunath Aradhya, M.S. Pavithra, "A comprehensive of transforms, Gabor filter and k-means clustering for text detection in images and video", Applied Computing and Informatics, Vol. 12, Issue. 2, Pages 109-116, 2016.
[84] L. Gomez, D. Karatzas,"Multi-script text extraction from natural scenes", in Proceeding IEEE International Conference in Document Analysis and Recognition, pp. 467–471, 2013.
[85] Gomez, L. & Karatzas, "A Fast Hierarchical Method for Multi-script and Arbitrary Oriented Scene Text Extraction", IJDAR, Vol. 19, Issue. 4, pp 335–349, December 2016.
[86] X. Liu, H. Fu, and Y. Jia, "Gaussian Mixture modeling and learning of neighboring characters for multilingual text extraction in images", Pattern Recognition, Vol. 41, No. 2, pp. 484–493, 2008.
[87] A. Negi, N. Kasinadhuni, "Localization and extraction of text in Telugu document images", TENCON. Conference on Convergent Technologies for Asia-Pacific Region, Vol. 2, pp. 749-752, 2003.
[88] A Negi, KN Shanker, CK Chereddi, "Localization, extraction and recognition of text in Telugu document images", In Document Analysis and Recognition, Proceedings. Seventh International Conference on, pp. 1193-1197. IEEE, 2003.
[89] U. Bhattacharya, S. K. Parui and S. Mondal, "Devanagari and Bangla Text Extraction from Natural Scene Images", 10th International Conference on Document Analysis and Recognition, Barcelona, pp. 171-175, 2009.
[90] "D.V. Sharma, S. Singh "Gurmukhi text detection and localization in natural scene images", International Journal of Advanced Research in Computer Science and Software Engineering Vol. 5, Issue 12, pp. 442–458, 426-428, 2015
[91] S. Kaur, G.S Josan, "Gurmukhi Text Extraction From Image Using Support Vector Machine (SVM)", International Journal of Engineering Research and Applications (IJERA), Vol. 3, No. 4, pp. 9-15, 2011.
[92] R Ghoshal, A Roy, TK Bhowmik, SK Parui, "Decision tree based recognition of Bangla text from outdoor scene images", In International Conference on Neural Information Processing, pp. 538-546. Springer, Berlin, Heidelberg, 2011.
[93] S. V. Seeri, S. Giraddi and B. M. Prashant, "A novel approach for Kannada text extraction", International Conference on Pattern

Recognition, Informatics and Medical Engineering (PRIME-2012), Salem, Tamilnadu, pp. 444-448, 2012.
[94] S. A.Angadi, M. M. Kodabagi. "A robust segmentation technique for line, word and character extraction from Kannada text in low resolution display board images", In Signal and Image Processing (ICSIP), Fifth International Conference on, pp. 42-49. IEEE, 2014.
[95] T. E. de Campos, B. R. Babu, M. Varma, "Character recognition in natural images", VISAPP Vol. 2, Issue 7, 2009.
[96] A. A. Desai, "Support vector machine for identification of handwritten Gujarati alphabets using hybrid feature space", CSI transactions on ICT, Vol. 2, No. 4, pp. 235-241, 2015.
[97] A. A. Desai, "Gujarati handwritten numeral optical character recognition through neural network", Pattern Recognition, Vol. 43, Issue. 7, pp. 2582–2589, 2010.
[98] A. A. Desai, "Handwritten Gujarati numeral optical character recognition using Hybrid feature extraction technique", In Proceeding of international conference on image processing, Computer vision and pattern recognition, (IPCV'10), Vol. 2, pp. 733–739, 2010.
[99] C. N. Patel, A. A. Desai, "Zone identification for Gujarati handwritten words", In Proceedings of international conference on emerging applications of information technology (EAIT 2011), IEEEXplore, pp. 19–20, 2011.
[100] J. Gllavata, R. Ewerth, B. Freisleben, "Text detection in images based on unsupervised classification of high-frequency wavelet coefficients", in Proceeding IEEE International Conference Pattern Recognition, pp. 425–428, 2004.
[101] Q. Ye, Q. Huang, W. Gao and D. Zhao, "Fast and robust text detection in images and video frames", Image and Visison Computing, Vol. 23, pp. 565–576, 2005.
[102] Sobel, I., "An Isotropic 3×3 Gradient Operator", Machine Vision for Three – Dimensional Scenes, Freeman, H., Academic Pres, NY, pp. 376-379, 1990.
[103] Haralick, Robert, M., Linda, G. S., "Computer and Robot Vision", Addison-Wesley Longman: Boston, MA, USA, pp. 28–48, 1992.
[104] N. Otsu, "A Threshold Selection Method from GreyLevel Histogram", IEEE transactions on systems, man, and cybernetics, Vol. 9, No. 1, pp. 62-66, 1979.

Authors Profile

Jagin M. Patel received MCA and M.PhiL degree from Veer Narmad South Gujarat University. He is currently pursuing Ph.D. and currently working as Assistant Professor in M. K. Institute of computer studies, Bharuch, Gujarat, India.. He has more than 15 years of teaching experience. His fields of interest are Image Processing and Optical Character Recognition.

Apurva A. Desai is the Dean of faculty of Computer Science and Information Technology of Veer Narmad South Gujarat University, Surat, Gujarat, India. He is HOD and professor in department of computer science of VNSGU. He earned his Ph.D. in the year 1997 in the joint fields of Operation Research and Computer Science. He is Ph.D. and M.Phil supervisor for many years and many students have completed their research under his guidance. He has delivered many lectures as resource person in various events and also given invited talks in national and international events. Dr. Desai has more than 50 research papers and four books to his credit. He is an Editor in Chief of VNSGU Journal of Science and Technology and also serving as a member of Editorial board for some of the national and international journals. His fields of interest are Optical Character Recognition, Image Processing, Computer Graphics, Data Mining and Soft Computing.

.

