

# A SURVEY ON VARIOUS APPROACHES OF TEXT EXTRACTION IN IMAGES

C.P. Sumathi<sup>1</sup>, T. Santhanam<sup>2</sup> and G.Gayathri Devi<sup>3</sup>

<sup>1</sup> Department of Computer Science, SDNB Vaishnav College for Women, Chennai, India

santsum@hotmail.com

<sup>2</sup> Department of Computer Application, DG Vaishnav College for Men, Chennai, India

santhanam\_dgvc@yahoo.com

<sup>3</sup> Department of Computer Science, SDNB Vaishnav College for Women, Chennai, India

mail2gg@yahoo.co.in

## ABSTRACT

*Text Extraction plays a major role in finding vital and valuable information. Text extraction involves detection, localization, tracking, binarization, extraction, enhancement and recognition of the text from the given image. These text characters are difficult to be detected and recognized due to their deviation of size, font, style, orientation, alignment, contrast, complex colored, textured background. Due to rapid growth of available multimedia documents and growing requirement for information, identification, indexing and retrieval, many researches have been done on text extraction in images. Several techniques have been developed for extracting the text from an image. The proposed methods were based on morphological operators, wavelet transform, artificial neural network, skeletonization operation, edge detection algorithm, histogram technique etc. All these techniques have their benefits and restrictions. This article discusses various schemes proposed earlier for extracting the text from an image. This paper also provides the performance comparison of several existing methods proposed by researchers in extracting the text from an image.*

## KEYWORDS

*Text Extraction, Document Text Images, Caption Text Images, Scene Text, Heterogeneous Images.*

## 1. INTRODUCTION

Text Extraction from image is concerned with extracting the relevant text data from a collection of images. Rapid development of digital technology has resulted in digitization of all categories of materials. Lot of resources are available in electronic medium. Many existing paper-based collections, historical manuscripts, records, books, journals, scanned document, book covers, video images, maps, manuscripts, pamphlets, posters, broadsides, newspapers, micro facsimile, microfilms, university archives, slides and films, book plates, pictures, painting, graphic materials, coins and currency, stamps, magazines, clipping files, educational, TV programs, business card, magazines, advertisements, web pages, mixed text-picture-graphics regions etc are converted to images. These images present many challenging research issues in text extraction and recognition. Text extraction from images have many useful applications in document analysis, detection of vehicle license plate, analysis of article with tables, maps, charts, diagrams etc., keyword based image search, identification of parts in industrial automation, content based retrieval, name plates, object identification, street signs, text based video indexing, video content analysis, page segmentation, document retrieving, address block location etc.

Images can be broadly classified into Document images, Caption text images and Scene text images. Figures 1-3 show some examples of text in images. A document image (Figure 1a, 1b) usually contains text and few graphics components. Document images are acquired by scanning journal, printed document, degraded document images, handwritten historical document, and book cover etc. The text may appear in a virtually unlimited number of fonts, style, alignment, size, shapes, colors, etc. Extraction of text in documents with text on complex color background is difficult due to complexity of the background and mix up of color(s) of fore-ground text with colors of background.

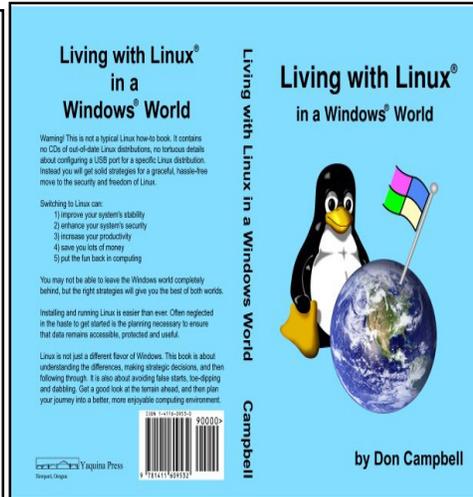
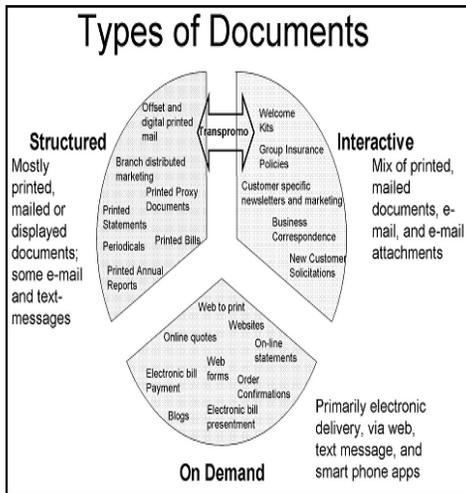


Figure 1a: Document Text Image  
(courierexpressandpostal.blogspot.com)

Figure 1b: Colored Text Image  
(athleticaid.com)

Caption text is also known as Overlay text or Cut line text. Caption text (Figure 2) is artificially superimposed on the video/image at the time of editing and it usually describes or identifies the subject of the image/video content. The superimposed text is a powerful source of high-level semantics. These text occurrences could be detected, segmented, and recognized automatically for indexing, retrieval and summarization. The extraction of the superimposed text in sports video is very useful for the creation of sports summary, highlights etc. These types of caption text include moving text, rotating text, growing text, shrinking text, text of arbitrary orientation, and text of arbitrary size.

Scene text (Figure 3) appears within the scene which is then captured by the recording device i.e. text which is present in the scene when the image or video is shot. Scene texts occurs naturally as a part of the scene and contain important semantic information such as advertisements that include artistic fonts, names of streets, institutes, shops, road signs, traffic information, board signs, nameplates, food containers, cloth, street signs, bill boards, banners, and text on vehicle etc. Scene text extraction can be used in detecting text-based landmarks, vehicle license detection/recognition, and object identification rather than general indexing and retrieval. It is difficult to detect and extract since it may appear in a virtually unlimited number of poses, size, shapes and colors, low resolution, complex background, non-uniform lighting or blurring effects of varying lighting, complex movement and transformation, unknown layout, uneven lighting, shadowing and variation in font style, size, orientation, alignment & complexity of background.



Figure 2: Caption Text Image  
(googlecode.blogspot.com)



Figure 3: Scene Text Image  
(ICDAR Dataset)

Due to very fast growth of available multimedia documents and growing requirement, studies in the field of pattern recognition shows a great amount of interest in efficient extraction of text, indexing and retrieval from digital video/document images. Intensive research projects are performed for text extraction in images by many scholars. Text extraction involves detection, localization, tracking, binarization, extraction, enhancement and recognition of the text from the given image. Several techniques have been developed for extracting the text from an image. The proposed methods were based on morphological operators, wavelet transform, artificial neural network, skeletonization operation, edge detection algorithm, histogram technique etc. The methods cited in this paper on text extraction in images are classified according to different types of images.

## **2. REVIEW**

A large number of approaches have been proposed for text extraction from images. The existing work on text extraction from images can be classified according to different criteria. This article classifies according to the different types of image, analyzes those algorithms and discusses the performance evaluation. The performance measure is presented in Table 1,2,3,4. The purpose of the survey is devoted to a remarkable growth of text extraction techniques.

### **2.1 Text Extraction Work**

Various methods are used for the extraction of text from colored journal images, camera captured images, video images, printed document, degraded document images, handwritten historical document, graphical and color document images, low resolution images, book cover and web pages.

#### **2.1.1 Document Text Images**

A robust approach to segment text from color images was put forth by Y. Zhan et.al [1]. The proposed algorithm uses the multiscale wavelet features and the structural information to locate candidate text lines. Then a SVM classifier was used to identify true text from the candidate text lines. This approach mainly included four stages. In preprocessing step text blocks were enhanced by using cubic interpolation to rescale the input text blocks and a Gaussian filter to

smooth the text blocks and remove noises. These image blocks were split into connected components and non-text connected components were eliminated by a component filtering procedure. The left connected components were merged using K-means clustering algorithm into several text layers, and a set of appropriate constraints were applied to find the real text layer. Finally, the text layer was refined through a post-processing step.

Thai et.al [2] described an approach for effective text extraction from graphical document images. The algorithm used Morphological Component Analysis (MCA) algorithm, an advancement of sparse representation framework with two appropriately chosen discriminative over complete dictionaries. Two discriminative dictionaries were based on undecimated wavelet transform and curvelet transform. This method overcame the problem of touching between text and graphics and also insensitive to different font styles, sizes, and orientations.

S.Audithan et.al [3] formulated an efficient and computationally fast method to extract text regions from documents. They proposed Haar discrete wavelet transform to detect edges of candidate text regions. Non-text edges were removed using thresholding technique. They used morphological dilation operator to connect the isolated candidate text edge and then a line feature vector graph was generated based on the edge map. This method exploited an improved canny edge detector to detect text pixels. The stroke information was extracted the spatial distribution of edge pixels. Finally text regions were generated and filtered according to line features.

Grover et.al [4] described an approach to detect text from documents in which text was embedded in complex colored document images. They proposed a simple edge based feature to perform this task. The image was converted to gray scale by forming a weighted sum of the R, G, and B components. Then edge detection was performed on the gray-scale image by convolving the image with Sobel masks, separately for horizontal and vertical edges. Convolution was followed by elimination of non-maxima and thresholding of weak edges. Next, the edge image was divided into small non overlapping blocks of  $m \times m$  pixels, where  $m$  depends on the image resolution. They performed block classification using pre-defined threshold which would differentiate the text from the image.

P. Nagabhushan et.al [5] proposed a novel approach to extract the text in complex background color document images. The proposed method used canny edge detector to detect edges. When dilation operation was performed on edge image, it created holes in most of the connected components that corresponds to character strings. Connected components without hole(s) were eliminated. Other non-text components were eliminated by computing and analyzing the standard deviation of each connected component. An unsupervised local thresholding was devised to perform fore-ground segmentation in detected text regions. Finally the noisy text regions were identified and reprocessed to further enhance the quality of retrieved foreground.

A robust and efficient algorithm to automatic text extraction from colored book and journal cover sheets was proposed by Davodet.al[6] based on wavelet transform. A dynamic threshold was used to detect edges from detail wavelet coefficient. Further effective edges were obtained by blurring approximate coefficients with alternative heuristic thresholding. Region of Interest (ROI) technique was applied and finally text was extracted. They evaluated the performance of their algorithm on 80 pictures collected from internet.

Another algorithm for Automatic text location and identification on colored book and journal covers was proposed by Karin et.al[7].The number of colors was reduced by applying a clustering algorithm. Text candidates were located using a top-down analysis based on successive splitting in horizontal and vertical direction. A bottom-up analysis detected homogeneous regions using a region growing method; grouping step was applied to find subsets of region. Finally text regions and non-text regions were distinguished.

Zhixin Shi et.al[8]proposed an extraction algorithm based on connectivity features for a complex handwritten historical document. This paper presented an algorithm using adaptive local connectivity map (ALCM) technique. Thresholding the gray scale image discloses clear text-line patterns as connected components. Grouping algorithm was used to group the connected components into location masks for each text line. Text line was extracted by mapping the location masks back onto the binary image to collect the text line components. Splitting algorithm overcame the problem of components touching multiple lines. This method dealt with fluctuating or skewed text lines and used for other types of images such as binary images, machine printed or even mixed script.

Syed Saqibet.al[9] described an approach for curled textline information extraction from grayscale camera-captured document images. The grayscale textline was enhanced by using multi-oriented multi-scale anisotropic Gaussian smoothing. Detection of central lines of curled textlines was found using ridges. This approach was based on differential geometry, which used local direction of gradients and second derivatives as the measure of curvature. Hessian matrix was used for finding direction of gradients and derivatives. By using this information, ridges were detected by finding the zero-crossing of the appropriate directional derivatives of smoothed image. Modified coupled snakes model was used for estimating x-line and baseline pairs from detected textlines. Their approach was robust against high degrees of curl and requires no post-processing.

Wafa et.al[10] suggested a new enhanced text extraction algorithm from degraded document images of both color and grayscale on the basis of the probabilistic models. Color document image was converted to YIQ colors space image and operate on Y luminance channel. Initial estimates and their corresponding mean and standard deviation vectors for expectation-maximization (EM) algorithm were calculated using k-means clustering method. The EM algorithm was used to estimate and improve the parameters of the mixtures of densities recursively. The maximum likelihood (ML) segmentation method estimates the probability that a pixel belongs to text or background.

Table 1 : Performance Analysis of Text Extraction in Document Text Images

<b>S. No</b>	<b>Author</b>	<b>Year</b>	<b>Method used</b>	<b>Accuracy</b>	<b>Benefits</b>
1	Davod et.al [6]	2011	Wavelet transform , Region of Interest (ROI)	91.20%	Robust to noise
2	Thai et.al[2]	2010	Sparse representation framework, Morphological Component Analysis (MCA) algorithm, Curvelet transform	94.76%	Overcomes the touching problem between text and graphics
3	Nagabhushan et.al [5]	2010	Canny edge detector, Dilation operation, Unsupervised local thresholding, Connected component analysis	97.12%	Handles degradations such as blur, wavy text

4	Grover et.al [4]	2009	Sobel edge detection, Thresholding technique, Block classification	99%	insensitive to color, fonts, language.
5	Syed et.al [9]	2009	Anisotropic Gaussian smoothing, Ridges based on differential geometry, Hessian matrix, Modified coupled snakes model	91%	Robust to high degrees of curl, needs no post-processing
6	Wafa et. al. [10]	2009	Expectation-maximization (EM) algorithm, Maximum likelihood (ML) segmentation method, k-means clustering method	96%	Works well on old degraded documents
7	S.Audithan et.al [3]	2009	Haar discrete wavelet transform, Morphological dilation operators, Canny edge detector	94.80%	Independent of contrast
8	Zhan et.al. [1]	2006	Multiscale wavelet features, SVM classifier, Cubic interpolation, Gaussian filter, K-means clustering algorithm	84.3%	Robust to text color, font size, languages
9	Zhixin et.al [8]	2005	Adaptive local connectivity map, Grouping algorithm, Splitting algorithm	95%	Handles mixed script.
10	Karin et.al [7]	1999	Clustering algorithm, Bottom-up and Top-down analysis	Promising results	Handles Web/video images

### 2.1.2 Scene Text Images

Angadi et.al [11] proposed a methodology to detect and extract text regions from low resolution natural scene images. Their proposed work used Discrete Cosine Transform (DCT) based high pass filter to remove and suppress the constant background. The texture feature matrix was computed on every 50x50 block of the processed image. A newly defined discriminant function was used to classify text blocks. The detected text blocks were merged to obtain new text regions. Finally, the refinement phase was a post processing step used to improve the detection accuracy. This phase used to cover small portions of missed text present in adjacent undetected blocks and unprocessed regions. The proposed methodology had been conducted on 100 indoor and outdoor low resolution natural scene images containing text of different size, font, and alignment with complex backgrounds containing Kannada text and English text. The approach

also detected nonlinear text regions and can be extended for text extraction from the images of other languages with little modifications.

Pan et.al[12] proposed a novel hybrid method where in a text region detector was designed to generate a text confidence map. A Local binarization approach was used to segment the text components using text confidence map. A Conditional Random Field (CRF) model was used to label components as text or non-text which was solved by minimum classification error (MCE) learning and graph cuts inference algorithm. A learning based method by building neighbouring components into minimum spanning tree (MST) and cutting off interline edge with an energy minimization model to group the text components into text lines.

Fabrizio et.al[13] offered a region based approach that starts by isolating letters, then groups them to restore words. The process was based on a new segmentation method based on morphological operator called Toggle Mapping Morphological Segmentation (TMMS) and a classification step based on a combination of multiple SVM classifiers. The training data base composed of 32400 examples extracted from various urban images and different configurations of classifiers have been tested to get the highest classification accuracy.

Kohei et.al [14] introduced a new approach to detect and extract text from commercial screenshot images. Their approach implemented edge-based method and connected component labeling method known as blob extraction method. Combination of homogeneity edge detection filter and appropriate threshold number separated the text from the image.

A method for localizing text regions within scene images was introduced by Luz et.al. [15]. A set of potential text regions was extracted from the input image using morphological filters. Connected Components (CC) were identified using ultimate attribute openings and closings, and selected a subset of text region after combining some of the CCs. Decision tree classifier were used to distinguish text or non-text regions.

Shivakumara et.al[16] proposed a new method based on Maximum Color Difference (MCD) and Boundary Growing Method (BGM) for detection of multioriented handwritten scene text from video. Average of RGB channel was calculated of the original frame to sharpen the text edges and increase the contrast of text pixels. Maximum Color Difference was computed to increase the gap between text and non-text pixel. Text clusters were obtained by K-means clustering algorithm. These clusters were used to obtain the text candidates and also help in eliminating false positives. To fix boundary for handwritten text, Boundary Growing Method (BGM) based on the nearest neighbour concept was used. The method made to appear the characters and words in regular spacing in one direction and it can grow based on orientation of text. The concept of intrinsic and extrinsic edges was used to eliminate false positives.

The unique approach of Shyamaet.al[17]projected a text segmentation technique to extract text from any type of camera grabbed frame image or video. Colour based segmentation methodology was used to link consecutive pixels in the same direction by exploiting the general text properties. Light Edge Enhancement (LEE) was used to find a set of consecutive candidate points and enhance the edge between them. Next, heavy edge enhancement (HEE) was applied to remove or reduce motion blur from camera image sequences. This helped to treat camera images and video frames in the same manner.

Table 2: Performance Analysis of Text Extraction in Scene Text Images

S.No	Author	Year	Method used	Accuracy	Benefits
1	Kohei et.al [14]	2011	Edge-based method, connected component labeling method, Morphology erosion filter, Comic text extraction method	94.66%	Works on complex background.
2	Shivakumara et.al [16]	2010	Maximum Color Difference (MCD), Boundary Growing Method (BGM), K-means clustering algorithm	89.67%	Insensitive to contrast
3	Luz et.al [15]	2010	Morphological filters , Decision tree classifier	85.93%	Insensitive to position
4	Shyama et.al [17]	2009	Colour based segmentation, Light edge enhancement, Heavy edge enhancement	94%	Insensitive to size, orientation
5	Fabrizio et.al[13]	2009	Multiple SVM classifiers, Toggle Mapping Morphological Segmentation	88.83%	Insensitive to lighting, orientations.
6	Pan et. al [12]	2009	Local binarization approach , Conditional Random Field (CRF) model, Minimum classification error (MCE) learning , Graph cuts inference algorithm, Minimum spanning tree (MST), Energy minimization model	83.44%	Robust and accurately localize texts
7	Angadi et.al [11]	2009	DCT based high pass filter , Discriminant functions	96.60%	Handles different type of size, alignment, nonlinear text region

### 2.1.3 Caption Text Images

A superimposed text extraction method was introduced by V.Vijayakumar et.al [18] for detecting video text regions containing player information and score in sports videos. Key frames from the video were extracted using Color Histogram technique to minimize the number of video frames and converted to gray images. Text image regions were cropped. Canny Edge Detection algorithm was applied to detect edges on the cropped image. From this edge detected images, text region was identified and fed to an Optical Character Recognition system which produces index-able keywords.

The goal of Min et.al [19] approach was to detect both low-contrast and high-contrast artificial texts invariant with language and font-size in a complex background video image. The sobel color edge detector was applied to detect edges. Non text points were eliminated by applying low threshold determined by the histogram of edge strength and selective local thresholding. Further enhancement was done using Edge-Strength Smoothing (ESS) operator and Edge-Clustering-Power (ECP) operator. To locate the text region, coarse-to-fine (horizontal and vertical) projection was used.

Yih-Ming et.al [20] proposed a scheme to extract the caption text from various sports videos. Iteratively temporal averaging approach was used in caption extraction process. To improve the image quality and to reduce noise spatial-image analysis was performed. Threshold value was determined using binarization process based on the global mean and the standard deviation of the gray level of the averaged video image. Binarization may lead to holes and disconnectivity on video captions with blurred background. This was cured by morphological processing. Each connected component was used to extract geometrical features to identify the captions. A model-based segmentation approach was applied to accurately extract the caption contents.

A technique for detecting caption text from videos for global indexing purpose based on hierarchical region-based image model was proposed by Leon et.al [21]. Binary Partition Tree (BPT) was created by combining color and contour homogeneity Criteria. Texture descriptors were estimated on the full image by means of a multi-resolution analysis using a Haar wavelet decomposition to highlight the candidate regions in the BPT. The largest connected component was selected as the area of support for computing geometric descriptors. Region evaluation was carried out by combining region-based texture information and geometric features. Final caption text nodes were selected by analyzing the various subtrees in BPT.

Zhong et.al [22] introduced a method to automatically localize captions in JPEG compressed images and the I-frames of MPEG compressed video. They proposed a texture-based caption text localization method that operates directly in the Discrete Cosine Transform (DCT) domain for MPEG video or JPEG images. The DCT coefficients which capture the directionality and periodicity of local image blocks were used as texture measures to identify text regions. Morphological operations and connected component analysis were performed to remove noisy blocks and merge disconnected text blocks.

An effective approach to extracting captions from videos was projected by Liu et.al [23]. Spatial localization and temporal localization were used in their approach. Candidate caption region were detected by exploiting the distribution of corners in spatial localization. The temporal localization for different captions in a video was performed by identifying the change of stroke directions and decomposed the video into a sequence of clips, each clip containing the same caption.

Table 3: Performance Analysis of Text Extraction in Caption Text Images

S.No	Author	Year	Method used	Accuracy	Benefits
1	Vijayakumar [18] et.al	2011	Histogram technique, Canny Edge Detection algorithm	84.89%	Extracts caption text from any type of video.
2	Liu et.al [23]	2010	Spatial localization, Temporal localization, clip-based Segmentation	91.1%	Insensitive to language

					s, symbols
3	Leon et.al [26]	2010	Generic indexing system, Wavelet Transform, Hierarchical image model	85.78%	Insensiti ve to different size, color, complex backgro und
4	Leon et.al [21]	2009	Haar wavelet decomposition and geometric information through hierarchical image model	86.35%	Insensiti ve to different size,com plex backgro und
5	Yih-Ming et.al [20]	2006	Temporal averaging technique,Spatial-image analysis, Binarization process, morphological operations, Model-based segmentation approach.	92.18%	Indepen dent of caption size, color, location, shape, layout
6	Luo et.al [24]	2003	Supervised classification of the temporal feature vector	94.2%	Indepen dent of size, font, alignme nt
7	Min Cai[19] et.al	2002	Sobel color edge detector, Histogram technique, Edge strength smoothing (ESS) operator, Edge-clustering- power(ECP) operator, Coarse-to-fine projection	93.6%.	Robust for contrast, font size, language , backgro und complex ity
8	Tang et.al [25]	2002	Fuzzy-clustering, neural network classifier, Minimum pixel search method, frame averaging methods Quantized spatial difference density, Morphological operations	99%	Indepen dent of size, shape, alignme nt.

9	Zhong et.al [22]	1999	Discrete Cosine Transform, Morphological operations, connected component analysis	98.42%	Method is very fast
---	------------------	------	---	--------	---------------------

Luo et.al [24] proposed a technique to extract the text information in video to create a summary of the video segment. The text information in video was extracted by using brightness values of a pixel to form a vector, called Temporal Feature Vector (TFV). The vector was formed by tracing the gray-level of each pixel in time over a sequence of consecutive frames. By analyzing the pixel changes in the sequence, they located the appearing frames of captions. Finally they extracted the captions to create a summary of the video segment.

Tang et.al [25] presented a video caption detection and recognition system based on a fuzzy-clustering neural network (FCNN) classifier. A self-organizing neural network and fuzzy clustering classifier were used to segment the video sequence into basic frame units representing continuous action. The frame difference metrics, the histogram difference metric (HDM) and the spatial difference metric (SDM) were used to detect boundaries. They proposed a new metric, quantized spatial difference density (QSDD), to detect the caption transition frame and located the caption image region

Leon et.al [26] projected a technique for caption text detection that combines texture information and geometric information. Texture features were estimated through Haar wavelet decomposition and geometric information were estimated through the analysis of the regions proposed by the hierarchical image model.

#### 2.1.4 Heterogeneous Text Images

Rama et.al [27] proposed a new text extraction algorithm which was insensitive to noise, skewness and text orientation, color or intensity, layout and orientation from a text/graphics heterogeneous document images. The edge detection operation was performed using the basic operators of mathematical morphology. The text candidate connected components were found using the edges. These components have been labeled to identify different components of the image. The variance was found for each connected component considering the gray levels of those components. Then the text was extracted by selecting those connected components whose variance was less than some threshold value.

A unified approach for text extraction to handle all kinds of image such as scene text images, caption text & document images was put forth by Gopala et.al [28]. They presented an algorithm, based on the non-subsampled contourlet transform NSCT for text extraction. The contourlet wavelet transform using multiscale and directional filter banks (DFB) captures smooth contours images that were the dominant feature in natural images. The original image was decomposed into eight directional sub band outputs using the DFB and the energy of each sub band were obtained. The Sub bands were categorized as Strong & weak based on the value of the computed energy. Weak sub bands were boosted to get the proper edges. Detected edges were dilated using morphological dilation to enlarge or group the identified text regions. Strong & boosted edges after dilation were combined with addition followed by logic AND operation to extract text regions. Finally, remaining non text regions were identified & eliminated.

Khelifi et.al in [29] proposed an approach for text extraction to categorize and index documents from heterogeneous documents. This approach was to identify similar text regions based on their fonts. SVM classifier, fractal descriptors were used to perform this task. To make image homogenous different color layers were extracted. The results were binarized, and text region candidates were grouped. An SVM classifier was used to detect if a connected component belonged to text entities or not. The CDB (Counting Densities per Box) was calculated to locate text region. Fractal descriptors were calculated for extracted local text zones and to perform a

font classification step. Experiments were evaluated for both maps taken from geological department database and ancient documents. The detection of text zones were done for every direction, but the font recognition was well achieved only for horizontal zones.

Sunil et.al [30] proposed a scheme for the extraction of textual areas from an image using Globally Matched Wavelet Filters(GMW) filters with Fisher classifiers.GMW filters was estimated using clustering-based technique. They have used these filters to segment the document images and classify them into text, background, and picture components. To improve the result Markov random field (MRF) based post processing had been applied.

Liu et.al [31]’s dealt with printed document images as well as with scene text. They proposed edge based method with edge strength, density and the orientation variance as distinguishing characteristics of text embedded in images to build a feature map. This method used multiscale edge detector for text detection and morphological dilation operator for text localization stages.

A new text extraction method which was insensitive to variations in font, color, or size of the text in mixed-type color documents was presented by C.Strouthopoulos et.al [32]. The method was based on a combination of an adaptive color reduction (ACR) technique and a page layout analysis (PLA) approach. The Adaptive tree clustering procedure using principal component analyzer (PCA) and self-organized feature map (SOFM) was used to achieve color reduction. On individual color planes the PLA technique based on a Run Length Segmentation algorithm (RLSA) and a neural network block classifier fed by suitable texture spatial features was applied to identify text regions. The text regions were merged to determine final text regions.

G. Sahooet.al [33] projected a set of sequential algorithms for text extraction and enhancement of image using cellular automata. The Luminance-based algorithm was used to convert the image in to grey scale image.Converted image have only luminosity attribute. The edge detection was performed using a  $3 \times 3$  Sobel operator and it was then followed by the elimination of non-maxima and thresholding of weak edges. The edge-bounded averaging was performed through Moore neighborhood to obtain smooth non-edge regions. The image was classified in to text based or non-text based region using constant threshold.

A combination of two learning mechanism, an artificial neural network (ANNs)-based approach and Non-negative Matrix Factorization (NMF)-based filtering was proposed by Keechul et.al [34] for text extraction in complex images. Multilayer Perceptron ANN classifier increased a recall rate and a precision rate with NMF-filtering-based Connected Component (CC) analysis. Detection of text was performed using neural networks without any explicit feature extraction stage. MLPs automatically generated a texture classifier that discriminates between text regions and non-text regions on three color bands. Bootstrap method was used in MLPs to learn a precise boundary between text and non-text classes. To overcome the locality property of the texture-based method, they used CC-based filtering using the NMF technique. Enhancement on time was done using CAMShift for video images and X-Y recursive cut algorithm for document images.

Table 4 : Performance Analysis of Text Extraction in Heterogeneous Text Images

S.No	Author	Year	Method used	Accuracy	Benefits
1	Rama et.al [27]	2010	Mathematical Morphology	84.01%	Insensitive to noise, skew orientation
2	Trung et.al [35]	2010	Skeletonization operation, Laplacian operator, Morphological operation	84.90%	Detect s multioriented text.

3	G. Sahoo et.al [33]	2009	Luminance-based algorithm ,Sobel Edge detection, Moore neighborhood	91.88%	Insensitive to color,orientations
4	Gopala et al. [28]	2008	Contourlet wavelet transform , Morphological dilation operator,	91%	Handle different languages
5	Khelifi et al. [29]	2008	SVM classifier, Fractal descriptors, Adaptive Wiener filter	Good Result	Handles horizontal Text zones
6	Sunil et al.[30]	2007	Globally matched wavelet filters, Fisher classifiers, clustering-based technique, Markov random field (MRF) based post processing	85.54%	Independent of scripts, font, font-size, geometric distortion
7	Liu et.al [31]	2006	Multiscale edge detector , Morphological dilation operator	94.30%	Insensitive to font style, alignment
8	Keechul et.al [34]	2004	artificial neural network,non-negative matrix factorization, Bootstrap method,	92.38%	No feature extraction stage.
9	C.Strouthopoulos et.al [32]	2002	Adaptive color reduction technique , Page layout analysis, , Neural network classifier, Self-organized feature map Principal component analyzer	majority of the text areas were correctly obtained	Robust to variations in text size, font, color

Phan et.al [35] proposed an approach based on the skeletonization operation for multi-oriented graphics text and scene text in video images. This method used the laplacian operator to highlight the transitions between text and background. K-means was used to classify text and non-text region. The morphological open operation was used to remove small artifacts from the text cluster. Each region was classified as either a simple or complex connected component depending on the number of intersection points in its skeleton. Complex connected components were then segmented into essential parts based on the skeleton segments in order to separate the text strings from each other. Finally, text string straightness and edge density were used for false positive elimination.

## 2.2 Performance Evaluation

There are several performance evaluations to estimate the algorithm for text extraction. Most of the approaches quoted here used Precision, Recall and F-Score metrics to evaluate the performance of the algorithm. Precision, Recall and F-Score rates are computed based on the number of correctly detected characters (CDC) in an image, in order to evaluate the efficiency and robustness of the algorithm. The performance metrics are as follows:

### **2.2.1 False Positives**

False Positives (FP) / False alarms are those regions in the image which are actually not characters of a text, but have been detected by the algorithm as text.

### **2.2.2 False Negatives**

False Negatives (FN)/ Misses are those regions in the image which are actually text characters, but have not been detected by the algorithm.

### **2.2.3 Precision rate**

Precision rate (P) is defined as the ratio of correctly detected characters to the sum of correctly detected characters plus false positives.

### **2.2.4 Recall rate**

Recall rate (R) is defined as the ratio of the correctly detected characters to sum of correctly detected characters plus false negatives.

### **2.2.5 F-score**

F-Score is the harmonic mean of recall and precision rates.

## **3.CONCLUSION &FUTUREWORK**

There are many applications of a text extraction such as Keyword based image search, text based image indexing and retrieval , document analysis, vehicle license detection and recognition, page segmentation ,technical paper analysis, street signs, name plates, document coding, object identification, text based video indexing, video content analysis etc. A number of methods have been proposed in the past for extraction of text in images. These approaches considered the different attributes related to text in an image such as of size, font, style, orientation, alignment, contrast, color, intensity, connected-components, edges etc. These attributes are used to classify text regions from their background or other regions within the image. This paper provides a broad study of the various text extraction techniques and algorithms proposed earlier. This paper also exposed a performance comparison table of different technique that was proposed earlier for text extraction from an image. Every approach has its own benefits and restrictions. Even though there are many numbers of algorithms, there is no single unified approach that fits for all the applications. The future work mainly concentrates on developing an algorithm for exact and fast text extraction from an image.

## **ACKNOWLEDGEMENTS**

This work is a part of UGC Minor Research Project. Reference Number-No F. MRP-3725/11(MRP/UGC-SERO), Dated: 08/09/2011. Authors thank the Management and Principal for extending their support for this project.

## **REFERENCES**

- [1] Y. Zhan, W. Wang, W. Gao (2006), "A Robust Split-And-Merge Text Segmentation Approach For Images", *International Conference On Pattern Recognition*, 06(2):pp 1002-1005.
- [2] Thai V. Hoang , S. Tabbone(2010), "Text Extraction From Graphical Document Images Using Sparse Representation" in *Proc. Das*, pp 143–150.

- [3] Audithan,,R.M.Chandrasekaran (2009), "Document Text Extraction From Document Images Using Haar Discrete Wavelet Transform" ,*European Journal Of Scientific Research*, Vol.36 No.4 , pp.502-512.
- [4] Sachin, Grover,Kushal Arora,,Suman K. Mitra(2009),“Text Extraction From Document Images Using Edge Information” ,*IEEE India Council Conference*.
- [5] P. Nagabhusan, S. Nirmala(2009) ,”Text Extraction In Complex Color Document Images For Enhanced Readability” ,*Intelligent Information Management*, pp: 120-133.
- [7] Davod Zaravi, Habib Rostami, Alireza Malahzaheh, S.S Mortazavi(2011),” Journals Subheadlines Text Extraction Using Wavelet Thresholding And New Projection Profile” , *World Academy Of Science, Engineering And Technology* .Issue 73.
- [8] Karin Sobottka, Horst Bunke and Heino Kronenberg(2009), “Identification Of Text On Colored Book And Journal Covers” , *ICDAR*.
- [9] Zhixin Shi, Srirangaraj Setlur And Venu Govindaraju(2005), “Text Extraction From Gray Scale Historical Document Image Using Adaptive Local Connectivity Map” , *Proceeding Of The Eighth International Conference On Document Analysis And Recognition*, Vol. 2, pp: 794–798.
- [10] Syed Saqib Bukhari , Thomas M. Breuel,Faisal Shafait(2009), “Textline Information Extraction From Grayscale Camera-Captured Document Images “ , *ICIP Proceedings Of The 16th IEEE International Conference On Image Processing*, pp: 2013 – 2016.
- [11] Boussellaa , Aymen Bougacha, Abderrazak Zahour, Haikal El Abed, Adel Alimi(2009) ,“Enhanced Text Extraction From Arabic Degraded Document Images Using Em Algorithm” , *10th International Conference On Document Analysis And Recognition*.
- [12] S. A. Angadi , M. M. Kodabagi(2009) , ”A Texture Based Methodology For Text Region Extraction From Low Resolution Natural Scene Images “ , *International Journal Of Image Processing (Ijip)* Volume(3), Issue(5).
- [13] Yi-Feng Pan, Xinwen Hou, Cheng-Lin Liu(2009), “Text Localization In Natural Scene Images Based On Conditional Random Field,” *ICDAR*,pp 6-10.
- [14] J. Fabrizio, M. Cord, And B. Marcotegui(2009), “Text Extraction From Street Level Images,” , *CMRT*, Vol. Xxxviii, Part 3/W4 , pp. 199–204.
- [15] Kohei Arai , Herman Tolle(2011),” Text Extraction From Tv Commercial Using Blob Extraction Method” , *International Journal Of Research And Reviews In Computer Science* Vol. 2, No. 3
- [16] Wonder Alexandre Luz Alves And Ronaldo Fumio Hashimoto(2010),”Text Regions Extracted From Scene Images By Ultimate Attribute Opening And Decision Tree Classification” , *Proceedings of the 23rd Sibgrapi Conference On Graphics, Patterns And Images*.
- [17] Shivakumara P, A Dutta, U Pal And C L Tan(2010), “A New Method For Handwritten Scene Text Detection In Video” , *International Conference On Frontiers In Handwriting Recognition*, pp: 16-18.
- [18] Shyama Prosad Chowdhury,Soumyadeep Dhar,Karen Rafferty,Amit Kumar Das,Bhabatosh Chanda(2009),”Robust Extraction Of Text From Camera Images Using Colour And Spatial Information Simultaneously” ,*Journal Of Universal Computer Science*,Vol. 15, No.18 , pp:3325-3342.
- [19] V.Vijayakumar,R.Nedunchezianm(2011),”A Novel Method For Super Imposed Text Extraction In A Sports Video” ,*International Journal Of Computer Applications*,Volume 15– No.1.
- [20] Min Cai, Jiqiang Song, Michael R. Lyu(2002),”A New Approach For Video Text Detection” ,*Proceedings International Conference On Image Processing* , Volume 1, pp: I-117-I-120.

- [21] Yih-Ming Su, Chaur-Heh Hsieh(2006), "A Novel Model-Based Segmentation Approach To Extract Caption Contents On Sports Videos", *IEEE International Conference On Multimedia And Expo*, pp:1829 - 1832 .
- [22] Miriam Leon, Veronica Vilaplana, Antoni Gasull, Ferran Marques(2009) , "Caption Text Extraction For Indexing Purposes Using A Hierarchical Region-Based Image Model", *,Proceedings Of The 16th IEEE International Conference On Image Processing*, pp:1869-1872.
- [23] Yu Zhong, Hongjiang Zhang, And Anil K. Jain(1999),"Automatic Caption Localization In Compressed Video", *International Conference On Image Processing*, pp: 96 - 100 Vol.2.
- [24] Xiaoqian Liu,Weiqliang Wang(2010) ,"Extracting Captions From Videos Using Temporal Feature",*Proceedings Of The International Conference On Acm Multimedia* ,pp:843-846.
- [25] Bo Lilo, Xaou Tang, Jianzhuang Liu, And Hongjiang Zhan(2003) ,"Video Caption Detection And Extraction Using Temporal Information", *International Conference On Image Processing*, Vol.1 , pp:I 297-300 .
- [26] Tang X, Gao X, Liu J, Zhang H(2002). "A Spatial-Temporal Approach For Video Caption Detection And Recognition",*IEEE Transactions On Neural Networks*, Vol. 13, No. 4.
- [27] Miriam Leon, Veronica Vilaplana, Antoni Gasull, Ferran Marques(2010),"Region-Based Caption Text Extraction",*11th International Workshop On Image Analysis For Multimedia Interactive Services (Wiamis)*.
- [28] G. Rama Mohan Babu, P. Srimaiyee, A.Srikrishna(201), "Text Extraction From Heterogeneous Images Using Mathematical Morphology",*Journal Of Theoretical And Applied Information Technology*,Vol.16,No.1,pp 39-47.
- [29] Chitrakala Gopalan , Manjula(2008) ,"Text Region Segmentation From Heterogeneous Images", *International Journal Of Computer Science And Network Security*, Vol.8 No.10, pp.108-113.
- [30] Badreddine Khelifi, Nizar Zaghden, Adel M. Alimi, Rémy Mullot(2008), "Unsupervised Categorization Of Heterogeneous Text Images Based On Fractals",*Proceedings of ICPR'*,pp.1-4.
- [31] Sunil Kumar, Rajat Gupta, Nitin Khanna, Santanu Chaudhury, Shiv Dutt Joshi(2007),"Text Extraction And Document Image Segmentation Using Matched Wavelets And MRF Model", *IEEE Transactions On Image Processing*, Vol. 16, No. 8.
- [32] X. Liu And J.Samarabandu(2006), "Multiscale Edge-Based Text Extraction From Complex Images," *Proc. International Conference Of Multimedia And Expo*, pp.1721-1724.
- [33] Strouthopoulos, N. Papamarkos, A.E. Atsalakis(2002)," Text Extraction in Complex Color Documents", *Pattern Recognition* 35 pp:1743–1758 .
- [34] G. Sahoo, T. Kumar, B. L. Raina, C. M. Bhatia(2009), "Text Extraction And Enhancement Of Binary Images Using Cellular Automata", *International Journal Of Automation and Computing*, Vol. 6, No. 3, pp. 254-260.
- [35] Keechul Jung, Eun Yi Ki(2004),. "Automatic Text Extraction For Content-Based Image Indexing",*Proceedings of PAKDD*,pp.497-507.
- [36] Trung Quy Phan ,Palaiahnakote ,Shivakumara Chew Lim Tan(2010)," A Skeleton-Based Method For Multi-Oriented Video Text Detection", *Das '10 Proceedings Of The 9th IAPR International Workshop On Document Analysis Systems*,pp :271-278.