IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 2, February 2016

A Survey: Text Extraction from Images and Video

Vivek Dhanapal Sapate

M.E. (Department of Computer Science), DKTE's Collage of Engineering, Ichalkaranji, India

Abstract: Text information present in pictures and video contain valuable info. Text extraction from image has stages of detection the text from given image, finding the text location, extraction, improvement and recognition of text from the given image. But variations of text just like the variations in orientation, size, style, alignment; low size image distinction and a lot of difficult background create the matter of automatic text extraction extraordinarily troublesome. The number of techniques and Methodology are planned to this downside, and thus the aim of this paper is to review the paper, discuss datasets and performance analysis, and to point the long run analysis.

Keywords: Text extraction, image, video, compression, text location, extraction, size, style, alignment.

1. INTRODUCTION

Growth of accessible multimedia documents and increasing demand for info categorization and retrieval, rather additional effort has been done on text extraction in pictures and videos. Applications like Mobile Text Recognition applications, Sign Detection and Translation, license plate reading, Content primarily based Image Search so on. Text consists of words that are parallel-defined models of concepts. Text objects embedded in video contain lots of correct knowledge related to the multimedia system. Text extraction techniques play an important role in multimedia knowledge categorization and retrieval.

Extracting text from pictures or videos may be a vital drawback in many applications like document processing, image assortment, video retrieval [14], video content summary [12] so on. Usually, texts embedded in an exceedingly image or a frame capture necessary media contexts like player's details, title of the films, date in news channels, story introduction etc. Studies on image content among the kind of text, face, vehicle, and act have to boot attracted some recent interest. Among them, text among an image is of specific interest as (i) really useful for describing the contents of Associate image; (ii) it permits applications like text-based image classification, keyword-based image search, automatic video work, and so on. After all as results of the variability of fonts, multi orientations, utterly completely different sizes and styles of fonts, alignment effects of uncontrolled illuminations, reflections, shadows, and the distortion due to perspective projection to boot as a result of the complex of image background, automatic localizing and extracting text might be a troublesome downside.

1.1 Types of text

Images could also be loosely classified into Document text footage, Caption text footage and Scene text footage. Figures 1-3 show some samples of text in footage.

A document image [20] (Figure 1) usually contains text and few graphics components. Document footage are comes from scanning journal, papers, degraded document, papers, and book cowl etc. The text may appear in a during in an exceedingly in a very nearly unlimited vary of fonts, style, alignment, size, shapes, colors, etc. Extraction of text in documents with text on difficult color background is hard as results of quality of the background associate of color(s) of fore-ground text with colors of background.

Caption text [13, 20] is known as Overlay text or Cut line text. Caption text (Figure 2) is artificially superimposed on the video/image at the time of writing and it invariably describes or identifies the subject of the image/video content. The superimposed text might be a powerful source of high-level linguistics. These text occurrences is detected, segmented, and recognized automatically for assortment, retrieval and summarization. The extraction of the superimposed text in sports video is extremely useful for the creation of sports define, highlights etc. These styles of caption text in video [19] embody moving text, rotating text, growing text, shrinking text, text of different orientation, and text of impulsive size.

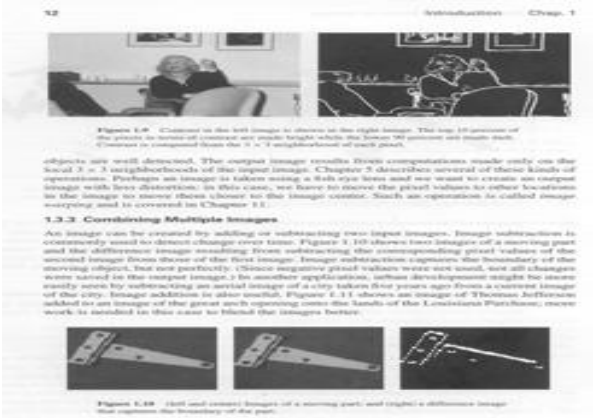

Figure 1

Scene text [13, 20] (Figure 3) seems at intervals the scene that's then captured by the recording device i.e. text that's present among the scene once the image or video is captured.

DOI 10.17148/IJARCCE.2016.5286

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 2, February 2016

Figure 2

Scene texts happens naturally as a area/region of the scene and contain very important linguistics knowledge like advertisements that embody artistic fonts, names of streets, institutes, shops, road signs, board signs, nameplates, food containers, street signs, bill boards, text on vehicle so on. Scene text extraction could also be utilized in detection text-based landmarks, vehicle license plate detection/recognition, and object identification rather than general assortment/indexing and retrieval. it's robust to seek out and extract since it ought to appear in unlimited form of poses, size, shapes and colors, low resolution, difficult background, heterogeneous lightning or blurring effects of variable lighting, difficult movement and transformation, unknown layout, shadowing and variation in font vogue, size, orientation

Figure 3

1.2 Properties of text

Text properties are the appearance and behavior of the text object.

Size: although the text size will vary plenty, assumptions can be lots looking on the application domain.

Alignment: The characters within the caption text seem in clusters and typically lie horizontally, although generally they'll appear as non-planar texts as results of computer graphics. This doesn't apply to scene text, which might have numerous perspective distortions. Scene text is aligned in any direction and may have geometric distortions.

Inter-character distance: characters in a text line have an identical distance between them.

Color: The characters in a text line tend to own an equivalent or similar colors. This property makes it potential to use a connected component-based approach for text detection. Most of the analysis reported till date has focused on finding 'text strings of one color (monochrome). However, video pictures and alternative complicated color documents will contain 'text strings with over 2 colors (polychrome)' for effective visual image, i.e. totally different colors at intervals one word.

Motion: an equivalent characters sometimes exist in consecutive frames during a video with or while not movement. This property is employed in text trailing and improvement. Caption text usually moves in a very uniform way: horizontally or vertically. Scene text will have arbitrary motion owing to camera or object movement.

Edge: Most caption and scene text are designed to be simply scan, thereby leading to robust edges at the boundaries of text and background.

Compression: several digital pictures are recorded, transferred, and processed during a compressed format. Thus, a faster TIE system is achieved if one will extract text without decompression.

1.3 Architecture Text Information Extraction

A Text data Extraction System Receive an input image and output the relevant text information. Images will be in gray scale or color, compressed or uncompressed. The TIE drawback are often divided into the subsequent sub-problems: (i) detection, (ii) localization, (iii) tracking, (iv) extraction and enhancement, and (v) Optical Character recognition (OCR)

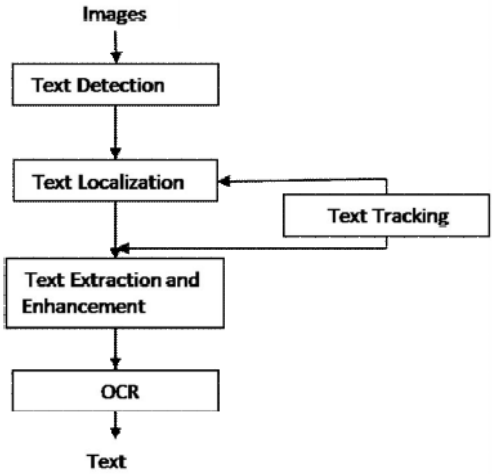

Figure4. Architecture of TIE system

Text detection determines the presence of text in a very given frame. Text detection to estimate text existing confidence in native image regions by classification Text localization is that the method of deciding the location of text within the image and generating bounding boxes around the text and text verification to get rid of non-text regions for more process.

Text tracking is performed to cut back the time interval for text localization. Although the precise location of text in a

DOI 10.17148/IJARCCE.2016.5286

picture will be indicated by bounding boxes, the text still has to be segmented from the background to facilitate its recognition. This implies that the extracted text image has got to be converted to a binary image and enhanced before it's fed into an OCR engine.

Text extraction is that the stage where the text components are segmented from the background. Enhancement of the extracted text parts is needed as a result of the text region typically has low-resolution and is suffer from noise. Thereafter, the extracted text images will be transformed into plain text applying OCR technology.

Text info extraction techniques principally consist of 5 vital phases: text region detection, text localization, tracking, character extraction, text recognition. From that initial 2 (text region detection, text localization) stages are more vital and also they're harder to implement. The output of text data extraction is especially obsessed with the output of those 2 phases. The techniques used for text data extraction falls in 5 categories are as follows:

1. Region -Based Technique
Region-based methods [13, 17] use the properties of the color or gray-scale in an exceedingly text region or their variations with the corresponding properties of the background. This methodology uses a bottom-up approach by grouping little components into in turn larger components until all regions are known in the image. A geometrical analysis is required to merge the text components using the placement of the parts therefore on separate out non-text components and mark the boundaries of the text regions. These strategies is further divided into 2 sub-approaches: connected component (CC)-based and edge-based. These 2 approaches add a bottom-up fashion; by distinctive sub-structures, like CCs or edges, then merging these sub-structures to mark bounding boxes for text. Note that some approaches use a mixture of each CC-based and edge-based strategies.

2. CC-based Technique
CC-based strategies [13, 20] use a bottom-up approach by grouping small components into in turn larger components until all regions are known within the image. A geometrical analysis is required to merge the text elements using the spatial arrangement of the components therefore on separate non-text elements and mark the boundaries of the text regions.

3 .Edge-based Techniques
Among the many textual properties in an image, edge-based strategies [16] specialize in the 'high contrast between the text and therefore the background'. The edges of the text boundary are known and merged, and then many heuristics are used to separate the non-text regions. Usually, an edge filter (e.g., a canny operator) is employed for the edge detection, and a smoothing operation or a morphological operator is used for the merging stage.

4. Texture-Based Techniques
Texture-based strategies [15] use the observation that text in images has distinct textural properties that differentiate them from the background. The strategies based on gabor filters, Wavelet, etc. are often wont to find the textural properties of a text region in an image.

5. Morphological based Technique
Mathematical morphology [11] may be a topological and geometrical primarily based approach for image analysis. It provides powerful tools for extracting geometrical structures and representing shapes in several applications. Morphological feature extraction techniques are with efficiency applied to character recognition and document analysis. It's used to extract necessary text contrast options from the processed images. The feature is invariant against numerous geometrical image changes like translation, rotation, and scaling. Even when the lighting condition or text color is modified, the feature still are often maintained. This technique works robustly underneath totally different image alterations

2. REVIEW OF RECENT PAPERS

Cong Yao, Xiang Tibeto-Burman and Wenyu Liu [2] introduced a unified framework model for text detection and recognition as a complete and performs every tasks in single unified pipeline. This paper are introduced in three parts: 1) text detection and recognition are accomplished practice completely same features and classification scheme; 2) that in the main consider horizontal or near-horizontal texts, the projected system is capable of localizing and reading texts of variable orientations; and 3) used a modified dictionary search technique supported Levenshein edit distance [21] , to correct the recognition errors generally caused by confusions among similar yet whole totally different characters.

The projected algorithm is prepared to discover and acknowledge texts of various scales, colors, fonts and orientations, curved, reversed words is successfully localized and read.

Limitations:
Failure cases of text detection. The misses are in the main as a result of non-uniform lighting condition, blur, low resolution and low diversity between text and background and variety of different typical failure cases of character recognition are partial misses in detection, improper word partitions, irregular fonts, connected characters, all turn out to recognition errors.

Jing Zhang and Rangachar Kasturi [1] projected novel technique by applying three new character choices i.e. Average agle distinction of corresponding pairs, Fraction of non-noise pairs and Vector of stroke width to observe text objects in images/videos then calculate the character energy, link energy and also the Text unit energy. a replacement text model is created to explain text objects. Each character may well be a half in the model and every two neighboring characters are connected by a link. Two characters and thus the link connecting them are made defined as a text unit.

Limitations:
(a) Continuance patterns. (b) Connected characters. (c) Single character. (d) Small characters. (e) Clear characters

A recent paper, Semiautomatic Ground Truth Generation for Text Detection and Recognition in Video frames [8] proposed by Trung Quy Phan, Palaiahnakote Shivakumara, Souvik Bhowmick, Shimiao Li,Chew Lim Tan, and Umapada Pal. They projected a semiautomatic system for ground truth generation for video text detection and recognition that has English and Chinese text of multi orientation at word level. Ground trothing for text detection and recognition involves text line segmentation, word segmentation, bounding box drawing, deciding field, graphics and scene text separation. The system includes a facility to allow the user to manually correct the bottom truth if the machine-driven technique produces incorrect results. Proposes eleven attributes at the word level, namely: line index, word index, coordinate values of bounding box, area, content, script kind, orientation knowledge, kind of text (caption/scene), condition of text (distortion/distortion free), begin frame, and end frame to judge the performance of the strategy.

Limitations:
This methodology generally produces false positives and integrated text lines (nearby text lines are connected to each other)

Scene Text Recognition applying Structure-Guided Character Detection and Linguistic Knowledge [7] projected by Cun-Zhao Shi, Chun-Heng Wang, Bai-Hua Xiao, and Song GAO, Jin-Long Hu projected a completely unique scene text-recognition technique combination of structure-guided character detection and linguistic info. Use of every global structure and native look information of characters, build a part-based tree structure to model each category of characters so on along observe and acknowledge characters at identical time.. For word recognition, mix the detection scores and language model into the posterior likelihood of character sequence from the Bayesian decision tree and the final word recognition result's obtained by finding most likelihood character sequence utilized by Viterbi algorithm and thus the various information a bit like the language model to eliminate the word recognition probable ambiguities.
The final word-recognition result's obtained by maximizing the character sequence posterior chance via Viterbi algorithm.

Limitations:
a) Huge Gap between the characters
b) As a results of blur, and have larger deformation or distortion.

Chucai Yi and Yingli Tian [6] given a method combines scene text recognition and scene text detection algorithms. In text detection, projected a Layout based primarily scene text detection algorithms are applied to get text regions from scene image. In scene text recognition schemes, structure based scene text recognition technique is used. First, style a discriminative character descriptor by combining several progressive feature detectors and descriptors. It combines several feature detectors (Harris-Corner, outside Stable Extremal Regions (MSER), and dense sampling) and Histogram (bar chart) of Oriented Gradients (HOG) descriptors. Second, model character structure at each character class by with stroke configuration maps. , to urge a binary classifier for each character class in text retrieval, a completely novel (unique) stroke configuration from character boundary and skeleton to model character structure. Algorithm is compatible with the appliance of scene text extraction in wise mobile devices.

Limitations:
Accuracy rates of text detection, Scene text extraction and add lexicon analysis to extend system to word level recognition.
Color uniformity and aligned arrangement area unit acceptable for the captured text knowledge from natural scene.

Xu-Cheng rule, Xuwang Yin, Kaizhu Huang, ANd Hong-Wei Hao [5] projected at an correct and sturdy technique for detection texts in natural scene photos. Throughout this paper propose a robust and proper Maximally Stable Extremal Regions MSER-based scene text detection technique. First, a designed a fast and effective pruning algorithm may well be a Maximally Stable Extremal Regions (MSERs); the amount of character candidates to be processed is reduced with high accuracy. Second, Character candidates are classified into text candidates by the single-link clustering algorithm, where distance weights and clustering threshold are learned automatically by a completely unique (novel) self-training distance metric learning algorithm. Third, use a classifier to estimate the chance of text candidates just like non text and remove text candidates. For Multi orientation text detection algorithm using a heuristic strategy- the forward-backward algorithm. In forward backward algorithm, calculate the orientation of text lines and convert the discretionary text lines into the horizontal text lines. Converted text line all over again fed into the horizontal text line detection algorithm.

Limitations:
First, the way to observe blurred texts in low resolution natural scene photos may well be a detailed to future issue. Second, some multi language texts have quite whole totally different characteristics from English texts.

Yao Li, Wenjing Jia, Chunhua Shen, and Anton van den Hengel [4] proposed this paper shows the detection methods to measures of abjectness. Throughout this paper describes the characterness model, regions are extracted by modified MSER-based region detector. Then computed noval characterness cues, then these cues unit of measurement utilized during a Bayesian framework where naïve Bayes is used to model the probability. Text is made from sets of characters, then vogue a markov random field model so on exploit the inherent dependencies between characters.

Limitations:
Extremely blur and Low resolution characters, character in uncommon fonts

Jerod J. Weinman,, Zachary manservant, Dugan hillock,

and Jacqueline Field [9] projected a bit, A system that handles many stages of scene text reading in probabilistic manner, from binarization to seem standardization to character segmentation and recognition. Throughout this work, describe a reading system that integrates a simple region grouping rule and probabilistic models for binarizing a given text region, distinctive baselines, along perform word and character segmentation throughout recognition technique. Use semi markov model that's utilized for integrate several knowledge like character fonts & styles, languages and pure mathematics.

Limitations:
Scene text recognition is difficult—the world's whole totally different colors, uncontrolled lighting, and unpredictable views conspire to form general machine reading a "grand challenge"-worthy task.

In this work projected by Ali Mosleh,Nizar Bouguila, Abdesamad mount Hamza [10] projected a bit to erase the unwanted text from the video. Throughout this work presents a two stages (i) automatic video text detection and (ii) restoration once the removal. Support Vector Machine (SVM) base video text detection technique is used to localize the text from video frames. Develop one frame text detection algorithm using a Stroke Width Transform (SWT) and unsupervised classification. Video caption detection is performed by multilayer perceptron theme and genetic algorithm. Bandlets based 3D video inpainting algorithm is employed for to revive the parts occluded by the removal texts.

Limitations:
Extend info relating to kinds of text to differentiate and remove from videos

Palaiahnakote Shivakumara, Trung Quy Phan, Shijian Lu, and Chew Lim Tan [3] presents a replacement technique supported gradient vector flow (GVF) and neighbor part grouping that extracts text lines of any orientations. GVF for characteristic text component applying Sobel edge map attributable to sobel provides fine details for text and fewer details for nontext on top of the canny edge map. They planned a two stage grouping criterion for Text Candidates. (i) Text Candidates are to be verify the nearest neighbor based on size and angle of the text candidate to cluster them. Introduced a skeleton conception on text elements to eliminate false text parts. This text is Candidate Text part (CTC).(ii)Tails of the CTC is use to analyze the direction of text information and notice nearest neighbor CTC. At this stage conclusion of Multi-Oriented text detection in video.

Limitations:
The projected technique may not provide sensible accuracy for horizontal text lines with less spacing between text lines

3. LISTING THE SURVEY OF TEXT EXTRACTION USING DIFFERENT APPROACHES

A listing of the published work on different approaches used for text extraction is presented in Table 1.

Table 1: Performance table on survey of Text Extraction of using different approaches

Sr. No	Author	Year	Method/Techniques	Datasets	Performance		
					Precision	Recall	F measure
1	Cong Yao, Xiang Bai, Wenyu Li	November 2014	SWT & Clustering, Component level Features and Classifiers, Error Correction	ICDAR11 MSRA-TD500 HUST-TR400	0.822 0.64 0.415	0.657 0.62 0.386	0.730 0.61 0.393
2	Jing Zhang, Rangachar Kasturi	September 2014	A Novel Algorithm for Text Detection, Zero-crossing based edge detector, Character Features, Maximum Spanning Tree(MST) and Prims algorithm	ICDAR2003/2005 MICROSOFT SVT VACE	0.74 0.52 0.495	0.62 0.38 0.383	0.67 0.44 0.589
3	Trung Quy Phan, Palaiahnakote Shivakumara, Souvik Bhowmick, Shimiao Li, Chew Lim Tan, Umapada Pal	August 2014	Ground Truthing, A Laplacian approach, A new Fourier Moments based word and character extraction	HORIZANTAL TEXT(MANUAL COUNTING) HORIZANTAL TEXT(AUTOMATIC COUNTING) NONHORIZANTAL TEXT(MANUAL COUNTING) NONHORIZANTAL TEXT(AUTOMATIC COUNTING)	0.82 0.52 0.68 0.25	0.75 0.53 0.51 0.28	0.78 0.53 0.11 0.27

DOI 10.17148/IJARCCE.2016.5286

4	Cun-Zhao Shi, Chun-Heng Wang, Bai-Hua Xiao, Song Gao, Jin-Long Hu	July 2014	Tree Structure Model (TSM), Bayesian decision view, Viterbi Algorithm	RECOGNITION RATES(WITH EDIT DISTANCE CORRECTION) ICDAR03 ICDAR11 SVT RECOGNITION RATES(WITHOUT EDIT DISTANCE CORRECTION) ICDAR03 ICDAR11 SVT		79.58 83.21 73.67 58.56 55.96 38.64	
5	Chucai Yi, Yingli Tian	July 2014	Character Descriptor, Stroke Configuration, Layout Based Scene Text Detection, Structure Based Scene Text Recognition	CHARS74K Sign ICDAR03		Accuracy Rates (AR) 0.726 0.868 0.536	False Positive Rates (FPR) 0.078 0.075 0.180
6	Xu-Cheng Yin, Xuwang Yin, Kaizhu Huang, Hong-Wei Hao	May 2014	A Novel MSER-based Scene Text Detection Method, Single Link Clustering, Distance Metric Learning Algorithm, Forward-Backward Algorithm	ICDAR11 MULTILINGUAL DB Epshtein db 	Horizontal Multi-Oriented Horizontal Multi-Oriented Horizontal Multi Oriented	0.862 0.82 0.826 0.687 0.66 0.54	0.682 0.6663 0.685 0.678 0.41 0.42	0.762 0.7352 0.746 0.683 0.51 0.47
7	Yao Li, Wenjing Jia, Chunhua Shen, Anton van den Hengel	April 2014	Egde-Preserving MSER, Characterness Evaluation, Labeling & Grouping	ICDAR11 ICDAR03		0.80 0.79	0.62 0.64	0.70 0.71
8	Jerod J. Weinman, Zachary Butler, Dugan Knoll,Jacqueline Field	February 2014	Region Grouping, Segmentation and Binarization, Text Line Normalization, Text Line Recognition	ICDAR11		0.41	0.36	0.33
9	Ali Mosleh,Nizar Bouguila, Abdesamad Ben Hamza	November 2013	SWT, Novel Bandlet Transform, K-means Clustering, CAMSHIFT Algorithm, Bandlet Based 3D Video Volume Inpainting Algorithm	ICDAR VACE		0.76 0.70	0.66 0.61	0.71 0.65
10	Palaiahnakote Shivakumara, Trung Quy Phan, Shijian Lu, Chew Lim Tan	October 2013	Sobel Edge Map, Gradient Vector Flow for Dominant Text pixel Selection	ICDAR03 HUA'S DATA		0.76 0.74	0.92 0.88	0.83 0.80

DOI 10.17148/IJARCCE.2016.5286

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 2, February 2016

4. PERFORMANCE EVALUATION

Precision, Recall and F-Score rates are computed supported the number of correctly detected characters (CDC) in an image, so as to evaluate the potency and strength of the algorithmic program.

The performance metrics are as follows:
False Positive (FP): Those regions within the image that are literally not characters of a text however are detected by the algorithm as text.
False Negatives (FN): Those regions parts in the image that are literally text characters however haven't been detected by algorithm.
Precision Rate (P): Defined as the ratio of properly detected characters to the sum of properly detected characters and false positives.

$$P = \frac{CDC(\text{true positives})}{CDC + FP} * 100$$

Recall Rate (R): Defined as the ratio of the properly detected characters to sum of properly detected characters and false negatives.

$$R = \frac{CDC(\text{true positives})}{CDC + FN} * 100$$

F-Score (F-Measure): The harmonic mean of recall and precision rates.

$$F = 2.\frac{\text{Precision. Recall}}{\text{Precion} + \text{Recall}}$$

5. LIST OF DATASETS

Listing of Datasets Used by different approaches in recent review papers in Table 2

Table 2: Datasets are used by different Text Extraction approaches

Datasets	Description	Location
MSRA-TD500	Contains 500 natural images, which are taken from indoor and outdoor scenes using packet camera. The resolution images vary from 1296*864 to 1920*1280	http://www.iapr-tc11.org/mediawiki/index.php/MSRA_Text_Detection_500_Database_(MSRA-TD500)
Street View Text(SVT) Dataset	Dataset was harvested from google street view. Image text in this data exhibits high variability and low resolution.	Vision.ucsd.edu/~kai/svt
ICDAR 2003/2005 Robust Reading Competitions	Robust Reading Competition is to find the best system able to read a complete word in a camera captured scenes. The dataset contains 258 training and 251 test images with various sizes from 307*93 to 1280*960.	algoval.essex.ac.uk/icdar/
ICDAR 2011	It includes 485 natural images. The database contains 229 training images and 255 testing images.	http://robustreading.opendfki.de/wiki/SceneText.
Chars74k	This database contains symbols of both English and kannada language and only part of the character In the original images are annotated.	http://www.ee.surrey.ac.uk/CVSSP/demos/chars74k/
Hust-TR400	This database is diverse in both text and background. The images from-1.Images taken by the volunteers, shot in various cities using different devices.2.Images from Flickr website 3.images from MSRA-TD500	http://www.flickr.com/ http://mc.eistar.net/
VACE (Video Analysis and Content Exploitation)	Dataset has 50 broadcast news videos from CNN and ABC. Format of source videos are in MPEG-2 standard, progressive scanned at 720*480 resolution, GOP(Group of Pictures) of 12and frame-rate at 29.97(frames per second)	Marathon.csee.usf.edu/vace-links.html

DOI 10.17148/IJARCCE.2016.5286

6. CHALLENGES

Although lots of approaches are developed on text detection in real applications. However a quick and strong algorithm [18] for detection text under numerous conditions ought to be more investigated. To develop a quick and robust text detection algorithm may be a nontrivial task since there exists such difficulties as:

Text is also embedded in complicated background; it's tough to search out effective features to discriminate text with alternative text-like things, like leaves, window curtains or different general textures; Text pattern varies with completely different font-size, font-color and languages; Text quality decreases due to noise. It's difficult to find text of discretionary orientations.

7. CONCLUSION

This paper provides a study of the varied text extraction techniques and algorithms planned earlier. The proposed system is capable of detecting and recognizing texts of various scales, colors, fonts and orientations. This paper additionally exposed a performance comparison table of various techniques that was projected earlier for text extraction from an image. Each approach has its own advantages and restrictions.

The purpose of our paper is to classify and review of varied recent papers, discuss comparison and performance analysis and to illustrate challenges for future analysis. Several researchers have already investigated text localization, text detection and tracking for images is needed for utilization in real applications (e.g., mobile hand-held devices with a camera and real time categorization systems). A text-image-analysis is required to modify a text info extraction system to be used for any kind of image, as well as each scanned document pictures and real scene pictures through a video camera.

The future work chiefly concentrates on developing an algorithm for actual and quick text extraction from an image

REFERENCES

1. "A Novel Text Detection System Based on Character and Link Energies" IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 23, NO. 9, SEPTEMBER 2014
2. "A Unified Framework for Multioriented Text Detection and Recognition" IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 23, NO. 11, NOVEMBER 2014 4737
3. "Gradient Vector Flow and Grouping-based Method for Arbitrarily Oriented Scene Text Detection in Video Images" IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 23, NO. 10, OCTOBER 2013 1729
4. Characterness: An Indicator of Text in the Wild Yao Li, Wenjing Jia, Chunhua Shen, and Anton van den Hengel IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 23, NO. 4, APRIL 2014
5. "Robust Text Detection in Natural Scene Images" IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 36, NO. 5, MAY 2014
6. "Scene Text Recognition in Mobile Applications by Character Descriptor and Structure Configuration" IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 23, NO. 7, JULY 2014
7. "Scene Text Recognition Using Structure-Guided Character Detection and Linguistic Knowledge" IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 24, NO. 7, JULY 2014 1235
8. "Semiautomatic Ground Truth Generation for Text Detection and Recognition in Video Images" IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 24, NO. 8, AUGUST 2014 1277
9. "Toward Integrated Scene Text Reading" IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 36, NO. 2, FEBRUARY 2014
10. "Automatic Inpainting Scheme for Video Text Detection and Removal" Ali Mosleh, Student Member, IEEE, Nizar Bouguila, Senior Member, IEEE, and Abdessamad Ben Hamza, Senior Member, IEEE, IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 22, NO. 11, NOVEMBER 2013
11. "Morphological Text Extraction from Images" Yassin M. Y. Hasan and Lina J. Karam, IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 9, NO. 11, NOVEMBER 2000
12. Chitrakala Gopalan and D. Manjula, "Contourlet Based Approach for Text Identification and Extraction from Heterogeneous Textual Images", International Journal of Computer Science and Engineering 2(4) pp.202-211, 2008.
13. A Hybrid Approach to Detect and Localize Texts in Natural Scene Images Yi-Feng Pan, Xinwen Hou, and Cheng-Lin Liu, Senior Member, IEEE, IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 20, NO. 3, MARCH 2011
14. Xin Zhang,Fuchun Sun, Lei Gu, "A Combined Algorithm for Video Text Extraction", 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery,2010
15. Chu Duc Nguyen, Mohsen Ardabilian and Liming Chen ,"Robust Car License Plate Localization using a Novel Texture Descriptor", Advanced Video and SignalBased Surveillance, 2009
16. Edge Detection Techniques for Image Segmentation, Muthukrishnan.R and M.Radha, nternational Journal of Computer Science & Information Technology (IJCSIT) Vol 3, No 6, Dec 2011
17. Survey of Region-Based Text Extraction Techniques for Efficient Indexing of Image/Video Retrieval Samabia Tehsin, Asif Masood and Sumaira Kausar, I.J. Image, Graphics and Signal Processing, 2014
18. Y. X. Liu, S. Goto, and T. Ikenaga, "A contour-based robust algorithm for text detection in color images," IEICE Trans. Inf. Syst., vol. E89-D, no. 3, pp. 1221–1230, 2006.
19. Y. Zhong, H. J. Zhang, and A. K. Jain, "Automatic caption localization in compressed video," IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 4, pp. 385–392, 2000.
20. A Survey on various approaches of text extraction in Images C.P. Sumathi, T. Santhanam and G.Gayathri Devi, International Journal of Computer Science & Engineering Survey (IJCSES) Vol.3, No.4, August 2012.
21. V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," Soviet Phys. Doklady, vol. 10, no. 8, pp. 707–710, 1966.