ALGORITHM FOR AUTOMATIC TEXT RETRIEVAL FROM IMAGES OF BOOK COVERS

A Dissertation submitted towards the partial fulfilment of requirement for the award of degree of

MASTER OF ENGINEERING
IN
WIRELESS COMMUNICATION

Submitted by

Niharika Yadav
Roll No. 801363021

Under the guidance of

Dr. Vinay Kumar
Assistant Professor, ECED
Thapar University Patiala

ELECTRONICS AND COMMUNICATION ENGINEERING DEPARTMENT
THAPAR UNIVERSITY
(Established under the section 3 of UGC Act, 1956)
PATIALA – 147004, PUNJAB, INDIA
JULY-2015

DECLARATION

I, **Niharika Yadav,** hereby declare that the dissertation entitled **"Algorithm for automatic text retrieval from images of book covers"** is an authentic record of my own work carried out towards the partial fulfilment for the award of degree of Master of Engineering in Wireless Communication at Thapar University, Patiala, under the supervision of **Dr. Vinay Kumar**, Assistant Professor, Electronics and Communication Engineering Department.

The matter presented in this dissertation has not been submitted in any other University/Institute for the award of any other degree.

Date: 06-07-2015

Niharika

Niharika Yadav

Roll No. 801363021

This is to certify that the above statement made by the student is correct to the best of my knowledge and belief.

Date: 06-07-2015

Dr. Vinay Kumar

Assistant Professor

ECED, TU, Patiala

Countersigned by:

Dr. Sanjay Sharma

Professor and Head ECED

Thapar University, Patiala

Dr. S.S. Bhatia

Dean of Academic Affairs

Thapar University, Patiala

ii

ACKNOWLEDGMENT

With deep sense of gratitude I express my sincere thanks to my esteemed and worthy supervisor, **Dr. Vinay Kumar,** Assistant Professor, Department of Electronics and Communication Engineering, Thapar University, Patiala for his valuable guidance in carrying out work under his effective supervision, encouragement, enlightenment and cooperation. Most of the novel ideas and solutions found in this dissertation are the result of our numerous stimulating discussions.

I shall be failing in my duties if I do not express my deep sense of gratitude towards **Dr. Sanjay Sharma,** Professor and Head of the Department of Electronics and Communication Engineering, Thapar University, Patiala who has been a constant source of inspiration for me throughout this work, and for the providing us with adequate infrastructure in carrying the work.

I am also thankful to **Dr. Amit Kumar Kohli**, Associate Professor and P.G. Coordinator, and **Dr. Hem Dutt Joshi**, Assistant Professor and Program Coordinator, Electronics and Communication Engineering Department, for the motivation and inspiration that triggered me for this work.

I am greatly indebted to all my friends who constantly encouraged me and also would like to thank the entire faculty and staff members of Electronics and Communication Engineering Department for their unyielding encouragement.

At last but not the least my gratitude towards my parents, who always supported me in doing the things my way and whose everlasting desires, selfless sacrifice, encouragement, affectionate blessings and help made it possible for me to complete my degree.

Place: TU, Patiala **Niharika Yadav**

Date: Roll No. 801363021

ABSTRACT

Text extraction is one of the major areas of research in the field of document image Analysis. Text retrieval is needed for bibliographic databases, structuring images etc. Text embedded in multimedia data, as a well-defined model of concepts for humans' communication, contains much semantic information related to the content. This text information can provide a much truer form of content–based access to the image and video documents if it can be extracted and harnessed efficiently.

Moreover, automation of this process will greatly reduce the human interference while converting books (specifically their covers where this task becomes extremely difficult) to readable and editable electronic format specifically for electronic book readers. However this is a challenging task because images contain text of different size, style, orientation, alignment, low contrast, noise and have complex background structure.

This dissertation propounds a method for extracting text from images of book covers and embedded text. A new text model is constructed to retrieve text regions from the scene text images. The image is first clustered to reduce the number of color variances, a suitable plane is identified and then text region is segmented using connected component based method. The text thus obtained is then enhanced to ameliorate the results. A detailed study of sundry techniques that have been proposed so far, along with their performance analysis has also been incorporated in the work.

The algorithm is evaluated comprehensively on various datasets including ICDAR -2011 dataset. The experimental results demonstrate that the proposed text detection method can capture the inherent properties of text and discriminate text from other objects efficiently. The proposed method gives a very high character recognition rate for monochrome images, however in cases where there is a drastic variation in the text features rejection is noticeable.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GROSSARY OF ACRONYMS

AT	Artificial Text
ASM	Angular Second Moment
BB	Bounding Boxes
BGM	Boundary Growing Method
CAMSHIFT	Continuously Adaptive Mean Shift Algorithm
CC	Connected Components
CFR	Conditional Random Field
EBM	Edge Based Method
EMST	Euclidian Mean Square Technique
FCM	Fuzzy C-mean
FFT	Fast Fourier Transform
GLCM	Grey Level Co occurrence Matrix
LEE	Light Edge Enhancement
MCD	Maximum Color Difference
MCE	Minimum Classification Error
MO	Morphological Operation
MST	Minimum Spanning Tree

OCR	Optical Character Recognition
RGB	Red, Blue, Green planes
PMF	Probability Mass Function
PR	Precision Rate
RR	Recall Rate
SE	Structuring Element
SVM	Support Vector Machine
TBM	Texture Based Method
TD	Text Detection
TIE	Text Information Extraction
TL	Text Localization
TS	Text Segmentation

CHAPTER-1
INTRODUCTION

1.1 Motivation

With the advent of e-books and the device to read those books has opened up a new field of research, that is; text recognition. While Gutenberg, openlibrary, Google play books etc provides free classical eBooks for readers at the same time Amazon, Google, etc have an option to buy electronic version of books. The classical eBooks are produced with the help of scanning already available printed material. Optical Character Recognition (OCR) algorithms are then applied to extract the text and generate eBooks.

In recent years, with the rapid development of multimedia technology and computer network, the competence of digital image around the world has increased. Digital images present on the web contain a lot of useful information. These document are generally polychromatic i.e. contain various colors, which in fact is an advantages in separating the required information. Due to the complexity of page-layout, multicoloured or textured background patterns or even background images, the automatic retrieval of relevant information from colored paper documents is a challenging task. In addition, the appearance of text elements may differ not only in color, character font and size, but also in orientation and alignment.

Current computer vision and artificial intelligence technology can not automatically mark the image but must rely on manual annotation for image marking, which is not only time consuming but also often erroneous or incomplete with inevitable subjective bias, that is different people have different understanding methods for the same image, which leads to identification error in image retrieval. This requires a technology that quickly and accurately searches for and access images.

There arise two major problems during this automated process of ebook generation

1. Recognition of the text in the exclusively text areas, and
2. Recognition of the text embedded in the image; for example at the cover of the book, on the sides of the book, in pictures, etc.

There exists number of solutions for the first task [1-5]. There exist a few solutions for the second task [6-10] .The current manuscript deals with problem faced in second situation.

1.2 Text Features

A number of variations are exhibited by the text present in the images with respect to the some properties. Most of them follow from the research done in the image analysis.

- **Size:**

  Depending upon the application, the size of text in an image is assumed. This is done as there is a lot of variation in the text size. Size of the text is set to a readable range and thus is always above a defined value however the upper bound of the text is loose. For example, in Figure 1.1 the size of the caption is variable.

- **Alignment:**

  Generally text is aligned horizontally in caption images, however sometimes due to addition of some special effects it can appear non planar as well. In case of scene direction of the text is uncertain and can have distorted geometries.

- **Color:**

  The variation in the color of the characters of a text line is miniature, they are almost equivalent. Connected component based approaches use this characteristic in their advantage for text detection. Enormous work has been reported about the images that are monochromatic. However such is not the case in videos and scene images. They contained two or more colors for better visuals.

- **Motion:**

  In case of videos motion is one parameter that is used for tracking. In the consecutive frames of a video characters do not change significantly. This property is well utilized in algorithms of text tracking. Movement or motion of text is also a criterion, while caption text moves either horizontally or vertically, movement of camera or the object itself leads to the arbitrary orientation of the scene text.

- **Rigidity :**

  Text strings or the caption text maintain their shape, orientation, size and font over multiple frames in case of videos. However this may not be true in cases where text has special graphic features incorporated. Generally this feature is maintained in several images.

Figure 1.1: Image with caption text

- **Edge:**

  Strong contrast edges are defined for the text regions of the image so that the text is easily readable in case of caption text. Hence text extraction could be made more efficient by identifying these edges.

- **Compression:**

  Generally digital images are stored, exported, and processed in a compressed format. Thus, a competent text extraction system is developed if text retrieval is performed without performing decompression.

Table 1: Properties of an image

Property		variant
Geo-Metry	Size	Constant text size
	Alignment	Horizontal/vertical
		Skewness in lines
		Curves
		3D perspective distortion
	Inter-character Distance	uniform distance between characters
Color		Gray
		Color (monochrome, polychrome)
Motion		Static
		Linear movement
		2D rigid constrained movement
		3D rigid constrained movement
		Free movement
Edge		High contrast at text boundaries
Compression		Un-compressed image
		JPG, MPEG-compressed image

1.3 Text Classification

Image text is categorized in two ways: one is called the scene text and the other is called artificial text

SCENE TEXT

Scene text refers to the two-dimensional image of scene texts in the real three-dimensional world received by a camera or video camera, such as the license in car

pictures, shop signs that occasionally appear on the video screen or texts on street advertisements.

Uneven illumination, camera or video camera distortion, lack of exposure and narrow dynamic range, titled shooting angle, uneven surface of three-dimensional texts and various degrees of pollution and other reasons will all lead to low-quality images [11]. Since character size, font, color, direction and background texture of the text is a priori unknown, it is difficult to extract and identify this type of text.

Scene images contain the text, which is obtained naturally when the scene images are taken by the camera, Complexity of the scene images is due to the presence of varying features such as the styling, alignment, sizes etc ,these features leads to a complex background, meanwhile the resolution of the images is low. Morover, scene text is generally affected by lighting conditions and perspective distortions. The current OCR software available is unable to handle the inference of the background and complex text. Therefore recognition of the scene text directly is impossible. Text extraction from natural scene images is full of challenges.

ARTIFICIAL TEXT

Artificial text is a man-made text, with more criterion features and purposes, such as the artificial text in synthesized images and the title sequence, epilogue and subtitles in videos [12]. It can be seen that compared to scene texts in videos, this kind of text has more noteworthy content, which makes it easier to detect and identify and play an important role in video retrieval. Therefore, we mostly use artificial text for the processing of video texts.

In order to have a more detailed understanding of text location and extraction in complex background, Figure 1.2 shows the dissimilarity in image source, color, background complexity and application between scene text and artificial text. (Note: the text location and extraction within a single frame of dynamic images is basically the same with that of static images.

Figure1.2: Scene text image

Figure 1.3:Multi-color document images

Table-2: Text Classification

Classification	Scene text	Artificial text
Sources of image	Static images (including digitized pictures, pictures received from digital camera or scanner)	Dynamic images(video sequence or film documents),synthesized images on www
Color	Grayness or color	Color
Complexity of background	Spinning, tilted, partially hidden, uncontrollable light	Both background and text can be mobile
Field of application	Color image database, car license identification	Content based video search, retrieval and web search

1.4 Text Information Extraction

The predicament of text information extraction needs to be defined before going any further. A TIE model has five stages. All the stages are explained thoroughly in the next chapter. Here we give a brief overview of the process. Input to this model is any image, generally a scanned is considered, and the input image can be colored image or a grayscale image. In the literature a variety of images have been considered. As described in the previous section. These images can be caption images, digital born images or naturally captured images called the scene text images.

The input image undergoes the following process i) detection ii) localization iii) segmentation iv) tracking v) recognition.

Text detection is the process of determining the existence of text in the image. This is generally suitable for sequence of images. Text localization is basically used to obtain the location of the text region in any image. Text tracking is performed in case of videos; it reduces the text localization time and also maintains the integrity of the position of the

text region. Text segmentation is needed to facilitate text recognition as segmentation separates the text from its background. Text enhancement is performed to improve the results as text regions are some times of low resolution and prone to noise.

1.5 Scope of the Dissertation

In this dissertation, we discuss the various stages of Text Information Extraction model including preprocessing, text detection, detailed study of the region based segmentation approaches etc in the Chapter-2.

In Chapter-3, we review the literature of the text extraction approaches that includes algorithms for clustering, text detection and text segmentation etc. A detailed description of all the text segmentation methods that includes, CC-based method, Edge based method, Texture based method, Morphological based method etc. Along with the methodologies used in the past a performance analysis of these methods is also described in this chapter.

In Chapter-4, we describe the proposed method for text retrieval from the scene text images. In the pre processing stage, color variation is reduced using histogram based clustering. In the next stage the text segmentation performed using the connected component based technique has been explained. Noise in the form of figures, non-text regions or logo in the book covers are removed using MOs. Final stage of the algorithm briefs about the text identification process. A fresh way of displaying the recognized text is worked upon and implemented on the various datasets.

In Chapter-5, we display the investigational results obtained after applying the proposed method on a various datasets, such as ICDAR 2011 etc. Performance analysis of this technique is also discussed in this chapter using the parameters of text localization i.e. recall rate, precision rate, f-mean etc.

Chapter-6, gives concluding remark of the dissertation, explaining all the finding and observations along with the improvements to consider for the future.

CHAPTER-2

TEXT INFROMATION EXTRACTION MODEL

Text retrieval process is defined generally using a text information extraction model. The input to this model is any RGB or grey digital image, image could be compressed or uncompressed, and text in the image can be motionless text or moving one**.** The model can be alienated into mentioned stages:

a) Text detection
b) Text localization
c) Text tracking
d) Text extraction and enhancement
e) Text recognition(OCR)

Figure 2.1 shows a generalized text information extraction model. A brief explanation of all the stages is given in the next section.

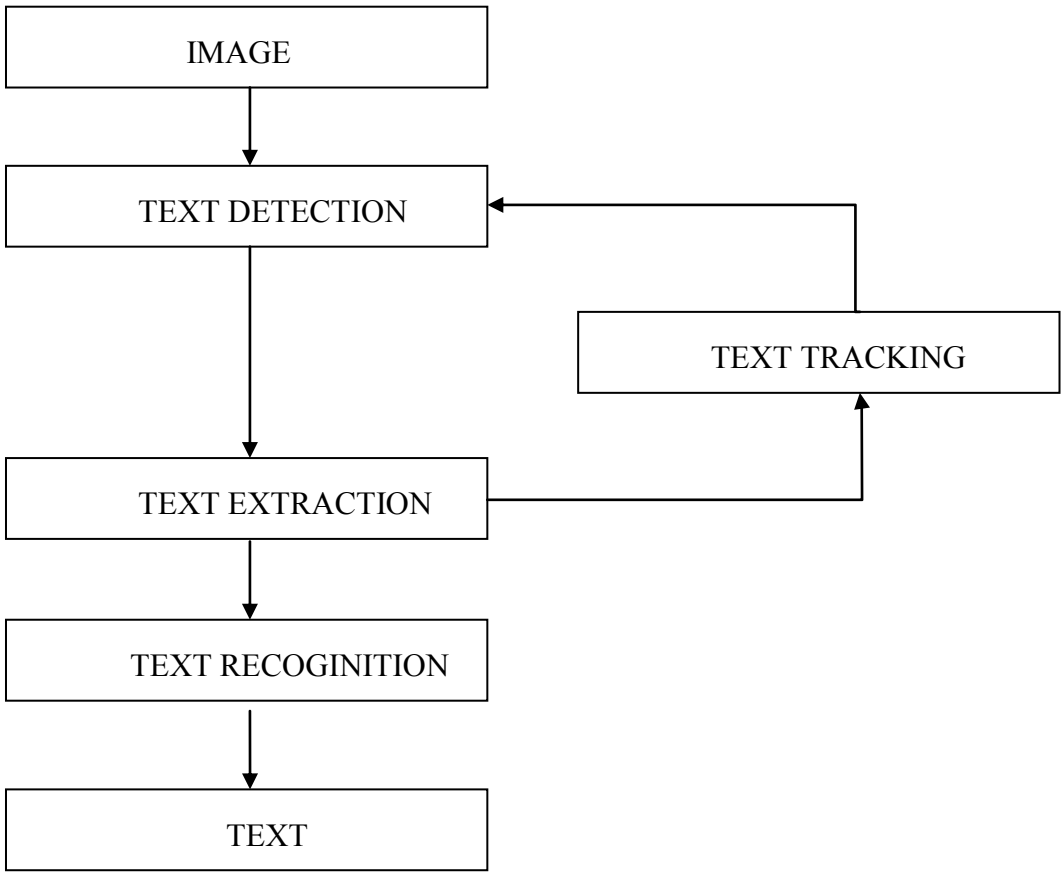


Figure 2.1: Text information extraction model

2.1 Text Detection

Text detection is the most fundamental step, it is significant as, in this step decision about the existence or non existence of the text in the image is made. With the researches going on it has been identified that certain type of text contains text, for example it is generally assumed that scanned images of book covers or compact disk covers will be text images.

The text detection stage seeks to perceive the occurrence of text in a given image. Many small color variations are shown by a color picture after digitization that shows minimal variation with naked eyes. These small not useful variations need to be eliminated. This is performed during the preprocessing stage. should be eliminated in a pre-processing step. Clustering algorithms are used for pre-processing. Depending on the clusters, a pre-processed image is computed by categorizing pixel depending upon the similarity with the original colors.

Clustering is a process for classifying objects or patterns such that samples of the same group are more similar to one another than samples belonging to different groups [13]. Various clustering techniques [15-17] have been proposed so far, each of which has its own special characteristics, some of the techniques of interest are, Euclidean Minimum-Spanning-Tree (EMST) [15] based clustering requires a pre defined threshold, Fuzzy C-Means (FCM) [16] is a fuzzy clustering method, it is a soft segmentation method, k-mean clustering [17] is a partitioned clustering technique that clusters objects based on attributes into k partitions. Histogram based clustering is a hard clustering approach. According to the application these techniques are used for different purposes. In the present approach we used histogram-based clustering technique. This clustering is based on spatial and feature space clustering [18] and significantly reduces the number of colors. It has the advantage that no critical threshold has to be defined; moreover it requires no prior knowledge of the expected number of clusters, which is required to solve the problem at hand.

2.2 Text localization

Text localization is the process of obtaining the location of text in the image and marking the text regions with the help of bounding boxes around the text [19]. Text localization methods utilize the features and other properties of the text and casiffy the methods in three ways: region-based methods, texture-based methods and hybrid methods.

2.2.1 Region-based methods

The properties of an image discussed in the previous section plays a crucial role in deciding various methods of text localization. One major property of any image is the color or the greyscale of the text region in an image and its variance as compared with the background of the image.

These methods can furthermore be fragmented in two sub-categories: connected component (CC)-based and edge-based. These approaches work in a bottom-up fashion; sub-structures, such as CCs or edges, are identified and then the text regions are marked with the help of bounding boxes.

a) CC based method

Small similar components of the input images are grouped to form successively large components using bottom up approach in the case of CC-based methods. This process iteratively continues until all the regions of the image are not identified.

Filtering of the non text regions of the image is performed by considering a geometrical analysis which merges the components using the spatial arrangement of the components and mark the boundaries of the text regions.

Due to their comparatively straightforward execution, CC-based methods are widely used. Generally the CC-based methods work upon four basic: (a) preprocessing (b) Generation of the CC's, (c) removal of non-text regions, and (d) Grouping of text regions. A CC-based method gives promising results for polychrome text strings with low resolution as it is able to defragment the image into multiple [20]. In addition, numerous threshold values are required so as to eliminate the various non text areas. These threshold values are dependent on the input image being considered decides the value of the threshold to be selected.

b) Edge-based methods

Edges are those portions of the image that corresponds to the object boundaries. There is a major contrast in the text and the background and this drastic variation forms the edges.

The non text region of the images is filtered out by firstly identifying the edges of the text boundaries and then applying various heuristics [21]. Generally, an edge filter is

incorporated for the identification of the edges in an image, and for the merging stage an averaging operation or a morphological operator is used.

Canny edge detection algorithm:

Canny edge detection algorithm is also known as optimal edge detector. This algorithm was designed to satisfy three criterions [22]. First criterion is a low error rate and removal of all the unwanted information while keeping the useful information intact. Second criterion to keep minimal variation in the original and proceeds image. Third criterion is to eliminate multiple responses to any edge.

Following this criteria, the algorithm first removes the noise element in the image. This is accomplished by performing convolution of the image with a using Gaussian filter; example of such filter is given in Figure 2.2. Regions having high spatial derivatives are than identified by calculating the image gradient. Sobel operators are used to calculate the gradient. In the next step a non-maximizing suppression is used to suppress any pixel which is not at its maximum [22]. Hysterics is then applied to further reduce the gradient array so as to remove streaking and thinning edges.

1/11

2	4	5	4	2
4	9	12	9	4
5	12	15	12	5
4	9	12	9	4
2	4	5	4	2

Figure 2.2: Gaussian filter

2.2.2 Morphological based methods

Mathematical morphology is a topological and geometrical based approach for image analysis. This method efficiently extracts geometrical structures from the image and helps representing different shapes depending upon the applications. It uses various operations

to perform this task [23]. Character recognition and document analysis uses the morphological features for the extraction of data and it has been very resourceful. It is useful in retrieving important text features from images.this method gives promising results because operations like translation, rotation and scaling do not have any effect on the geometrical shape of the image. Change in text color and the light conditions also do not have any effect on the features .This method works robustly under different image alterations

A brief description of some of the popularly used MOs is given below;

a) **Erosion:** This operation erodes away the edges or the boundary region of foreground pixels, in simpler words the white pixels. This leads to shrinking of the area of foreground pixels, which in return enlarged the holes present within the area [24].The erosion of a binary image f by a structuring element S (denoted by f Θ S) produces a new image g which consists of all one's located in structuring element's origin at which that structuring element S fits the input image f.
b) **Dilation:** This operation enlarges the edges or boundary region of the foreground pixels. In case dilation the structuring element consists of all one's. The new binary image formed by a structuring element S (denoted f ⊕ S) gives a output 1 where the structuring element coincides the input image *f*, i.e. g(x,y) =1, otherwise value is 0.Dilation is a reverse process of erosion; a layer of pixels is added to both the inner and outer boundaries of regions. Figure shows the two processes.
c) **Opening:** Opening is hybrid process and consists of erosion followed by dilation and this operation is capable of removing pixel values that are too diminutive to contain the structuring element.

$$I(x,y) \circ S_{m,n} = (I(x,y) \odot S_{m,n}) \oplus S_{m,n} \qquad (1)$$

a) Greyscale image

c) binary image by thresholding

b) Dilation

d) Erosion

Figure 2.3: Morphological operations [24]

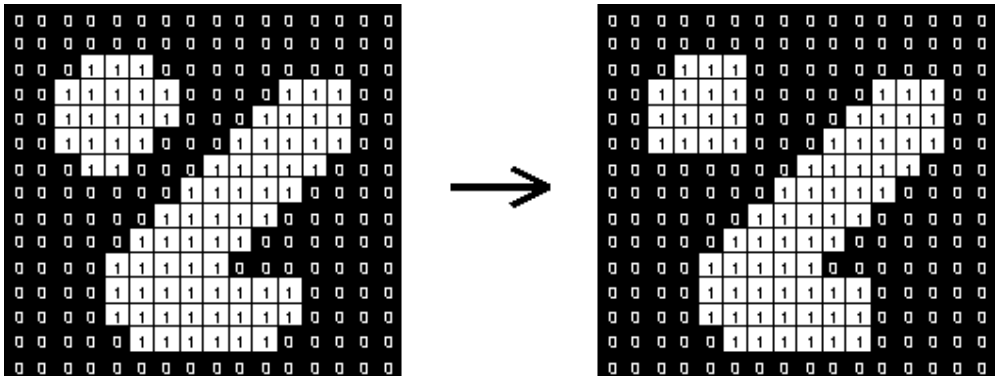


Figure 2.4: Effect of opening using a 3×3 square structuring element

d) Closing: Closing consists of a dilation followed by erosion and can be used to fill in holes and small gaps. The closing operation has the effect of filling in holes and closing gaps.

$$I(x,y) \bullet S_{m,n} = (I(x,y) \oplus S_{m,n}) \odot S_{m,n} \quad (2)$$

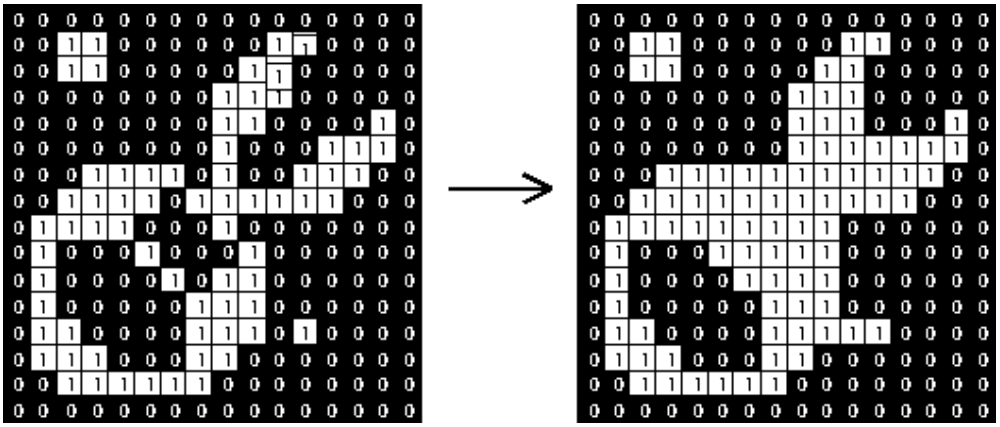


Figure 2.5: Effect of closing using a 3×3 square structuring element

2.2.3 Texture-based methods

Texture-based methods use the inspiration that texts in images have distinctive textural properties that distinguish the text and the background [25]. The textural properties of the text regions can be perceived by using techniques based on Gabor filters, Wavelet, FFT, spatial variance, etc. The following parameters define the texture of an image [26].

- *Energy*: This value also called Uniformity or Angular Second Moment. ASM describes image smoothness. It calculates the textural uniformity that is repetition in the pair of pixels.
- *Contrast*: Contrast evaluates local gray level variation. This statistic measures the spatial frequency of an image and is a difference moment of GLCM
- *Inverse Difference Moment:* This statistic is also called as homogeneity. It considers the homogeneity of the image by comparing different grey tone pair elements.
- *Entropy:* This is a measure of the complexity of image. A non uniform textural image that has many small valued GLCM elements will have high entropy.

The predicament with conventional texture-based methods is they require more processing time, which is due to the computational complexity in the texture classification stage. A thorough scan of the input image is required by the Texture-based filtering to detect and localize text regions [27]. This leads to the running of a convolution operation, which is a very expensive operator. A number of solutions have been proposed for this. However an effective method is to classify the pixel on regular intervals and then perform interpolation to obtain the missing pixels. However, this is also not a very

effective method.

2.4 Performance Analysis

In order to calculate the robustness and performances of the proposed algorithms two methods are defined.

The Precision rate is the ratio of words that have been correctly detected to the sum of correctly detected words plus false positives. False positives words are those area of the image that were not text but were detected as text region by the algorithm implemented.

$$\text{Precision rate} = \frac{\text{correctely detected characters}}{\text{correctely detected characters} \times \text{false positive}} \times 100 \quad (3)$$

The Recall rate is defined as the ratio of words correctly detected to the sum of correctly detected words plus false negatives. False Negatives are the regions of the mage that have text but the algorithm used is unable to identify them.

$$\text{Recall rate} = \frac{\text{correctely detected characters}}{\text{correctely detected characters} \times \text{false negative}} \times 100 \quad (4)$$

CHAPTER-3
LITERATURE REVIEW

This chapter presents a survey of the previous work done in the field of text extraction. A detailed study of various methods and algorithms has been done to formulate this work. A variety of work has been done on the artificial text present in the images. Here however we are studying the scene text images and approaches that have been used to successfully detect, localize and extract these texts from the images of book covers or book stacks. In this chapter we have presented the literature of the Text extraction model explained in the Chapter-2.

3.1 Preprocessing Techniques

Many clustering strategies have been used for the purpose of text retrieval, such as the hard clustering scheme and the fuzzy clustering scheme, each of which has its own special characteristic. As a result of which this type of clustering gives crisp results. Some of the work done in the progress of these techniques is explained below.

Fuzzy c-means (FCM) algorithm [15] is used in the preprocessing because it has stout features for uncertainty and is able to store much more information as compared to other hard segmentation methods. The algorithm uses an iterative clustering approach that part ions the by minimizing the weighted within group sum of squared error group functions.

k-mean clustering [16] is similar to the expectation-maximization algorithm implemented for the mixtures of Gaussians where they both challenge to locate the centers of natural clusters in the data. In this algorithm to cluster n objects, k partitions are formed based on some attributes, where $k < n$. It considers that the object attributes form a vector space.

Histogram-based clustering technique [17] drastically reduces the number of colors. Compared with other clustering techniques, e.g., the Euclidean minimum-spanning-tree (EMST) or fuzzy c-means (FCM), histogram-based clustering has the advantage that no critical threshold has to be defined. For the EMST [15] technique, for example, a distance has to be defined to eradicate edges in the EMST. Furthermore; the histogram-based clustering technique is an unsupervised method that needs no a priori acquaintance about the number of expected clusters as is often necessary for FCM clustering techniques. Moreover, histogram-based clustering is a fast technique with reasonable storage requirements.

3.2 Connected Component Based Methods

Ohya *et al*. [28] proposed a method that was subdivided in following satges: (a) thresholding for binarization, (ii) using the grey level of the image to identify the differences (iii) Character identification to compare the results with the already presented values. and (iv)Operations to perform relaxation. in addition, there are few restrictions to this approach related to the text alignment as well as the color of the text. Experimental results showed a recall rate of text localization on 100 images of 85.4% while the character were correctly recognizes at a rate of 66.5%

Sobertte *et al.* [29] compared two CC-based methods with respect to their processing time, precision rate recall rate. Dataset for this experimental setup was images of books covers and journals. He gave a text segmentation method that used a start region, pixels were merged if they belong to the region else were discarded, this let to formation of homogenous regions. These regions were surrounded by rectangular boxes known as the bounding boxes. This approach is bottom up analysis later explained in Chapter-3. Method was applied on variety of images including, images of book cover and journals. Results of the approach are shown in Figure 3.1.

Zhong *et al.* [30] implemented an algorithm that utilized a CC-based method that incorporated color reduction. The peaks of the color histograms were quantized to achieve the reduction. This was based on the fact the quantization will regroup the text containing regions and a significant portion if the image will be covered. Filtering of the text regions was then performed by passing the reduced image through various filters of area, alignment etc. The inputs for this algorithm were images of CD covers and Book covers.

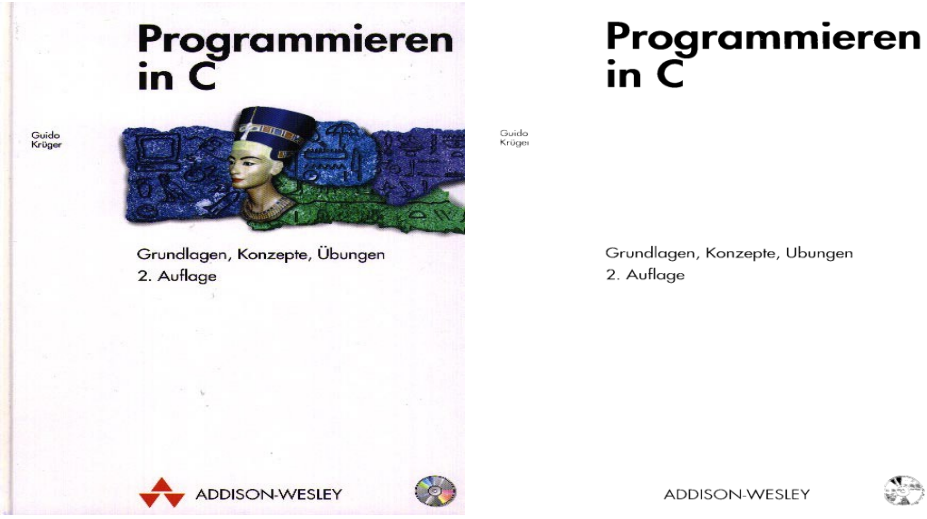


Figure 3.1: Result of text localization

Kim *et al.* [31] segmented an RGB image using color clustering. All the Non-text components including long horizontal lines and image boundaries were eliminated. an iterative projection profile analysis was then implemented to extract horizontal text lines and text segments. In the post-processing stage, different heuristics were used to combine these regions.

Shim *et al.* [32] proposed to consider the homogeneity of pixel intensity for the text regions of the images. Similar grey level pixels were merged to form groups. Once these groups were formed they were identified as backgrounds and thus removed. The grey level contrast was then used to perform region boundary analysis that sharpened the text regions. The applicant regions were then subjected to authentication using features such as size, area, fill factor, and contrast.

Kim *et al.* [33] proposed an approach to filter the non-text regions for the hetrosegment characters to alleviate the difficulty level of the heuristics that are defined to filter the non-text regions. **Ohya *et al.*** [28] also gave a similar approach. Geometrical properties such as alignment, area, size etc are used along with the clustering. These cluster-templates are developed by implementing the clustering technique defined as the k-mean from the actual images.

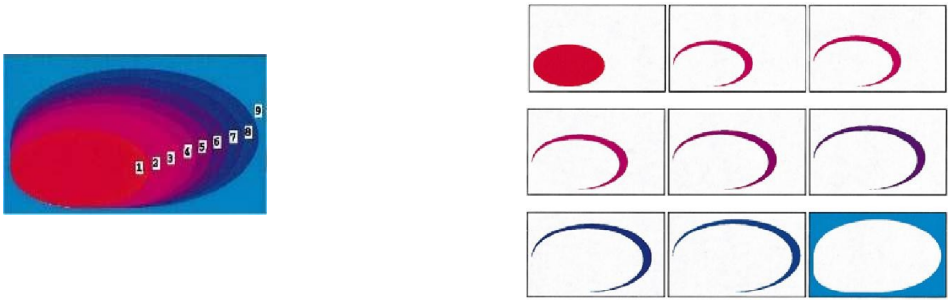
Jain and Yu [35] applied a model that included following stages. In the first stage, preprocessing, which includes bit dropping a 24-bit image was bit dropped to 6-bit

image., in the second stage of color clustering, image was quantized After the decomposition of the image in multiple foregrounds ,same localization technique is implied on each foreground image.

Figure 3.2 shows an example where decomposition of the multi-valued image. CCs are generated in corresponding for all the foreground images using a block adjacency graph. An output image is formed by merging the localized text components in the individual foreground. Horizontal and vertical text is efficiently extracted using this algorithm however such is not observed in case of skewed text.

3.3 Edge Based Methods

Messelodi and Modena [36] gave a method that has three stages: (i) extracting the elementary objects, (ii) filtering and (iii) selecting the text line. In the preprocessing stage noise reduction, deblurring, contrast enhancement, quantization, etc., are performed



(a) Color input image (b) Nine element images

Figure 3.2: A multi-colours image and its element images

After the preprocessing, pixel intensity was normalization and then the image binarization was performed, and CC generation were performed. The decision regarding the selection of filters is based on the value of thresholds and the applications. The approach of text line selection is initiated by a single region and then it is expanded until the criterion for termination is satisfied. The criterion uses features using such as closeness, alignment, and comparable height. The algorithm was applied on various images of book covers having different colors, sizes and orientations. In addition, this can also be applied to images that have variable skewness. The method achieved a localization rate of 91.2% when experiments were performed on a dataset of 100 images.

Smith and Kanade [37] worked on an edge-based method that to identify the vertical edges in the image. To serve the purpose they applied a 3×3 horizontal differential filter to an input image and perform thresholding .After a smoothing operation that eliminated small edges, adjacent edges were connected and a bounding box were computed. Then the non text regions were filtered by applying heuristics. In the final stage, the histogram value of the cluster is analysed to obtain the similar clusters.

Chen *et al*. [38] implemented canny edge operator that identified the high contrast regions of the image. In order to reduce the computational complexity a single edge point was used in one window for the estimation and orientation. The information obtained is then used to enhance the image MOs were performed on the image to connect the edges into clusters. Two types of Gabor asymmetric filters were implemented to generate a general estimation: an edge-form filter and a stripe-form filter. The obtained information about the edge is then enhanced using a suitable scale. This resulted in the blurring of the area that had no specific scale. Image enhancement was then obtained using the localization techniques.

Xiaoqing Liu [39] used a method that is sub divided in three steps: Text region identification, localization and text extraction. In the first step the strength of edges is measured using the magnitude of the second derivative of intensity. The average edge strength is used to calculate the density of the edge. Considering effectiveness and efficiency, four orientations $(0°, 45°, 90°, 135°)$ are used to evaluate the variance of orientations, where 0° denotes horizontal direction, 90°, denotes vertical direction, and $45°$and $135°$ are the two diagonal directions. In the second step, text localization is performed implying clustering techniques. In the third step, already proposed existing OCR engine where used, that were only efficient with printed documents having plane background and are unable to give promising results in case of text embedded in the image or in case of complex backgrounds..

a) Original image b) Result after extraction

Figure 3.3: Edge detection results [39]

3.4 Texture Based Methods

Zhong *et al.* [40] gave a method that property of variance in an image to locater the text regions of the image. He considered a horizontal window of size 1×21 that calculated the spatial variance of the pixels. Canny edge detector was then applied to identify the horizontal edges. From resulting image, edges that have opposite directions are paired into the lower and upper boundaries of a text line. However, the method was only able to identify the components with significant variations. The method was able to achieve a promising processing time of 6.6 second with a 256×256 image on a SPARC station 20.

Kim *et al.* [41] proposed a method that could detect texts using texture properties. A support vector machine (SVM) was implemented to study the properties of textural basis of the image. The method used the raw pixels of the image that confided with the textural pattern rather than adding new external texture feature extraction model. This worked fine even for the high dimensional spaces. Next step of the method was to implement, continuously adaptive mean shift algorithm (CAMSHIFT) that could locate the text regions by using the analysis of the texture properties. The amalgamation of CAMSHIFT and SVMs produced very resourceful results as text detection was significant. The performance of the system was based on the criterion that the SVM was able to classify the tet and non text regions and not on the final text detection. The author used a set of 100 images that were divided in two sets. One set had 70 images that were used for training and the other 30 were used to validate the results obtained.

Shiva kumara *et.al* [42] proposed a new method that was based on Maximum Color Difference (MCD) and Boundary Growing Method (BGM). The method was

implemented to obtain text from handwritten scene text. To sharpen the edges of the original frame and to increase the contrast of text pixels, calculation was of mean value of the RGB channel was performed. To raise the gap between the text and the non text region of the image Maximum Color Difference was computed. k-means clustering algorithm was used to cluster the text .These clusters were used to obtain the text candidates and also help in eliminating false positives. The boundary of the text region was setup using the nearest neighbour approach implanted by Boundary Growing Method (BGM). To eliminate the false positive edges the researchers used the concept of intrinsic and extrinsic edges.

The unique approach of **Shyama *et.al*** [43] projected a text segmentation technique to extract text from any image or video captured with a camera. To link consecutive pixels in the same direction Colour based segmentation was used which exploited the general text properties. Light Edge Enhancement (LEE) was used to locate the regions having consecutive pixel and to enhance the underlying edges. Next, the motion blur was removed with the help of heavy edge enhancement (HEE) from camera image sequences.

Pan *et.al* [44] implemented a hybrid method that implanted a text region detector. This detector could generate a text confidence map. Text components are segmented using a local binarization approach which uses 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. An iterative learning based method is worked upon 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.

3.5 Morphological Method

Jui-Chen Wu [45] presented a morphology-based text line extraction algorithm for extracting text regions from cluttered images. A novel set of MOs was implementing for extracting important contrast regions as possible text line candidates. The skewness of the text lines was detected using a moment based method. Depending upon the orientation, an x-projection technique is implemented to obtain the geometries that are analogous to segments for verification. However, fragmentation of the text region in to different segments is performed due to noise. Therefore, after applying projection, to obtain a complete text line a recovery algorithm is required. Hence a new recovery algorithm was

proposed. Subsequently, after the recovery verification is required. For this purpose a verification scheme was proposed that was able to perform verification of the text geometries obtained. The performance analysis of the method was performed by taking in consideration a set of 100 images. After testing this method, these images have various appearance changes like contrast changes, complex backgrounds, lightings, different fonts, and sizes.

Luz *et.al* [46] introduced a method for localizing text regions within scene images by defining a set of probable text regions which were extracted from the input image using morphological filters. Connected Components (CC) were identified using ultimate attribute openings and closings. The subset was selected by combining the CCs, the non-text regions were distinguished by the decision tree classifier.

(a) (b) (c)

Figure 3.4: Result of text line detection [45]

CHAPTER-4
METHODOLOGY

4.1 Text Information Extraction Model

This dissertation defined a generalized text information extraction model that included various stages in the Chapter-2. In this Chapter we introduce the modified text extraction model that specifically coincides with the algorithm we are proposing for the text extraction. Figure 4.1 shows the model defining all its stages. Raw image are the input to this model and after passing through various stages finally text is obtained from the image.

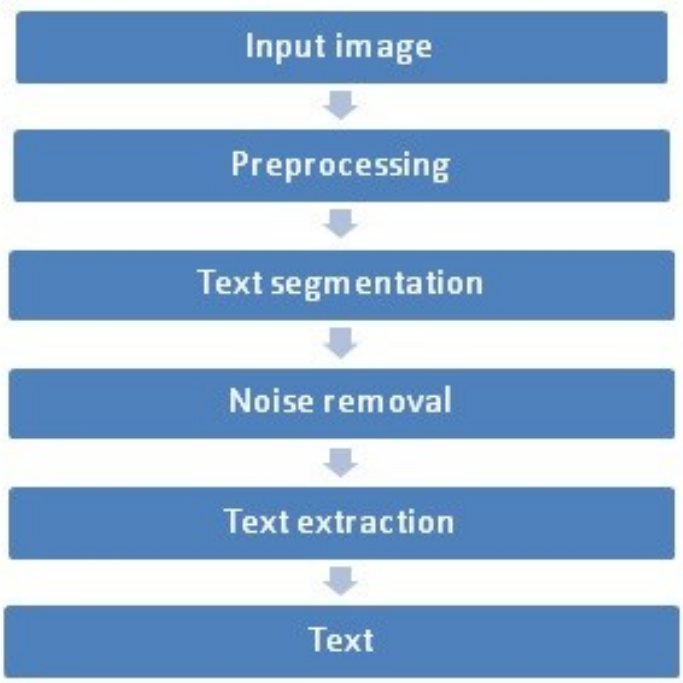


Figure 4.1: Text information extraction model

4.2 Preprocessing technique

The images typically have a size of about 1200 × 1600 pixels and include thousands of colors due to digitalization. To reduce the amount of small variations in color, color clustering is applied in a preprocessing step. Histogram-based clustering technique significantly reduces the number of colors. Compared with other clustering techniques, e.g., the Euclidean minimum-spanning-tree (EMST) or fuzzy c-means (FCM), histogram-based clustering has the advantage that no critical threshold has to be defined, moreover it requires no prior knowledge of the expected number of clusters. Hence it is a fast technique with reasonable storage requirements.

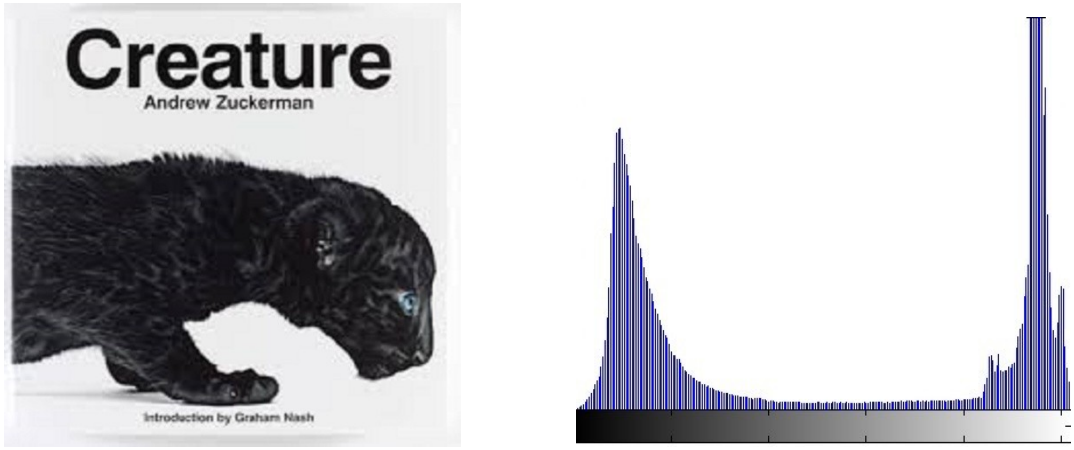
In the present approach we have used histogram-based clustering technique, this type of clustering is based on spatial and feature space clustering [15], and this technique significantly reduces the number of colors. It has the advantage that no critical threshold has to be defined; moreover it requires no prior knowledge of the expected number of clusters. [17] It is a fast technique with reasonable storage requirements.

4.2.1 Clustering

Consider an image I_{MxN} defined by Equation-5

$$I_{M\times N}(x,y) = (I_R, I_G, I_B) \; \forall \; x \in \{1, M\}, y \in \{1, N\} \tag{5}$$

Where $I_R(x,y)$,$I_G(x,y)$,$I_B(x,y)$ are intensity of pixel (x,y) in the red, green and blue plane respectively.



a) Image b) P_{pmf} of image

Figure 4.2: Example of probability mass function

Probability mass function (P_{pmf}) of the individual planes of the image is obtained, using Equation-7. Figure 4.2 shows the image and it's PMF. Now each P_{pmf} cell value is compared to its neighbouring cell; the cell is pointed the value that is maximum (n_{max}), refer Equation-8 in the neighbourhood, in this way a local maximum is calculated for all the cell values of the histogram. For any RGB image applying this technique reduces the color variation significantly. Figure 4.3 shows the clustered histogram.

$$R(\mathrm{i}) = \mathrm{card}\{(\mathrm{u},\mathrm{v})|\mathrm{I}(\mathrm{u},\mathrm{v}) = \mathrm{i}\} \forall \; 0 < i < k \tag{6}$$

Where k is no. of different types of gray scale values, for an 8 bit image k=256.

$$P_{pmf}(i) = \frac{R}{MN} \tag{7}$$

$$n_{max} = \max\left(P_{pmf}(j)\right) \tag{8}$$

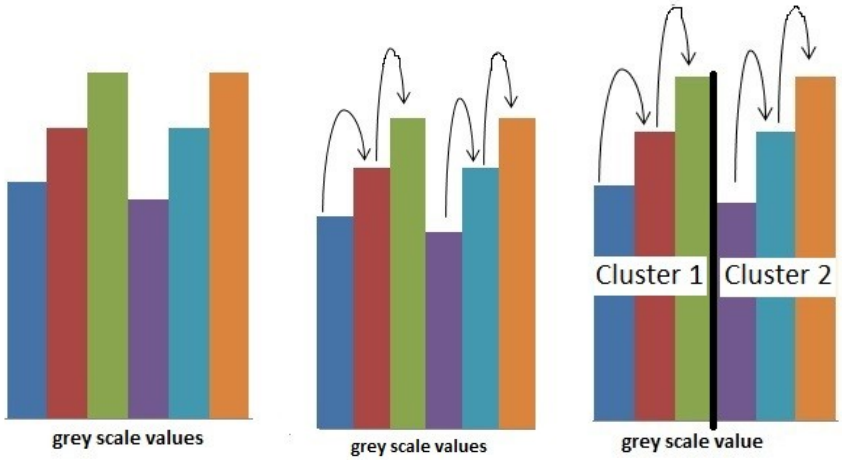
Where R is histogram array, j= $\{((\alpha - 1) * 3): ((\alpha - 1) * 3) + 3\ \}$,

$\alpha \in \{1,2,3 \ldots \ldots\}$

Individual RGB plane clustering will lead to three clustered histograms $R_{H.}$

$$R_{H,i} = card\{j | P_{pmf}(j) = \ n_{max}\} \tag{9}$$

Figure 4.4 to 4.6 shows an example of non clustered and clustered histograms for the three planes for a RGB image. The next task is to identify the best plane for extraction of text from its background.



a)Histogram b)Chain of pointers c) Clustered histogram

Figure 4.3: Example of histogram clustering

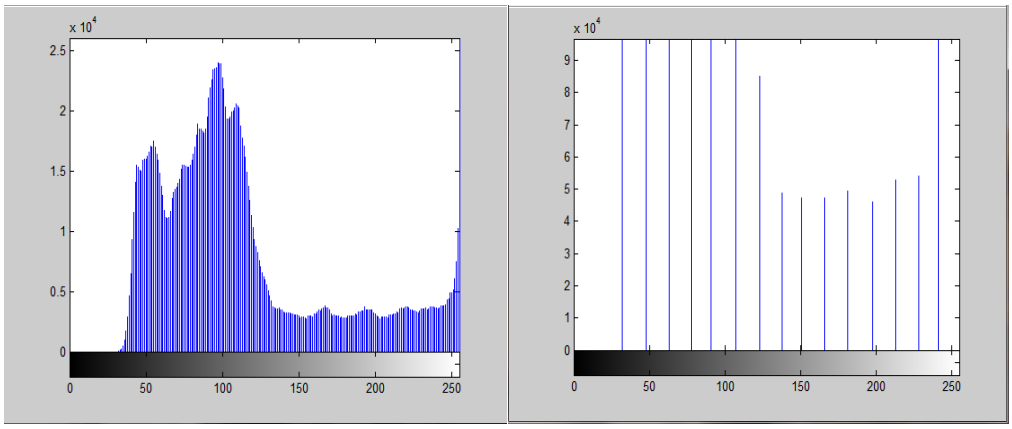


Figure 4.4: Histogram clustering: R-plane

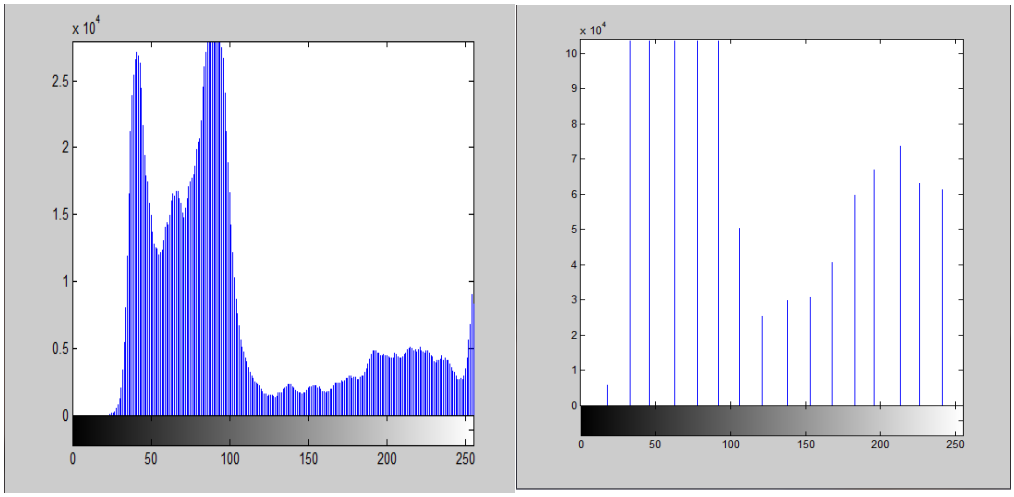


Figure 4.5: Histogram clustering: G-plane

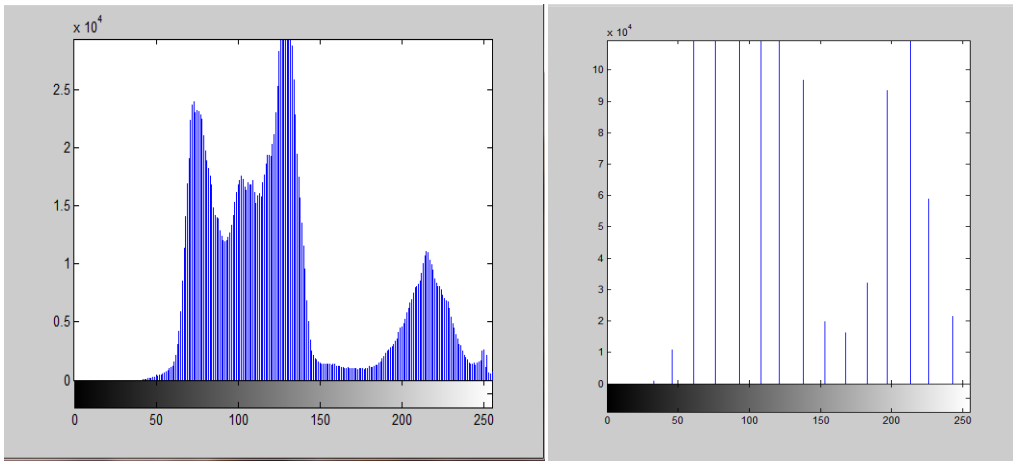
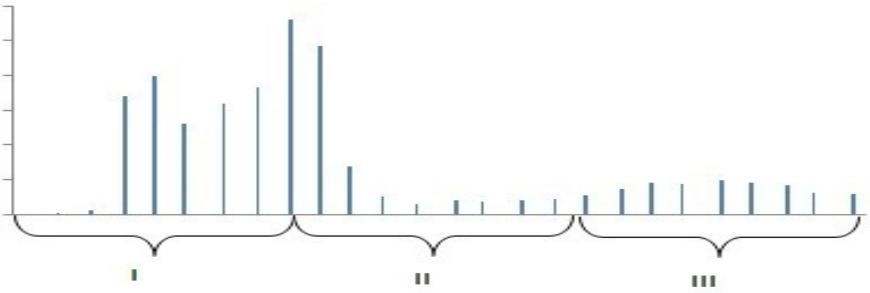


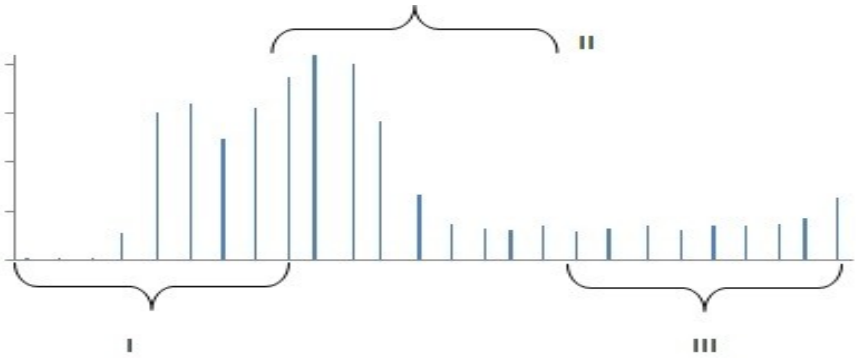
Figure 4.6: Histogram clustering: B-plane

4.2.2 Best Plane Identification

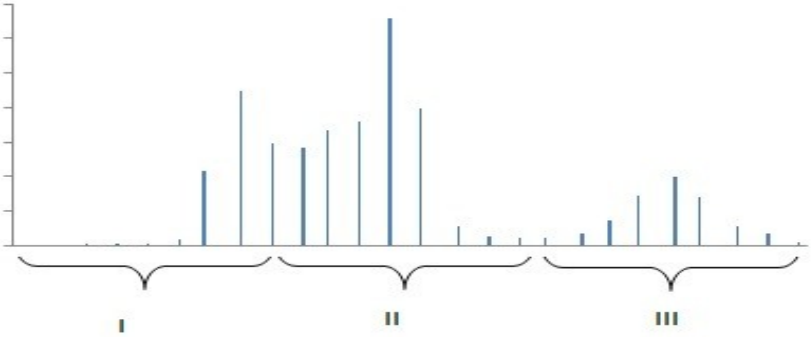
Any text in an image can be separated only if the background and the text are differentiable; i.e., the diachronic interpretation of background with respect to text. Amongst I_R, I_G, I_B, maximum variant plane will serve the purpose. To calculate this variation, an intra plane deviation of the distribution of pixels of each plane in the clustered histogram (refer Equation- 9) is calculated. This deviation (θ_i) can be found with the help of Equation-10. Equation-11, calculates the maximum variant plane. This can be better understood from the example given in the Figure-4.7.The clustered histogram of each plane is divided in three equal regions and using Equation 10 & 11 optimum plane is identified, from the figure 4.7 B-plane has the maximum variation hence is the best plane.



a) Clusterd Histogram R-plane



b) Clusterd Histogram G-plane



c) Clusterd Histogram B-plane

Figure 4.7: Example: Best plane identification

$$\theta_i = \min\left(\sum_0^{\frac{S}{3}}(R_i) - \sum_{\frac{s}{3}}^{\frac{2*S}{3}}(R_i), \sum_{\frac{s}{3}}^{\frac{2*S}{3}}(R_i) - \sum_{\frac{2*s}{3}}^{S}(R_i)\right) \quad (10)$$

$$\beta = \underbrace{max}_{i}\ (\theta_i\ ), i \in \{R, G, B\} \quad (11)$$

4.3 Text segmentation

To identify the text region of the image, segmentation is the next step. Segmentation is a difficult process as it has to consider different aspects of any ST image. ST image would contain information, such as a title, headings, names of authors, editors or publisher, content information, etc; together with these it will have various colors, fonts, sizes, orientations, spacing's and styles. [47] Moreover, ST images can also contain non-text information such as logos, graphic elements, drawings, images, etc. Further difficulties for segmentation arise due to textured or multicoloured background or even background images.

We use Connected Component (CC) based approach for segmentation. Any CC-based method [29] has four fundamental processing stages: (i) pre-processing, (ii) generation of C, (iii) removal of non-text areas, and (iv) grouping of text regions.

The binarization of the selected plane (β, refer Equation 4) is performed by selecting a threshold t to binary image using Equation-12

$$t = \arg\left(min\{w_1(t){\sigma_1}^2(t) + w_2(t){\sigma_2}^2(t)\}\right) \quad (12)$$

where w_i the probabilities of the two classes separated by a threshold t, ${\sigma_i}^2(t)$ are the variances of these classes.

This is also called as Otsu's method of thresholding [48]. A region growing method is then applied; image is scanned, beginning with a start region, the start region, pixels are merged if they belong to the same cluster in color space and homogeneous regions are formed. Depending upon the number of connected similar pixels, the image is divided in different regions, each region containing similar value pixels. Rectangular boxes are then formed around the homogeneous regions; we call them bounding box and are represented by BB_i, $i \in \{1, S\}, S$ is number of bounding box in image.These boxes differentiate the data to be extracted from the image. Figure 4.8 shows image with BB's.

Figure 4.8: Image with Bounding Boxes

After the pre-processing step in the CC- Based method, next task is to generate connected regions. there are several methods to accomplish this task, here we have described two methods a) Bottom-up analysis b) Top-down analysis in details, both these technique have some pros and cons .A comparison of the two methods is also included at the end of the description

4.3.1 Bottom up Analysis

The bottom-up analysis is used to detect homogeneous regions of a image using a region-growing method. The method commences by choosing a start region, pixels are fused to one cluster if it belongs to an identical color space. In this method arbitrary shape are obtained for machine printed text as these texts do not touch each other.

Procedure: Bottom up analysis uses region growing, technique generally used in image processing. As the start region three horizontally or vertically adjacent pixels are selected which belong to the same cluster and which are not yet assigned to a region. Initializing with the start region, shown in Figure 4.9 pixels within a 3×3 neighbourhood are iteratively amalgamated if they belong to the same group [29]. The process terminates when all pixels are merged into one of the regions, or no further start region can be found. The regions resulting from this procedure are represented by their bounding boxes. For each region the color is stored as possible text color

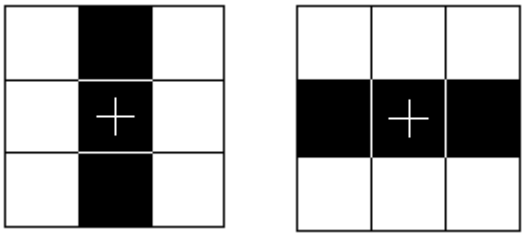


Figure 4.9: Start regions for region growing process [29]

4.3.2 Top Down Analysis

In the top-down segmentation procedure, the image is split alternately in horizontal and vertical directions. Regions obtained under this procedure are always of rectangular shape, and regions containing text include at least two colors. This knowledge is used to reject homogeneous regions as non-text elements during segmentation. Since the characters of machine printed text generally do not touch each other, several regions of rectangular shape result for a line of text, or even word.

Procedure: The image is processed in an iterative manner. A region is split along rows or columns in each iteration step. Commencing with the whole image as the start region, each row (column) is tested to see whether it contains one or more different colors of a region. A row (column) in which all pixels has the same color is rejected as non-text, i.e., the considered region is split along this row (column). The direction of splitting alternates in each iteration. For potential text regions, splitting terminates if at least n≥ 2 Different colors arise for all rows and columns of the region. This top-down procedure, also called the X-Y-tree decomposition [7], is a popular segmentation algorithm in document analysis.

4.4 Noise Removal

After segmentation of plane β, BB_i encloses all the connected components of the image. However, they may contain some non-text information such as logos, graphic elements, drawings, images, etc embedded into the text regions. Hence there is a need to suppress these no-text regions while maintaining text intact. Morphological operators can be used to achieve these results.

Table 3: Comparison of Top down And Bottom Up analysis [29]

Criteria	**Top-Down Analysis**	**Bottom-Up Analysis**
Small-sized text	Accurately segmented	Difficult to segment
Over segmentation	Not possible	Possible
under segmentation	Possible	Not possible
Inclusions of characters	Not possible	Possible
Images, graphics elements	Not split	Split into sub regions

MOs help removing imperfections by accounting for the form and structure of the image [49]. We use Opening (refer Equation-14) and Closing (refer Equation-13) operations in the present manuscript.

Closing operation

$$I(x,y) \bullet SE_{m,n} = (I(x,y) \oplus SE_{m,n}) \odot SE_{m,n} \quad (13)$$

Opening operation

$$I(x,y) \circ SE_{m,n} = (I(x,y) \odot SE_{m,n}) \oplus SE_{m,n} \quad (14)$$

where $SE_{m,n}$, is a structuring element, m and n are odd integers greater than zero.

Any SE which is rotationally independent can be chosen. We use the SE given in Figure 4.10 as this reduces the noise irrespective of its orientation.

1	1	1
1	0	1
1	1	1

Figure 4.10: Example of structuring element

On applying these MOs, majority the non-text regions of the image are removed.

4.5 Text extraction and identification

The image obtained from the previous step contains most of $BB_i's$ around the text region of the image. Next we rearrange the $BB_i's$'s according to the text style; for example, an English document will have the style of writing words from left-to-right.

Figure 4.10 shows the arrangement of $BB_i's$.A Matrix K is generated (refer Equation-15), that stores the entire bottom left coordinates of all the $BB_i's$ '. Matrix K consists of two columns, m is the x coordinate; n is the y-coordinate of these bounding boxes. For the first iteration K is sorted in increasing order .In the second iteration sorting is done with respect to x coordinate. Sorting is performed based on the Equation -16.

$$\mathbf{K} = \begin{vmatrix} m_1 & n_1 \\ m_2 & n_2 \\ m_3 & n_3 \\ \vdots & \vdots \\ m_s & n_s \end{vmatrix} \tag{15}$$

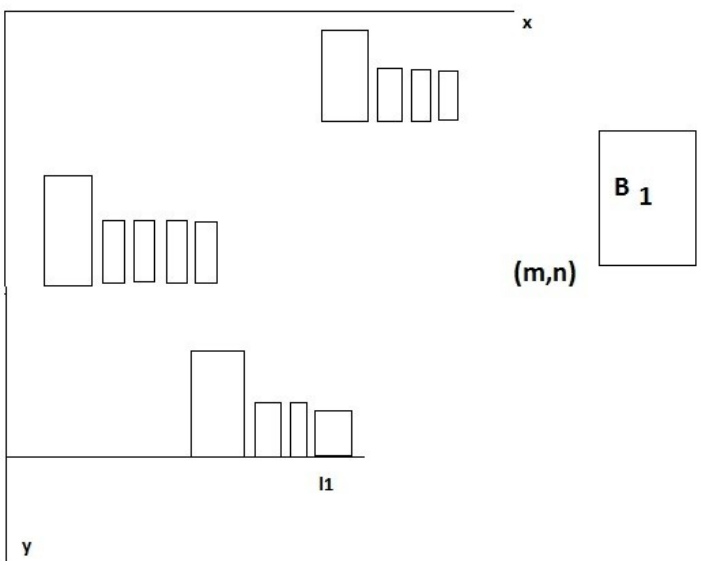


Figure 4.11: Image with Bounding Boxes

$$Q_n = (n-1) + \frac{2}{n}\sum_{k=0}^{n-1} Q_k \tag{16}$$

where Q_n denotes average number of comparisons, n is size of array. k ∈ $\{0, \ldots\ldots, n-1\}$

The algorithm extracts almost all of the alphabets of the image, however to structure the words accurately, appropriate spaces need to be added after end of every word, this decision is taken by a threshold value, Equation-17 gives the condition for threshold, if the extracted consecutive BB_i are separated by a value greater than the threshold, a space is inserted and vice versa. This threshold has been modelled experimentally by executing the algorithm on a variety of images.

$$\delta_{\text{threshold}} = \sum_{\text{i=1}}^{\text{no.of Bounding box}} \text{K(i, 2))/(2} * \text{no of bounding box)} \tag{17}$$

Figure 4.12: Example: Text Identification

Figure 4.12 shows an example of the text identification process. All the characters of the images are marked by the bounding boxes. Matrix K is generated by identifying the left bottom coordinate of each BB.

Figure 4.13: Example: Text Extraction

CHAPTER-5

RESULTS AND PERFORMANCE ANALYSIS

This chapter of the dissertation presents the experimental results. A detailed analysis of the results is also done in this chapter.

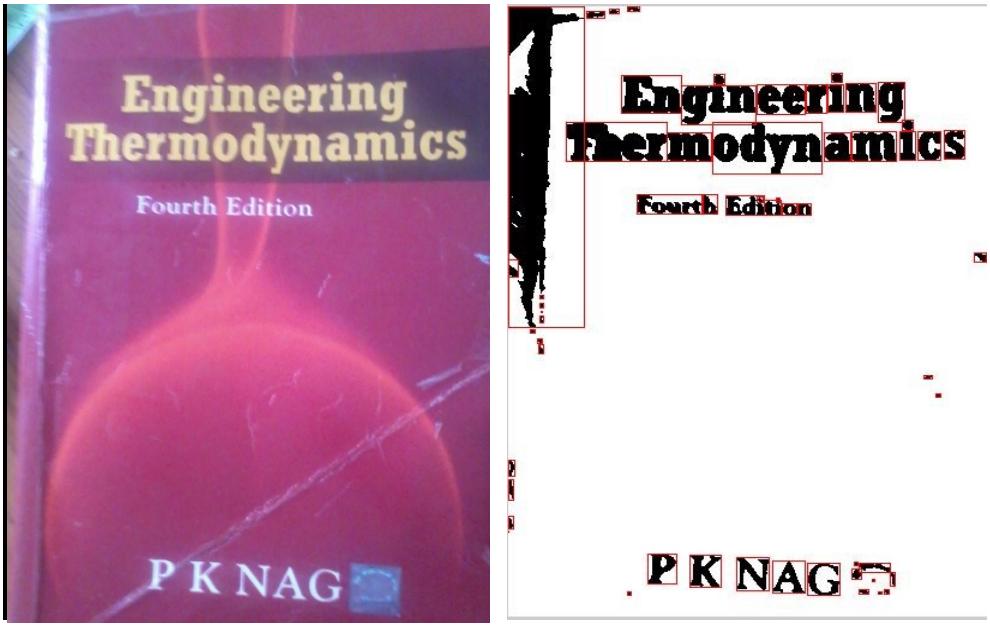
5.1 Dataset and Experimental Results

We have implemented our algorithm on the dataset that consists of book cover and book stack images. A diversified dataset has been considered which includes images with different types of background, font sizes, colors, orientations etc.

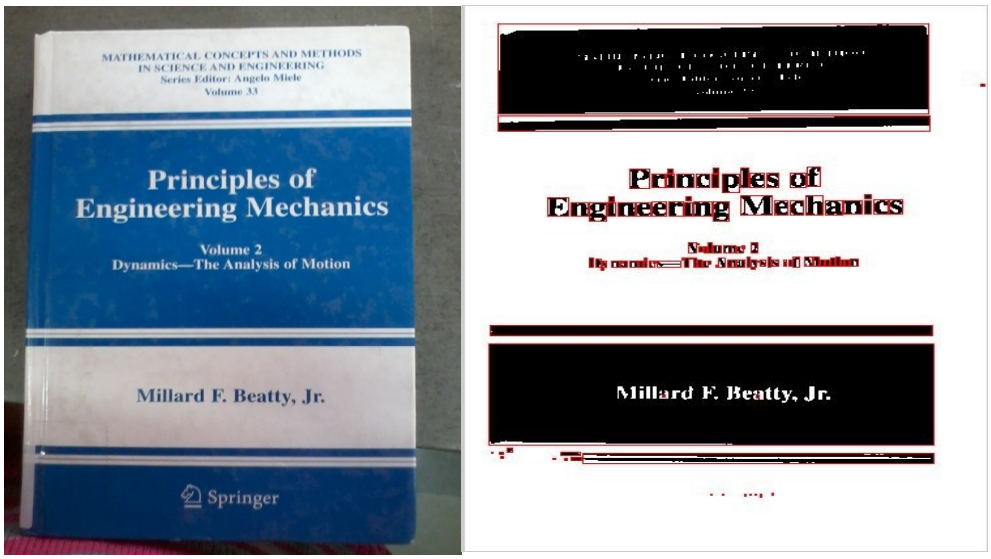
The proposed algorithm is implemented using Matlab R2010a programming. The execution time for an image is within 120 seconds depending upon the size of the image and also on the number of closed boundaries the image contains.

The proposed algorithm gives promising results for images of book covers with single color background as shown in Figure 5.1; algorithm is able to locate all the characters with 100 % efficiency. Images with logos are shown in Figures 5.2, this algorithm give significantly better results. Logos which have only graphics are rejected however where there are alphabets bounding boxes are created.

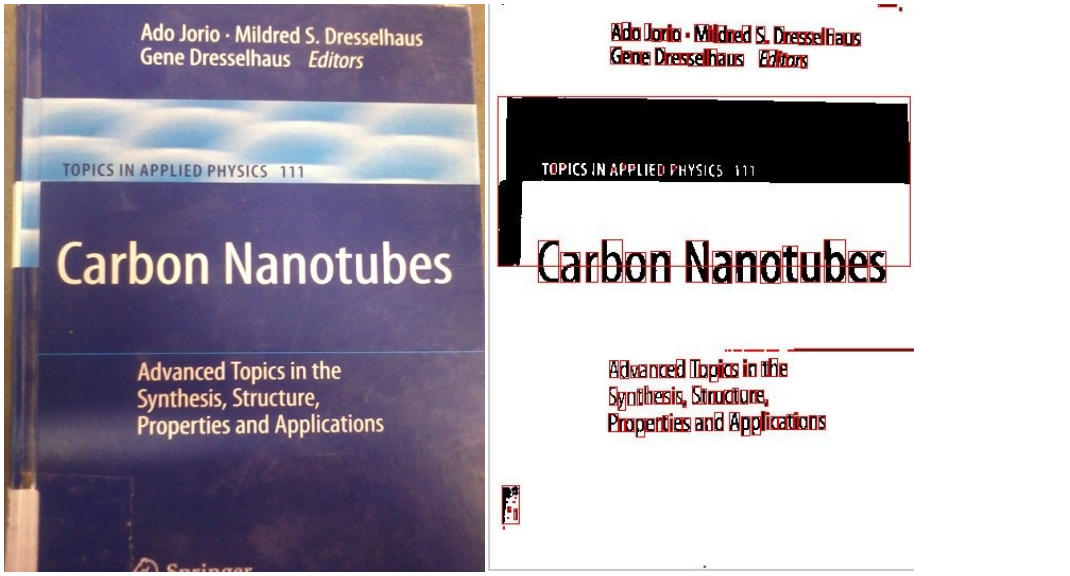
ICDAR 2011: dataset was used in text locating competitions of International Conference on Document Analysis and Recognition (ICDAR) 2011. This dataset is the most widely used benchmark for scene text detection. It contains 251 test images, including household objects, road signs, shop signs, bill-boards and posters, and book covers. The resolutions are from 307×93 to 1280×960. Four example images from this dataset are shown in Figure 5.1. The ground truths are bounded by green boxes. This dataset is freely downloadable at http://algoval.essex.ac.uk/icdar/Datasets.html.


Engineering
Thermodynamics
Fourth Edition
P K NAG
Engineering
Thermodynamics
Fourth Edition
P K NAG

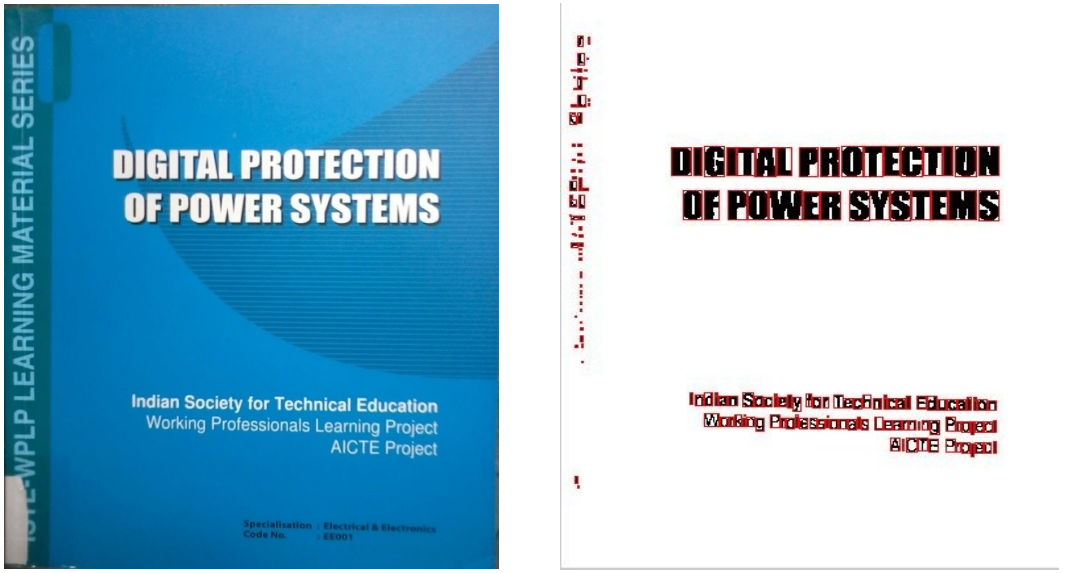
(a)


MATHEMATICAL CONCEPTS AND METHODS
IN SCIENCE AND ENGINEERING
Series Editor: Angelo Miele
Volume 33
Principles of
Engineering Mechanics
Volume 2
Dynamics—The Analysis of Motion
Millard F. Beatty, Jr.
Springer
Principles of
Engineering Mechanics
Millard F. Beatty, Jr.

(b)

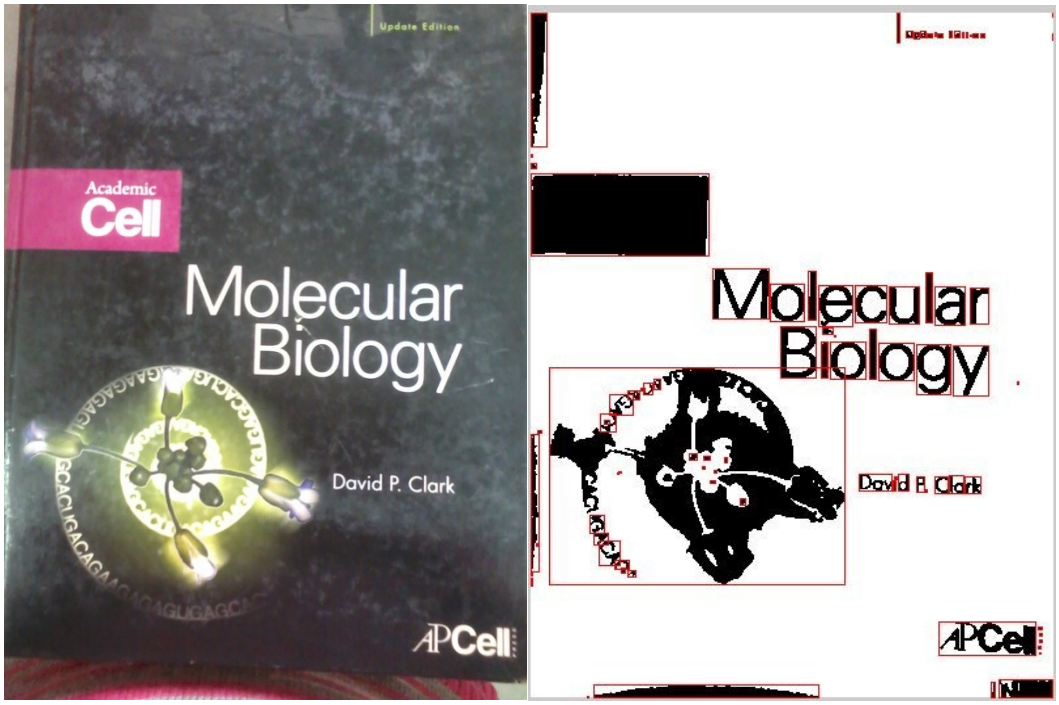


(c)



(d)

Figure 5.1: Results on single background images



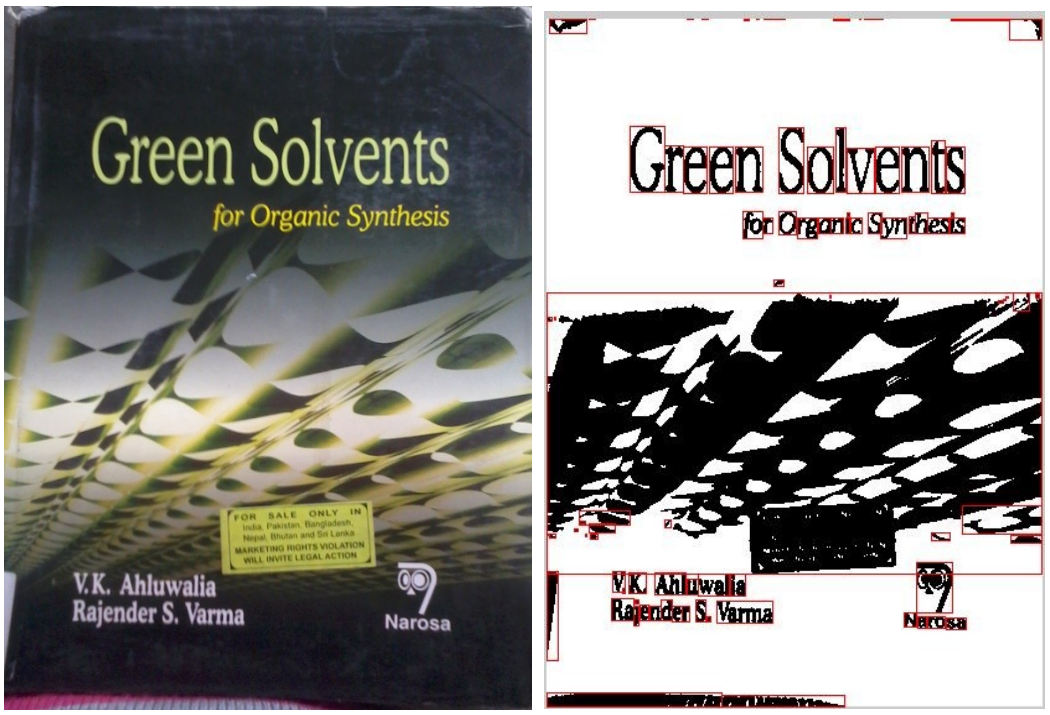
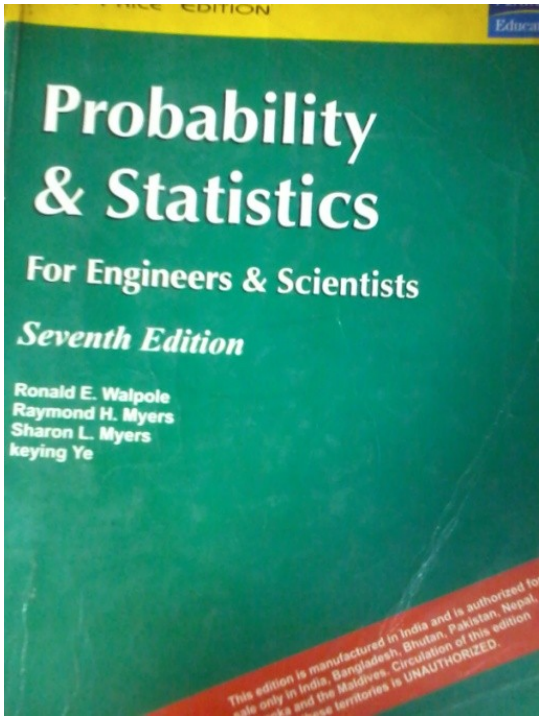
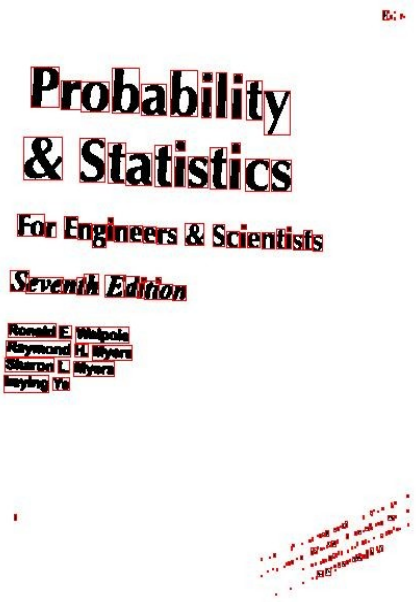
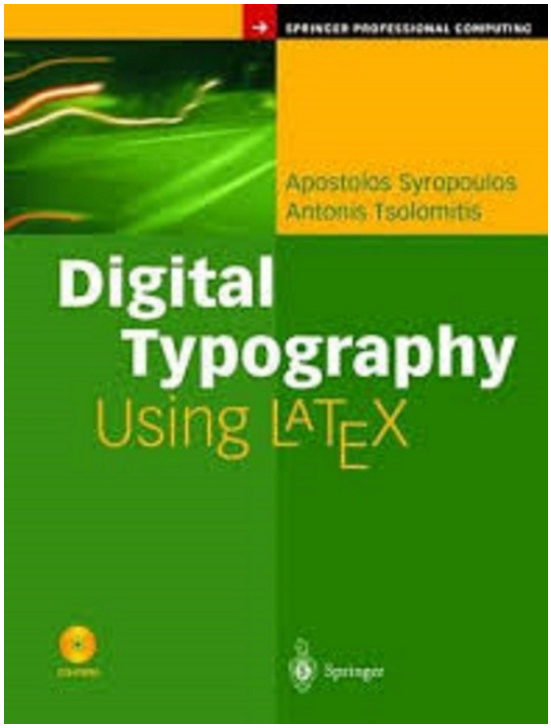


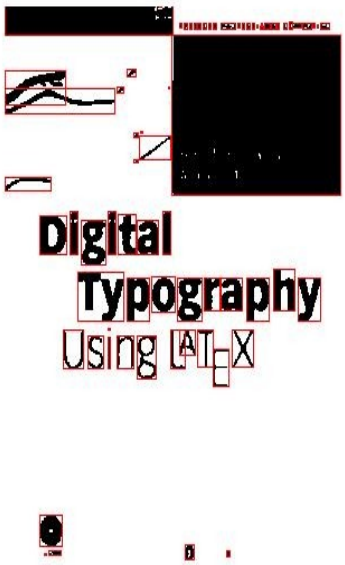
Figure 5.2 : Results on images with logos


Probability
& Statistics
For Engineers & Scientists
Seventh Edition
Ronald E. Walpole
Raymond H. Myers
Sharon L. Myers
keying Ye

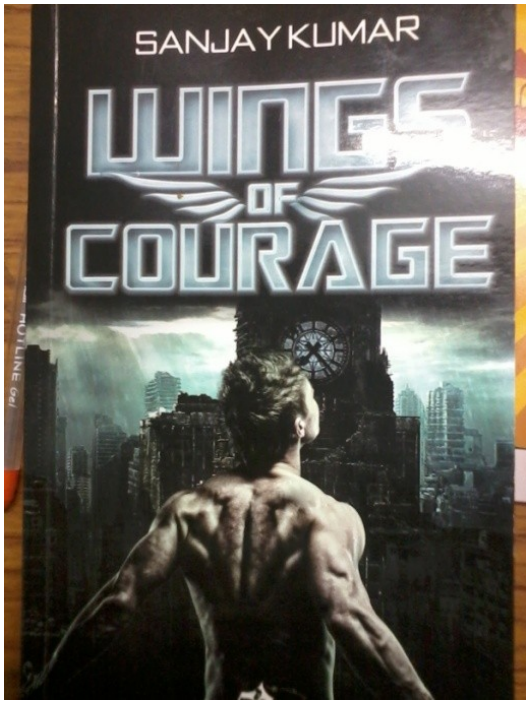

Probability
& Statistics
For Engineers & Scientists
Seventh Edition

(a)


SPRINGER PROFESSIONAL COMPUTING
Apostolos Syropoulos
Antonis Tsolomitis
Digital
Typography
Using LaTeX
Springer


Digital
Typography
Using LaTeX

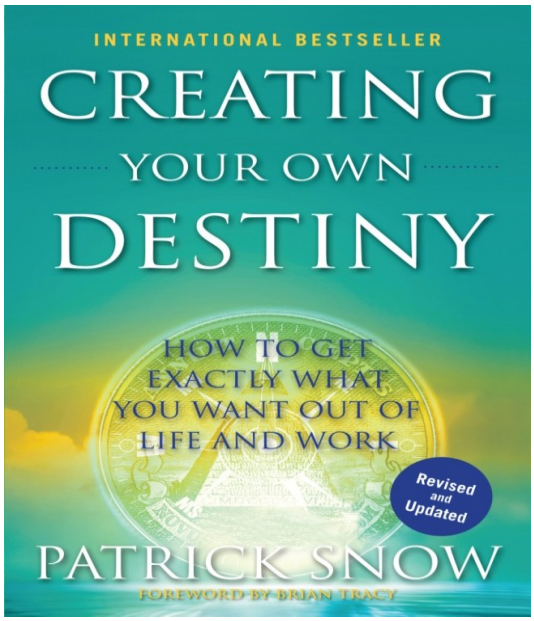
(b)

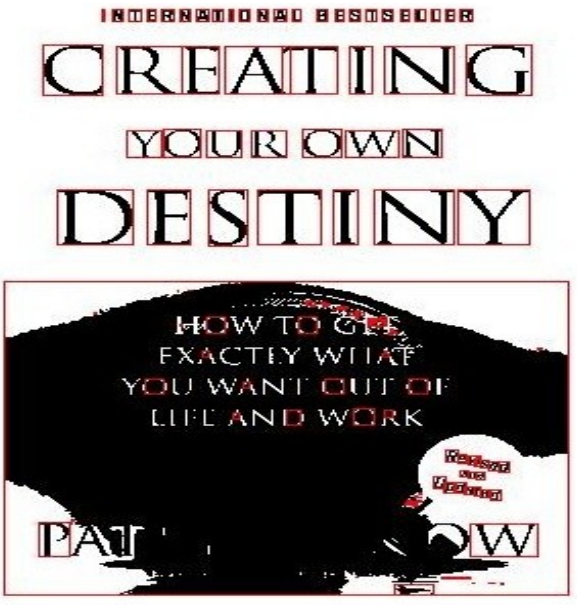

SANJAY KUMAR
WINGS OF COURAGE


SANJAY KUMAR
WINGS OF COURAGE


SANJAYKUMAR

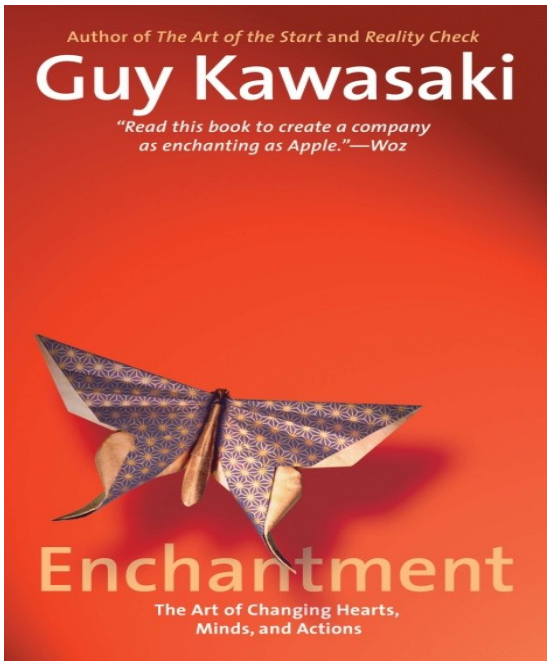
(c)

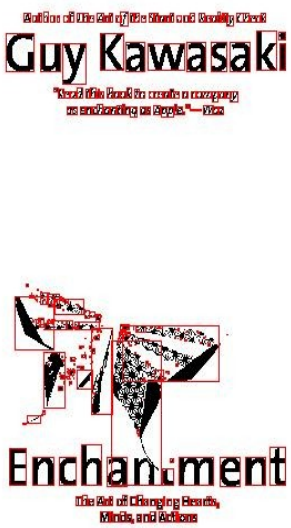

INTERNATIONAL BESTSELLER
CREATING
YOUR OWN
DESTINY
HOW TO GET EXACTLY WHAT YOU WANT OUT OF LIFE AND WORK
Revised and Updated
PATRICK SNOW
FOREWORD BY BRIAN TRACY

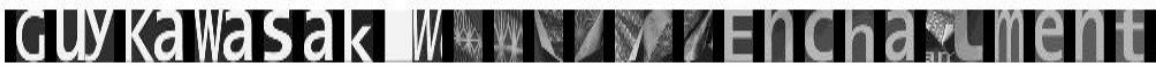

INTERNATIONAL BESTSELLER
CREATING
YOUR OWN
DESTINY
EXACTLY WHAT
YOU WANT OUT OF
LIFE AND WORK


INTERNATIONAL BESTSELLER CREATING YOUR OWN DESTINY

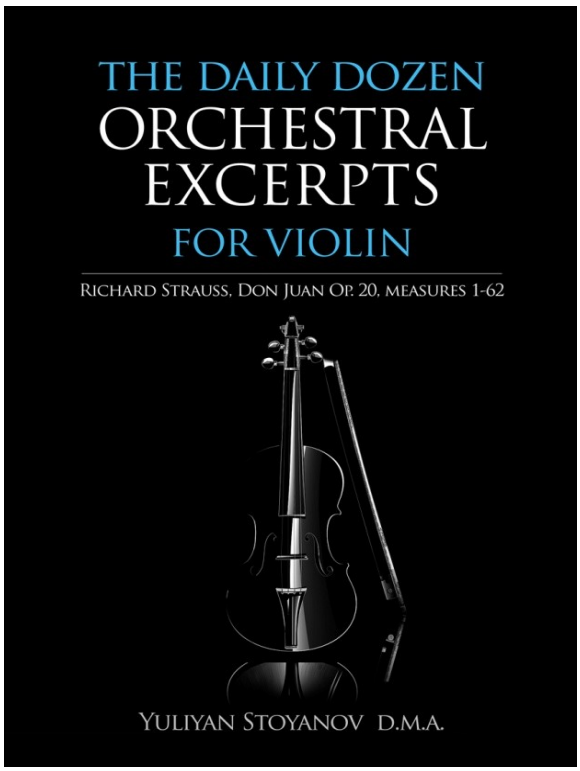
(d)

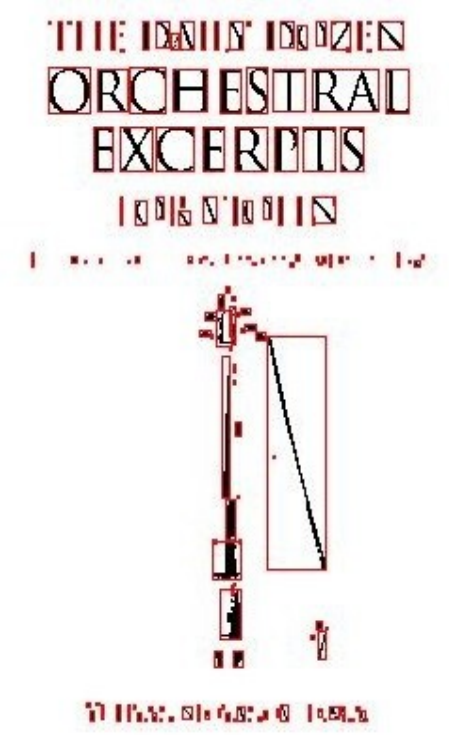

Author of The Art of the Start and Reality Check
Guy Kawasaki
"Read this book to create a company as enchanting as Apple."—Woz
Enchantment
The Art of Changing Hearts, Minds, and Actions


Guy Kawasaki
Enchantment

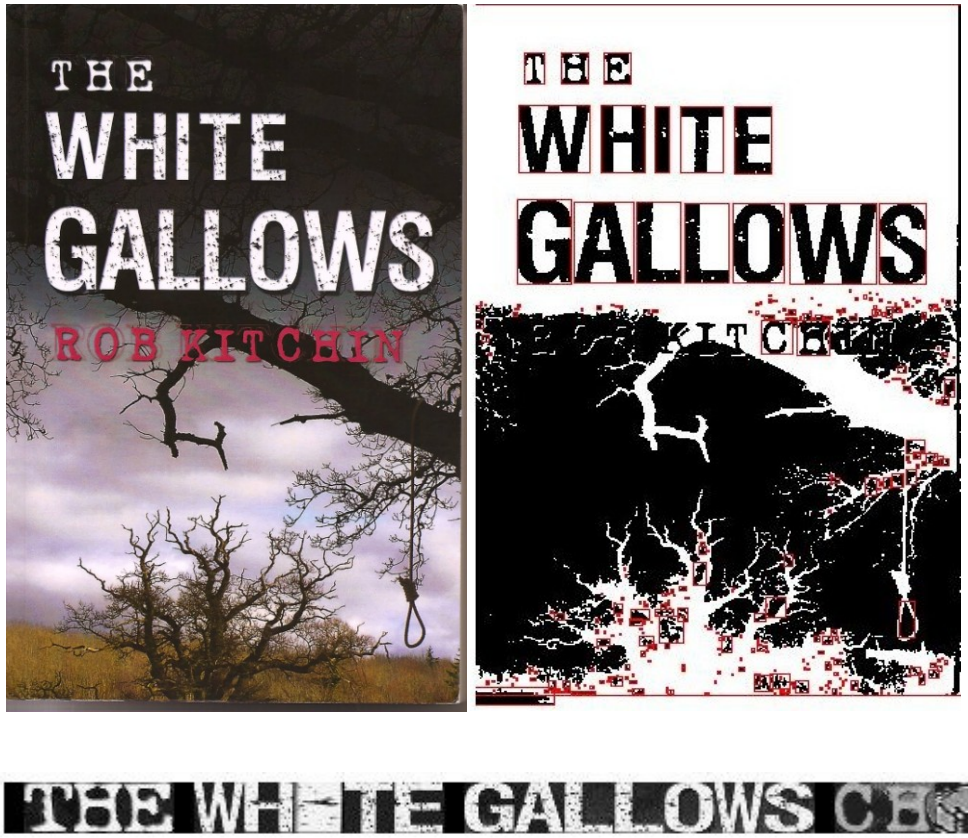


(e)


THE DAILY DOZEN
ORCHESTRAL EXCERPTS
FOR VIOLIN
RICHARD STRAUSS, DON JUAN OP. 20, MEASURES 1-62
YULIYAN STOYANOV D.M.A.


ORCHESTRAL
EXCERPTS

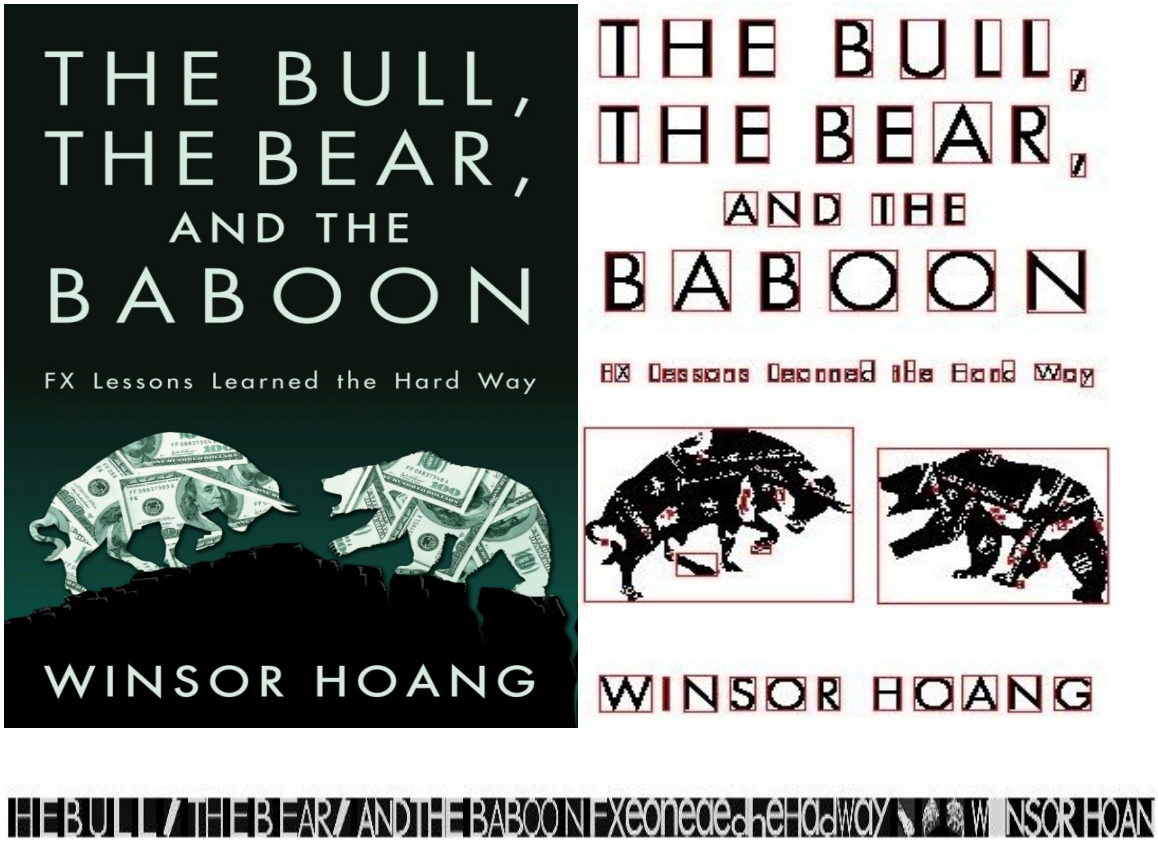
(f)


THE
WHITE
GALLOWS
ROB KITCHIN

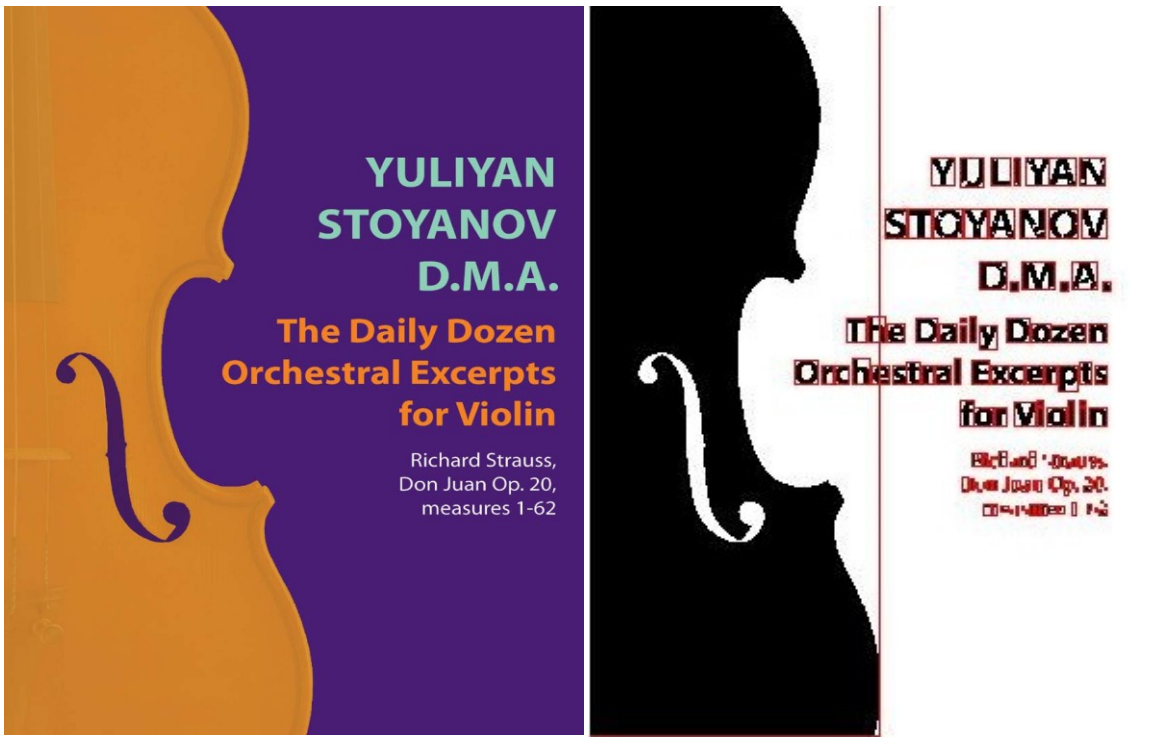
(g)


Go Creative!
It's Your Native State
Applying The Creative Process to Life
ORNA ROSS
go creative!

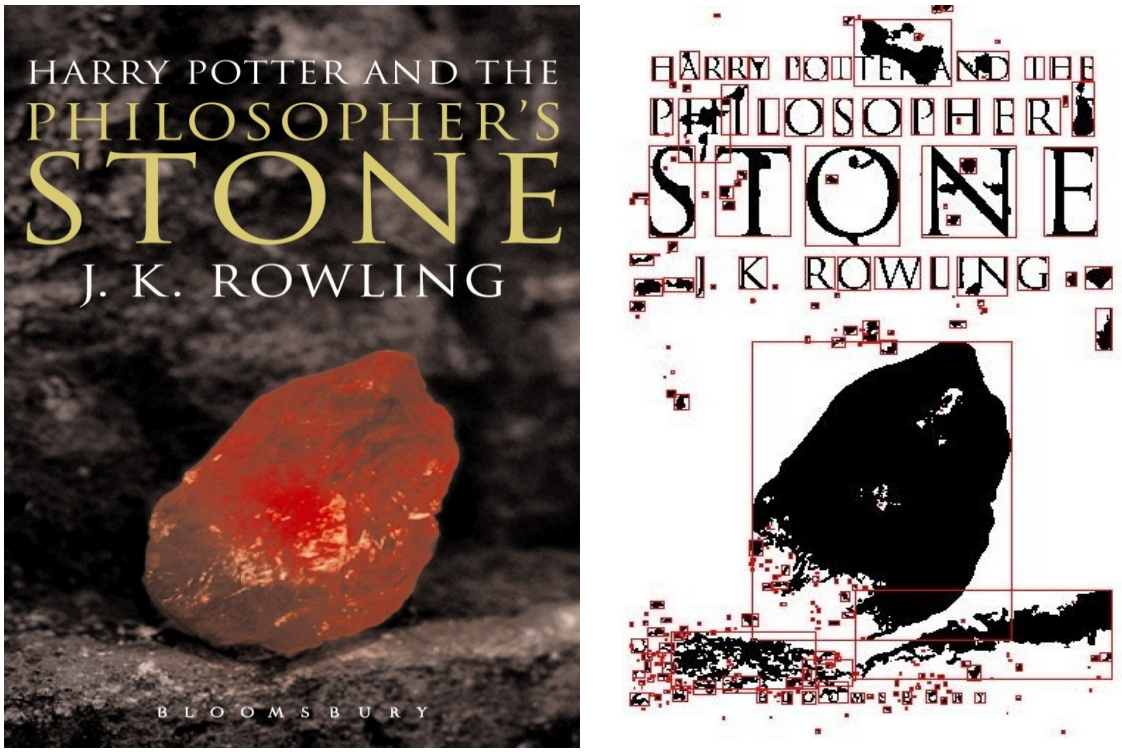
(h)


THE BULL, THE BEAR, AND THE BABOON
FX Lessons Learned the Hard Way
WINSOR HOANG
THE BULL, THE BEAR, AND THE BABOON
FX Lessons Learned the Hard Way
WINSOR HOANG

(i)


YULIYAN STOYANOV D.M.A.
The Daily Dozen Orchestral Excerpts for Violin
Richard Strauss, Don Juan Op. 20, measures 1-62
YULIYAN STOYANOV D.M.A.
The Daily Dozen Orchestral Excerpts for Violin

(j)



(k)

(l)

Figure 5.3 : Results for multi colored images

Figure 5.4 : Results on book stack images

(a)

(b)

Figure 5.5 : ICDAR 2011 Dataset results

5.2 Performance Analysis

Table-1 lists the actual values as well as the experimental results. Correctly deduced characters along with the errors in the form of false positive and false negative regions are presented. The data incorporated in the table has been obtained by executing the algorithm on images of multicolored images, images of book stacks and images containing logos. With the help of Equation (15) and (16) we calculated the recall rate and the precision rate of every test image considered. The average of these results was than calculated to compute the performance of the algorithm proposed.

Table-2 lists the error in the finally obtained results. This table shows the results of the text extraction and enhancement stage. The algorithm is able to extract the characters and then forming correct words by providing correct spaces. The current method gives a precision rate of around 83.7% and a recall rate of 87.4% of in case of complex images. Characters are identified as text regions at a rate of 98%, while the text region in the book stack images are rejected at rate of 45%.

Table 4: Results of text localization

Image	Total char.	Correctly detected char.	False positive	False negative
5.1 (a)	117	117	4	0
5.1 (b)	60	60	1	10
5.1 (c)	70	65	1	6
5.1 (d)	50	50	0	0
5.3 (a)	100	90	10	0
5.3 (b)	50	44	7	10
5.3 (c)	70	65	3	20
5.3 (d)	100	98	2	3
5.3 (e)	45	45	6	0
5.3 (f)	50	50	1	0

Table 5: Results of Text Recognition

Image	Total words	Correctly recognized words	Error (%)
5.1 (a)	17	17	0
5.1 (b)	11	8	20
5.1 (c)	7	7	0
5.1 (d)	12	8	33
5.3 (a)	17	17	0
5.3 (b)	20	16	20
5.3 (c)	16	16	0
5.3 (d)	20	17	15
5.3 (e)	13	13	0
5.3 (f)	10	7	30

CHAPTER-6
CONCLUSION AND FUTURE SCOPE

Texts in images contain valuable information that can be used for indexing applications and content based information extraction. There are mainly two types of text present in images. Artificial text present as captions etc in videos or digital born images is comparatively easier to retrieve than the scene text present in the naturally captured images using cameras, as scene text images vary in color, orientation, and size. Varieties of image extraction models and techniques have been given by various researchers. A detailed study of these techniques has been done during the course of research and work has been done to enhance the performance of some of the previously defined techniques.

This dissertation proposes a novel algorithm that can automatically retrieve text from images of books. Algorithm works for any image i.e., book covers, book stacks etc of size ranging from 500 ×700 to 1200 × 1200. Histogram based technique used in pre-processing step significantly reduced the color variations. Color quantization of 15×15×15 lead to about 10-12 clusters in case of a RGB image. It was also observed that the number of clusters not only depend on the content of the image but also on the color quantization. Most of the techniques that extract text do not consider the grayscale images, rather they convert the images to binary, and however this algorithm works to find out the best grayscale among the three planes in a RGB image. An optimized plane is identified using the intra plane variation. Preferring this plane makes the segmentation task much easier as the plane that has maximum variation in text and background is selected.

The algorithm implemented a connected component based method that merged similar regions together and these homogeneous regions were marked by bounding boxes. The segmentation technique gave effectual results by locating almost all the characters of the image on a case of book cover images. However for images having multicolored background error percentage was higher. The unwanted regions of the image i.e. the non-text regions for example logos etc, were removed significantly using the MO's.

The performance of the proposed algorithm was analyzed using two commonly defined parameters. The algorithm gives a precision rate of 83.7% and a recall rate of 87.4%. Character recognition rate is about 98% however the corrected recognition of words is about 80%. The 20% error in recognizing the correct spaces is due to various regions that require to be worked upon.

With the proposed method promising results were obtained on our data set however there is scope of improvement in case of the book stack images. It is observed that in a book stack image variation in the font size, color and background is significant. Hence; in the future our focus will be to bring innovation by considering these drastic changes and focusing on techniques that can efficiently retrieve the information. The other area to improve is the selection process of the optimized plane. In the current stage the algorithm is evaluated only on English datasets, in further work datasets of other languages will considered.

REFERENCE

[1] S. Mori, H. Nishida, and H.Yamada, *Optical character recognition*. John Wiley and Sons, Inc., 1999.

[2] J. M. White and G. D. Rohrer, "Image thresholding for optical character recognition and other applications requiring character image extraction," *IBM Journal of Research and Development*, vol. 27, no. 4, pp. 400-411, 1981.

[3] M. Farhad, S. Hossain, A. S. Khan, and A. Islam, "An efficient optical character recognition algorithm using artificial neural network by curvature properties of characters," in *Informatics, Electronics & Vision (ICIEV), 2014 International Conference on*, pp. 1-5, 2014

[4] K. Mohiuddin and J. Mao, "A comparative study of different classifiers for hand printed character recognition," *Pattern Recognition in Practice IV*, pp. 437-448, 2014.

[5] U. Pal and B. Chaudhuri, "Indian script character recognition: a survey," *Pattern Recognition,* vol. 37, no. 9, pp. 1887-1899, 2004.

[6] M.-H. Yang, D. Kriegman, and N. Ahuja, "Detecting faces in images: a survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 24, no. 1, pp. 34-58, 2002.

[7] E. K. Wong and M. Chen, "A new robust algorithm for video text extraction," *Pattern Recognition,* vol. 36, no. 6, pp. 1397-1406, 2003.

[8] I.Ben Messaoud, H. Amiri, H. El Abed, and V. Margner, "New binarization approach based on text block extraction," in *Document Analysis and Recognition (ICDAR)*, *2011 IEEE International Conference on* , pp. 1205-1209, 2014.

[9] K. Wang, B. Babenko, and S. Belongie, "End-to-end scene text recognition," *in Computer Vision (ICCV), 2011 IEEE International Conference on*, pp. 1457-1464, 2011.

[10] A. Gonzalez and L. M. Bergasa, "A text reading algorithm for natural images," *Image and Vision Computing*, vol. 31, no. 3, pp. 255-274, 2013.

[11] K. Jung, K. I. Kim, and A. K. Jain, "Text information extraction in images and video: a survey," *Pattern Recognition*, vol. 37, no. 5, pp. 977-997, 2004.

[12] V.Govindan and A.Shivaprasad, "Character recognition-a review," *Pattern Recognition*, vol. 23, no. 7, pp. 671-683, 1990.

[13] E. Kim, K. Jung, K. Jeong, and H. Kim, "Automatic text region extraction using cluster-based templates," in *Proc. of International Conference on Advances in Pattern Recognition and Digital Techniques*, pp. 418-421, 2000.

[14] S.Impedovo, L.Ottaviano, and S.Occhinegro, "Optical character recognition-a survey," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 5, no. 1, pp. 1-24, 1991.

[15] O. Grygorash, Y. Zhou, and Z. Jorgensen, "Minimum spanning tree based clustering algorithms," in *Tools with Artificial Intelligence, 2006. ICTAI'06. 18th IEEE International Conference on,* pp. 73-81, 2006.

[16] R.L.Cannon, J.V.Dave, and J.C.Bezdek, "Efficient implementation of the fuzzy c-means clustering algorithms", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 8, no. 2, pp. 248-255, 1986.

[17] A. Ahmad and L. Dey, "A k-mean clustering algorithm for mixed numeric and categorical data," *Data & Knowledge Engineering*, vol. 63, no. 2, pp. 503-527, 2007.

[18] H.Cauleld and W.Maloney, "Improved discrimination in optical character recognition", *Applied Optics*, vol. 8, no. 11, pp. 2354-2356, 1969.

[19] D. Chen, K. Shearer, and H. Bourlard, "Text enhancement with asymmetric filter for video ocr", in *Image Analysis and Processing, 2001. Proceedings.11th International Conference on*, pp. 192-197, 2001.

[20] A. Shahab, F. Shafait, and A. Dengel, "Icdar 2011 robust reading competition challenge 2: Reading text in scene images," in *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, pp. 1491-1496, 2011.

[21] Y. Zhong, K. Karu, and A. K. Jain, "Locating text in complex color images," in *Document Analysis and Recognition, 1995, Proceedings of the Third International Conference on*, vol. 1, pp. 146-149, 1995.

[22] J. F. Canny, "Finding edges and lines in images," M.I.T. Artificial Intelligence Lab., Cambridge, MA, Rep. Al-TR-720, 1983.

[23] S. R. Sternberg, "Grayscale morphology," *Computer vision, graphics, and image processing*, vol. 35, no. 3, pp. 333-355, 1986.

[24] J.-C. Wu, J.-W. Hsieh and Y.-S. Chen, "Morphology-based text line extraction," Machine Vision and Applications, vol. 19, no. 3, pp. 195-207, 2008.

[25] K. Mohiuddin and J. Mao, "A comparative study of different classifiers for hand printed character recognition," Pattern Recognition in Practice IV, pp. 437-448, 2014.

[26] J. Gao and J. Yang, "An adaptive algorithm for text detection from natural scenes," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 2, pp. II-84, 2001.

[27] U. Pal and B. Chaudhuri, "Indian script character recognition: a survey," *Pattern Recognition,* vol. 37, no. 9, pp. 1887-1899, 2004.

[28] J.Ohya, A. Shio, and S. Akamatsu, "Recognizing characters in scene images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 16, no. 2, pp. 214-220, 1994.

[29] K. Sobottka, H. Kronenberg, T. Perroud, and H. Bunke, "Text extraction from colored book and journal covers," *International Journal on Document Analysis and Recognition*, vol. 2, no. 4, pp. 163-176, 2000.

[30] Y. Zhong, H. Zhang, and A. K. Jain, "Automatic caption localization in compressed video," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 4, pp. 385-392, 2000.

[31] K. Kim, H. Byun, Y. Song, Y.-W. Choi, S. Chi, K. K. Kim, and Y. Chung, "Scene text extraction in natural scene images using hierarchical feature combining and verification," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 2, pp. 679-682, 2004.

[32] J.-C. Shim, C. Dorai, and R. Bolle, "Automatic text extraction from video for content-based annotation and retrieval," in *Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on.* vol. 1, pp. 618-620, 1998.

[33] J. Fan, D. K. Yau, A. K. Elmagarmid, and W. G. Aref, "Automatic image segmentation by integrating color-edge extraction and seeded region grow-ing, Image Processing," *IEEE Transactions on*, vol. 10, no. 10, pp. 1454-1466, 2001.

[34] A. K. Jain and B. Yu, "Automatic text location in images and video frames," *Pattern Recognition,* vol. 31, no. 12, pp. 2055-2076, 1998.

[35] S. Messelodi and C.M. Modena, "Automatic Identification and Skew Estimation of Text Lines in Real Scene Images," *Pattern Recognition,* vol. 32, pp.791-810, 1992.

[36] M. A. Smith and T. Kanade, *Video skimming for quick browsing based on audio and image characterization.* Citeseer, 1995.

[37] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform", in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. pp. 2963-2970, 2010.

[38] X. Liu and J. Samarabandu, "Multiscale edge-based text extraction from complex images," in *Multimedia and Expo, 2006 IEEE International Conference on*, pp. 1721-1724, 2006.

[39] R. Zanibbi and D. Blostein, "Recognition and retrieval of mathematical expressions," *International Journal on Document Analysis and Recognition (IJDAR),* vol. 15, no. 4, pp. 331-357, 2012.

[40] A. Risnumawan, P. Shivakumara, C. S. Chan, and C. L. Tan, "A robust arbitrary text detection system for natural scene images," *Expert Systems with Applications*, vol. 41, no. 18, pp. 8027-8048, 2014.

[41] K. I. Kim, K. Jung, and J. H. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm*," Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 12, pp. 1631-1639, 2003.

[42] P. Shivakumara, T. Q. Phan, and 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, 2011.

[43] S. P. Chowdhury, S. Dhar, K. Rafferty, A. K Das, and B. Chanda, "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.

[44] R.M.Haralick, S.R.Sternberg, and X.Zhuang, "Image analysis using mathe-metical morphology," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 9, no. 4, pp. 532-550, 1987.

[45] C.-M. Lee and 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.

[46] Y. K. Jain, S. Chugh, "Charecter Localization From Natural Images Using Nearest Neighbours Approach," *International Journal of Science & Engineering Research,* vol. 2, no. 12, 2011.

[47] M. Sawaki, H. Murase, and N. Hagita, "Automatic acquisition of context-based images templates for degraded character recognition in scene images," *in Pattern Recognition, 2000. Proceedings.15th International Conference on,* vol. 4, pp. 15-18, 2000.

[48] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285-296, pp. 23-27, 1975.

[49] M. Leon, V. Vilaplana, A. Gasull, and F. Marques, "Caption text extraction for indexing purposes using a hierarchical region-based image model," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pp. 1869-1872, 2009.

LIST OF PUBLICATIONS

1. “A novel algorithm for extraction of embedded text from images of books” communicated to International Journal for Light and Electron Optics.
2. “A review of various text localization techniques for natural scene images” communicated to International Journal of Digital Multimedia Broadcasting.