CHAPTER 1

INTRODUCTION

1.1 TEXT INFORMATION EXTRACTION (TIE)

Terabyte volumes of multimedia data in various modalities (as text, image, video and animation) are prevailing around the world, with the advent of the Internet and the World Wide Web. These are valuable resources and are highly desirable for indexing, archival and retrieval purposes. In addition, the popularity of the digital video is growing at an explosive rate. Surveillance cameras are everywhere, capturing videos for future scrutiny. Streaming video clips is increasingly popular on the Internet content. Digital Versatile Discs (DVDs) are quickly replacing analog video tape as the preferred medium for viewing movies at home.

Thus, the information era has brought us vast amounts of digitized text that are generated, propagated, exchanged, stored, and accessed through the Internet each day across the world. The accumulation of this data is making information acquisition increasingly difficult. Text Information Extraction (TIE) is a platform used to extract key information from web pages, images / videos. The extracted information can be used to support content based image indexing, information retrieval and search engines, and summarization.

Texts in various forms are frequently embedded into images providing important information about the scene, like names of people, titles,
locations or the date of an event in news video sequences, etc, requiring to be detected for semantic understanding and image indexation. Subsequently, the increased availability of online information has necessitated intensive research in the area of automatic Text information extraction (TIE) from images within the digital image understanding community.

The rapid rise in the quantities of digital video available on the Internet gives users access to unlimited image/video data from their home personal computers. It is quite probable that a user wants to see a video clip which exists somewhere in digital form, and search engines are required that can automatically identify the relevant video clips based on the user's query. The capability of the search engine is directly limited by the criteria and scope of the manually-created index. It is impossible for the human indexer to identify all possible keywords that describe a given video sequence. Hence algorithms need to be developed, that can automatically extract semantic information from an image / video using the content alone to provide better video search technologies.

A text embedded / inserted in images and videos is used to describe the contents of an image. An extracted text in a video sequence provides natural, meaningful keywords indicating the video's content. Also, the text can be easily extracted compared to other semantic contents, and it facilitates applications such as keyword-based image search, automated processing and reading of documents, text-based image indexing, pre processing for the OCR technique and multimedia processing.

A TIE system receives an input in the form of a still image or a sequence of images. The images can be in the gray scale or color, compressed or un-compressed, and the text in the images may or may not move. The extraction of this text information involves text detection / localization (TD/TL), tracking, binarization and extraction, enhancement, and recognition
of the text from a given image. It is evident that discriminating between the
text and non-text regions of an image is a complex and challenging task,
because natural scenes consist of complex objects, sometimes highly textured
buildings, trees, window frames and so on, giving rise to false text detection
and misses. In this context, the extraction of a scene and artificial text from
images and videos is an important research problem and receives growing
attention.

Precision and recall parameters are essentially a measure of the
ability of the TIE system to detect a text region, while at the same time,
avoiding a non-text one. It is normally considered that recall, precision and F-
score are sufficient to measure TD/TL effectiveness. The Character/Word
recognition rate (CRR / WRR) is used to evaluate the performance of the text
binarization system. In this work, methods to design a domain- independent
and robust TIE system have been attempted with an eye to improve the
TD/TL performance in terms of recall and precision, and Text binarization in
terms of CRR / WRR.

### 1.2 DIFFERENT TYPES OF TEXT IN IMAGES AND VIDEOS

Compared to low-level features (color, intensity, shape, texture, and
their temporal changes) and to other high-level features (present objects,
events, and their relations), the text is a powerful source of information about
the content of an image. Images used in multimedia databases might include
those captured from conventional digital cameras or video recording devices
or scanned documents. Document images are Electronic documents or images
of paper documents or images acquired by scanning book covers, CD covers,
or other multi-colored documents. The document images are categorized
(Jung et al 2004) as
i) Grayscale document images (Figure 1.1)

ii) Multi-colored document images (Figure 1.2)

The page layout analysis usually deals with document images (Jain and Zhong 1996) of Figure 1.1; it cannot be directly applied to Figure 1.2.

The texts in images/video images are classified into Caption text and Scene text images (Jung et al 2004).

- A caption text image is one which is an inserted text / artificially overlaid on the image, and otherwise called as superimposed/artificial text as shown in Figure 1.3 (Video titles, Name of the speaker/ place). The caption text usually underscores or summarizes the video's content. This makes caption text particularly useful for building a keyword index.

- An artificial text in news videos usually provides information about the names of related people e.g. anchors, location, subject of discussion, date and time when an event has occurred. The artificial text provides an abstract of the program and is often not available in other media like the audio track.

- Titles and credits displayed at the beginning or end of movies provide information, like the names of actors, producers, contributors, etc. Captions in sport programs often contain the names of the teams, players, scores, etc.

- Natural images /embedded text are called as scene text/graphics text image, which exist naturally in the image as in Figures 1.4 and 1.5 (Name of the Bottle drink in advertisements). The scene text occurring on signs, banners, etc. gives natural indications as to the content of a video sequence.
Text information can be found in the scene, e.g. the players’ numbers and names, the name of the team, brand names, location and commercials.

Displayed maps, Figures and tables in videos contain a text about locations, temperatures, and certificate items. Titles, logos, and names of programs displayed in videos are important for annotation, with respect to program types and names. Moreover, more and more web sites choose text pictures to improve the design of their web presentations.

In this context, the text appearing in images in general, and artificial text and scene text in particular, is of special interest due to two reasons:

- A text is very useful for describing the content of an image.
- The successful extraction of a text enables automatic text-based annotation, and subsequent keyword-based searching or content oriented processing.

![Figure 1.1 Samples of Document images](image-url)
Figure 1.2 Samples of Multi colored Document images

Figure 1.3 Samples of Caption text images

Figure 1.4 Samples of Scene text images
Many characteristics of the text in a document image are known a priori. For example, the text color in a document is nearly always black, and the background is known to be uniform white. There is high contrast between the background color and the text color. The orientation of the text can be assumed to be horizontal, or can easily be inferred by analyzing the structure of the document. In contrast, the text in scene / caption text images can have arbitrary and non-uniform stroke color. The background may be non-uniform, complex, and changing from frame to frame. The contrast between the background and foreground may be low. Text size, location, and orientation are unconstrained.

### 1.2.1 Properties of Text in Images

The text in images can exhibit many variations with respect to the following properties:

1. **Geometry**

   - Size: Although the text size can vary a lot, assumptions can be made depending on the application domain.
• Alignment: The characters in the caption text appear in clusters and usually lie horizontally, although sometimes they can appear as non-planar texts as a result of special effects. This does not apply to the scene text, which can have various perspective distortions. The scene text can be aligned in any direction and can have geometric distortions (Figure 1.5).

• Inter-character distance: characters in a text line have a uniform distance between them.

2. **Color:** The characters in a text line tend to have the same or similar colors. This property makes it possible to use a connected component-based approach for text detection. Most of the research reported till date has concentrated on finding the ‘text strings of a single color (monochrome)’. However, video images and other complex color documents can contain ‘text strings with more than two colors (polychrome)’ for effective visualization, i.e., different colors within one word.

3. **Motion:** The same characters usually exist in consecutive frames in a video with or without movement. This property is used in text tracking and enhancement. A caption text usually moves in a uniform way: horizontally or vertically. The scene text can have arbitrary motion due to camera or object movement.

4. **Edge:** Most caption and scene texts are designed to be easily read, thereby resulting in strong edges at the boundaries of the text and background.
5. **Compression**: Many digital images are recorded, transferred, and processed in a compressed format. Thus, a faster Text extraction system can be achieved if one can extract a text without decompression.

6. **Background**: The complexity of the background where the text appears can vary from simple to considerably difficult. The background can be composed of different colors, textures and other objects may be embedded in it as well.

Properties of a text in video/scene images are summarized in (Jung et al 2004) Table 1.1.

**Table 1.1 Properties of a text in images**

<table>
<thead>
<tr>
<th>Property</th>
<th>Variants or sub-classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geometry</td>
<td>Size</td>
</tr>
<tr>
<td></td>
<td>Regularity in size of text</td>
</tr>
<tr>
<td></td>
<td>Alignment</td>
</tr>
<tr>
<td></td>
<td>Horizontal/vertical</td>
</tr>
<tr>
<td></td>
<td>Straight line with skew (implies vertical direction)</td>
</tr>
<tr>
<td></td>
<td>Curves</td>
</tr>
<tr>
<td></td>
<td>3D perspective distortion</td>
</tr>
<tr>
<td></td>
<td>Inter-character distance</td>
</tr>
<tr>
<td></td>
<td>Aggregation of characters with uniform distance</td>
</tr>
<tr>
<td>Color</td>
<td>Gray</td>
</tr>
<tr>
<td></td>
<td>Color (monochrome, polychrome)</td>
</tr>
<tr>
<td>Motion</td>
<td>Static</td>
</tr>
<tr>
<td></td>
<td>Linear movement</td>
</tr>
<tr>
<td></td>
<td>2D rigid constrained movement</td>
</tr>
<tr>
<td></td>
<td>3D rigid constrained movement</td>
</tr>
<tr>
<td></td>
<td>Free movement</td>
</tr>
<tr>
<td>Edge</td>
<td>Strong edges (contrast) at text boundaries</td>
</tr>
<tr>
<td>Compression</td>
<td>Un-compressed image</td>
</tr>
<tr>
<td></td>
<td>JPEG,MPEG-compressed image</td>
</tr>
</tbody>
</table>
1.3 AREAS RELATED TO TIE

• Script Recognition

The text that appears in images or videos can be of a different script, such as Latin, Ideographic, etc. Languages such as English, German, Albanian, Italian, French etc., are classified under the Latin script, whereas Chinese, Japanese and Korean languages are known as Ideographic scripts. The process of script recognition (SR) deals with the problem of identifying the category of the script that a localized text belongs to. The identification of the used script can help in improving the segmentation results, and in increasing the accuracy of the OCR by choosing the appropriate algorithms.

• Handwritten recognition

Over the last few years, the number of academic laboratories and companies involved in research on handwriting recognition has continually increased. Simultaneously, commercial products have become available. Once the handwritten text has been localized, it will be recognized with the help of handwritten recognition methods. Methods and recognition rates depend on the level of constraints on handwriting. The constraints are mainly characterized by the types of handwriting, the number of scriptors, the size of the vocabulary and the spatial layout. Most efforts have been devoted to mail sorting, bank check reading, and forms processing in administration and insurance. These applications are of great economic interest, each of them concerning millions of documents.

• Text Matching

In the content image/video retrieval domain, it may be interesting to have the possibility to search for images of video frames where a text visually similar to the input text image appears. Visually similar means that the two
text images look similar to each other to a certain degree, which does not necessarily, imply that the same text appears in them. Given two different text images, a text matching (TM) algorithm tries to evaluate the visual similarity between the two.

- **Semantic Browsing of Images**

  Content-based media retrieval has received a lot of attention during the last few years, and ‘query by example’ is the most widely used methodology. Recently, relevance feedback methods have attracted researchers due to the possibility they offer to interact with the user, in order to increase the performance of a CBIR system. However, due to the increasing number of images and the need of the user to explore the media before taking a decision, the employment of techniques to visualize or browse a collection of images is becoming important. In this context, browsing methods to facilitate the interactive exploratory analysis of image data sets and to assist the user during semantic search, can be explored.

- **Citation analysis**

  When reliable text extraction techniques are available, then it is possible to automatically perform citation analysis. This task has a long tradition based on the manual annotation of bibliographic data and cross-references. Some well-known examples are the Science Citation Index (a commercial system) and the DBLP server that is fed by volunteers (Petricek et al 2005).

- **Layout recognition**

  Meta-data, “data about data”, provide high level information about a set of data. Meta-data are usually divided into three main categories:
administrative (e.g. the ISBN code), descriptive (e.g. the number of pages of a book), and structural (e.g. the title of a chapter). Structural meta-data can be extracted from a digital book only after an accurate analysis of the book content, and layout analysis techniques can be used to obtain this information. Layout analysis techniques have been widely investigated (Jain and Yu 1998, Marinai 2005) and are now used to process historical documents (Lebourgeois et al 2004).

- **Text summarization**

  Text summarization aims at the creation of a condensed version of a document or a document collection (multi document summarization) that should contain its most important topics. Most approaches still focus on the idea to extract individual informative sentences from the localized text. The summary then consists simply of a collection of these sentences. However, recently refined approaches try to extract semantic information from documents and create summaries based on this information (Jure et al 2004). In a paradigm more tuned to information retrieval (IR), one can also consider topic-driven summarization, that assumes that the summary content depends on the preference of the user and can be assessed via a query, making the final summary focused on a particular topic.

- **Word indexing and keyword spotting**

  Keyword spotting, whose goal is to locate user-defined words from an information flow (e.g. audio streams or sequences of digitized pages, such as faxes) (Curtis and Chen 1995, Williams et al 2000), is one of the first examples of the recognition-free paradigm. In the first approaches, the similarity computation took place considering the image or low level features, and demonstrated the feasibility of the general idea with low expectations
concerning the scalability toward large data-sets. Some recent applications concern the processing of historical documents (Terasawa et al 2005).

1.4 ISSUES IN TIE

Some of the issues related to the various processes involved in text information extraction are summarized as follows:

- **Variability in text size, color and position /alignment**

  Current text detection/localization approaches place several restrictions related to text size, color and alignment. The size of the text can vary in different ranges. Depending on the language, an English text requires at least 8-pixels-high font size, while a Chinese text requires at least a 20-pixels-high font size, due to the large number of strokes.

  Text strings can have different colors, but the characters belonging to an artificial text usually have a homogenous color (monochrome). However, a scene text can contain characters with different colors (polychrome) or sometimes characters of the same color, but with very different illuminations. Alignment/Orientation: The scene text can be aligned in arbitrary directions and can have perspective distortions. An artificial text is usually horizontally aligned, although sometimes due to special effects, it can appear as a non-planar text. Further investigation is required to handle these variations.

- **Text detection in video**

  Researchers have not given much attention to the text detection stage, mainly because most applications of TIE are related to scanned images, such as book covers, compact disk cases, postal envelopes, etc. which are supposed to include text. However, when dealing with video data, the text
detection stage is indispensable to the overall system for reducing the time complexity.

- **Different gray levels and high levels of noise and variations in illumination**

  These issues impose problems specifically for text segmentation and binarization techniques, where the foreground text is represented as black and the background as white. They give high false alarms for images with high noise and uneven illumination.

- **Poor quality and complex images**

  Although most texts with a simple background and high contrast can be correctly localized and extracted, images with low resolution and complex background will be difficult to extract. Text enhancement techniques can be investigated to solve these issues.

- **Text Extraction in Compressed Domain**

  Most digital images and videos are usually stored, processed, and transmitted in a compressed form. TIE methods that directly operate on images in MPEG or JPEG compressed formats have recently been presented. These methods only require a small amount of decoding, thereby resulting in a faster algorithm.

- **Performance Evaluation**

  Although various studies in computer vision and pattern recognition (CVPR) have investigated the issue of objective performance evaluation, there has been very little focus on the problem of TIE in images.
and video. Some of the issues related to the evaluation of text localization methods have been summarized (Antani et al 2001).

(i) **Ground truth data:** Unlike the evaluation of the automatic detection of other video events, such as video shot changes, vehicle detection, or face detection, the degree of precision of TIE is difficult to define. This problem is related to the construction of the ground truth data. The ground truth data for text localization is usually marked by bounded rectangles that include gaps between characters, words, and text lines. However, if an algorithm is very accurate and detects the text at the character level, it will not include these gaps, and thus will not have a good recall rate.

(ii) **Performance measure:** After determining the ground truth data, a decision has to be made, on which measures to use in the matching process between localized results and ground truth data. Normally, the recall and precision rates are used. Additionally, a method is also needed for comparing the ground truth data and the algorithm output: pixel-by-pixel, character-by-character, or rectangle-by-rectangle comparison.

(iii) **Application dependence:** The aim of each text localization system can differ. Some applications require that all the text in the input image must be located, while others focus on extracting only the important text. In addition, the performance also depends on the weights assigned to a false alarm or false dismissal.

(iv) **Public database:** Although many researchers seek to compare their methods with others, there are no domain-specific or general comprehensive databases of images or videos containing a text. Therefore, researchers use their own databases for evaluating the performance of the algorithms. Further, since many algorithms include specific assumptions and
are usually optimized on a particular database, it is hard to conduct a comprehensive objective comparison.

(v) **Output format:** The results from different localization algorithms may be different, which also makes it difficult to compare their performances. Some approaches only present a rectangular region containing the text and are unconcerned about the text skew. Other localization algorithms just present text blocks. A localized text region has to be segmented into a set of text lines before being fed into an OCR algorithm. Some algorithms localize text regions while considering the skew, and others perform more processing to extract the text pixels from the background.

### 1.5 APPLICATIONS OF TIE

Automatic Region of Interest (ROI) detection is an active research area in the design of machine vision systems, and is used in many applications. Researchers find that two levels of content always draw human visual attention, namely the low level perceptual content, such as color/intensity, texture, edges and geometric shapes, and high level semantic content such as human faces, text, and vehicles. In vision-based mobile robot navigation, many potential landmarks, such as nameplates and information signs contain the text. By locating those text landmarks, it can provide feature correspondence in a sequence of video frames to reconstruct 3D models.

In this context, text extraction has gained a lot of attention in the last few years and has potential applications in several areas such as

- Indexing, querying and retrieving of multimedia information
- Providing assistance to visually impaired persons
- Intelligent transport systems
Video indexing for better video search technologies
Pre processing for OCR,
Content-based retrieval,
Summarization of the visual information,
Mobile robot navigation,
Vehicle license detection and recognition,
Page segmentation and text-based image indexing
Text recognition systems to identify building addresses, street signs, business names, and restaurant menus
Searching museums for images
Tourist assistant systems
Navigational aids for automobiles

Current research in the field of text extraction in images/videos focuses on developing appropriate algorithms to extract different types of text from complex images or videos with the purpose of enabling the above services.

1.6 RESEARCH MOTIVATION

Multimedia data bases, both personal and professional, are continuously growing and the need for automatic solutions becomes mandatory. The effort devoted by the research community to content-based image indexing is also growing, but the semantic gap is difficult to bridge; the low level descriptors used for indexing are not efficient enough for an ergonomic manipulation of big and generic image data bases. The text present in a scene is usually linked to the image semantic context, and constitutes a relevant descriptor for content-based image indexing. Images used in
multimedia databases might include those captured from conventional digital cameras or video recording devices.

A distinguishing feature of a digital library is that, it has Terabyte volumes of multimedia resources which present a challenge for the application of multimedia technologies, such as video summarization, multimedia semantic annotation, multimedia cross indexing and retrieval, and 3-D visualization. Video summarization is a compact representation of video content. Video captions could also be used to generate video summarization with high-level semantics, since it implied lots of semantics inherently. For example, captions in news broadcasts and documentaries usually annotate information about the reported events. There are two kinds of text in videos: video caption and scene text. Scene text is part of the environment and is captured by the camera along with the rest of the scene. Video captions contain more semantic information than scene text.

The text embedded in images contains large quantities of useful semantic information which can be used to fully understand images. Text recognition is very useful in high level robot navigational tasks, such as path planning and goal-driven navigation. Therefore, a scene text is an important feature to be extracted. Captions present in video frames play an important role in understanding the video content and contain more semantic information than scene texts.

Page layout analysis usually deals with document images (Jain 1996) and cannot be applied directly to Artificial and Scene text images and multi-colored documents, because the contents of document images are far less complex than those of Artificial and Scene text images. In contrast to a caption text, a scene text can have any orientation and may be distorted by perspective projection. Characters in captions normally have only a horizontal
left-to-right orientation, with almost the same font size in a text line, while a scene text can be slanted or distorted by camera view angles with tapering characters. **Moreover, the scene text is often affected by variations in scene and camera parameters, such as lighting, focus etc.** Thus, algorithms designed for caption detection can not be directly used for scene text detection. **These variations make the design of unified Text extraction from various kinds of textual images extremely difficult.**

Specific application-oriented text extraction systems (Jung et al 2004) have been reported in literature such as page segmentation (Jain et al 1996, Tang et al 1996), address block location (Yu et al 1997), license plate location (Cui and Huang 1997, Kim and Chien 2001), and content-based image/video indexing (Zhang et al 1994, Shim et al 1998). Although a lot of work has been reported, it is still not easy to design domain independent text extraction systems. This is because there are so many possible sources of variation when extracting a text with different font sizes, colors, orientation and alignment, which could possibly be embedded in shaded or complex backgrounds. These variations make the problem of domain independent text extraction very challenging.

Therefore, designing a general purpose texture classifier or text detector that can distinguish the text with text-like texture from heterogeneous images with a unified framework, is still an open problem. Consequently, this research has been performed, guided by the motivation of Text extraction from heterogeneous images due to the explosion of Multimedia data and increasing number of camera based images in Digital Libraries, World Wide Web and Business Domain applications, and lacking a unified and domain independent text extraction system for heterogeneous textual images.
1.7 RESEARCH OBJECTIVE AND PROBLEM STATEMENT

Text extraction and recognition, which include text detection, localization, segmentation and binarization, and recognition, is a useful process for text-based image indexing. Furthermore, it is a very important task in searching for information in web sites or digital multimedia libraries (e.g. databases of images, videos or document images). The first three processing stages are important to achieve high-quality text recognition results when applying an OCR system.

The automatic extraction of a text is a challenging job due to variations in the font style, size, orientation, alignment and complexity of the background. In this context, the extraction of a scene and artificial text from images and videos is an important research problem and receives growing attention.

The methods and techniques developed for document images cannot be applied directly to caption and scene text images, as document images are far less complex. In contrast to a caption text, a scene text can have any orientation and may be distorted by the perspective projection. Moreover, it is often affected by variations in scene and camera parameters such as lighting, focus etc. These variations make the design of unified text extraction from various kinds of images extremely difficult.

The observation from literature that a separate text detection/localization algorithm needs to be designed for different kinds of images such as scene text, caption text and document images, sparked the idea to develop a unified scheme to distinguish the text and non text from heterogeneous images which takes care of variations in illumination, transformation/ perspective projection, font size, angular and radially changing text, so as to make the system suitable for heterogeneous images.
Binarization is the process of separating character strokes from the background. Binarization is necessary to bridge the gap between localization and recognition. A text occurring in caption / scene images can be of any color, and can appear against complex backgrounds. Thus, text segmentation methods which include the separation of the text from a complex background, and the binarization of the text image are crucial. It is the responsibility of a binarization algorithm to convert the complicated text regions occurring in images to the simple binary images required by the OCR. The characteristics of a scene text include uneven lighting, variability of character color, textured background and presence of text like objects, which demand the development of new methods to address these issues. Therefore, a text binarization system which could produce more discrimination across textured background images, with variations in font size, illumination, presence of text like objects and variability of character color, could well suit the needs.

Thus the primary objective of this research is to develop different algorithms to deal with the challenging problem of text detection, localization, segmentation and binarization in heterogeneous images as a unified framework to deal with Scene text, Caption text and Document images without devising a separate methodology. To achieve this objective, it is proposed to

- Develop a unified Text detection and localization (TD/TL) system to extract the text from heterogeneous images with an improved recall rate.
- Develop a robust TD/TL system to handle textual images with variations in text such as font sizes, location and orientation of text, Complex background, lighting conditions and Perspective
projection so as to make the system work across all kinds of images.

- Develop a text segmentation and binarization system which could produce more discrimination across textured background images with variations in font size, illumination, presence of text-like objects and variability of character color.

The input of the discussed problems can be a colored or gray-scale image. The input image can be any of the following types:

a) Scene text image
b) Caption text image
c) Document image
d) Hybrid image with Scene and Caption text image

1.8 CONTRIBUTIONS OF THESIS

Major contributions arise from the motivations and which brings novelty to our research has been listed below:

- A Unified approach for efficient text detection/localization from heterogeneous images
  o As a part of this idea, a **Sub band Texture Analysis algorithm (SBTA)** has been proposed in this thesis, to localize the text from heterogeneous images. This approach supports independent working with a limited range of font size and orientation, and can easily be extended to other languages.

- Robust Text detection/localization for heterogeneous textual images with variations in text
This idea gives birth to the development of **Multi level Feature Priority** (MLFP) technique, as a feature selection algorithm to isolate the textual blocks from image/picture blocks. This is robust to variation in illumination, Transformation/Perspective projection, Font size, and radially changing/angular text.

- **Text segmentation and binarization for complex textual images with variations in font size, illumination, presence of text-like objects and variability of character color**
  - As a part of this idea, **Integration of Edge and Color analysis (IECA)** algorithm has been proposed, where edge and color information are integrated to binarize the text to improve the character recognition rate.

- **Minimization of Background interferences in Text binarization system**
  - This idea gives birth to the contribution of **Character size Uniformity Check (CUC) based false edge removal** technique in binarization. The Sliding window approach based CUC technique has been proposed to remove non character EBs. This technique minimizes the background interferences due to a complex background.

- **Character segmentation in Text binarization system**
  - This idea gives birth to the development of **Edge Quadrant Coverage analysis (EQCA)** based character segmentation in binarization to filter out non text EBs having similar behavior (size) to character.
• Text Binarization from the segmented character
  o This idea leads to the development of Corner Vertices Color analysis (CVCA) in which the color information of corner vertices is used as an attribute to remove the background patches inside the segmented character.

• Integration of all factors to build complete Unified Text Extraction system
  o This idea gives birth to the development of a Unified Framework for Text detection, localization and extraction from heterogeneous textual images by integrating various factors such as text detection, text localization, text segmentation and binarization.

1.9 SIGNIFICANCE OF THESIS

Various algorithms implemented in this thesis have its own significance and play a major role in the following:

1. Mobile reading systems for the visually impaired or mobile translation devices for foreign visitors from camera-based images with blur, uneven lighting and complex backgrounds.

2. Text recognition systems to identify building addresses, street signs, business names, and restaurant menus which include scene text and caption text images in uncontrolled illumination conditions.

3. Vehicle license detection and recognition system, Video indexing and Embedded text detection system with anti-spam filtering.

4. Page segmentation system for newspaper headline extraction.
5. Text segmentation and binarization system will facilitate the conversion of the complicated text regions occurring in images to the simple binary images required by the OCR when the text appear in image with textured background, with variations in font size, illumination, presence of text like objects and variability of character color.

6. This system could be integrated with software before OCR, which builds and distributes digital library collections through the Internet, to access the information from web images which can be scene text / caption text/document images.

1.10 ORGANIZATION OF THESIS

The organization of the thesis is as follows:

- Chapter 1 gives the general introduction to the thesis.

- Chapter 2 discusses a literature survey of the topics associated with this work and the observations on literature conclude the chapter.

- Chapter 3 explains the system overview.

- Chapter 4 provides details of the proposed Image analysis based SBTA-TD/TL approach for text detection and localization. A quantitative performance evaluation is performed to compare this algorithm with others in the literature.

- Chapter 5 describes an Image analysis and the Machine learning based approach MLFP-TD/TL algorithm for text detection and localization. The experimentation results and a comparison with other methods conclude the chapter.
• Chapter 6 presents an IECA based text segmentation and binarization algorithm and its evaluation follows.

• Chapter 7 concludes the thesis with a summary of the achievements and limitations of the proposed solutions and suggests some future research directions.