DESIGN AND DEVELOPMENT OF MEDICINE TEXT IDENTIFICATION SYSTEM

A Dissertation Submitted in Partial Fulfillment of the Requirement for the Award of the Degree of

MASTER OF ENGINEERING

IN

ELECTRONICS AND COMMUNICATION

Submitted By

TANVI MEHRA

Roll no. 801561029

Under Supervision of

Dr. VINAY KUMAR

Assistant Professor, ECED

Thapar University, Patiala

ELECTRONICS AND COMMUNICATION ENGINEERING DEPARTMENT

THAPAR UNIVERSITY, PATIALA, PUNJAB

JUNE, 2017

DECLARATION

I, **Tanvi Mehra**, hereby declare that the dissertation entitled **"Design and Development of Medicine Text Identification System"** is an authentic record of my study carried out towards the partial fulfilment as requirement for the award of degree of Master of Engineering in Electronics and 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 to any other University/Institute for the award of any other degree.

Date: 21-08-2017

Tanvi Mehra
Roll No. 801561029

It is certified that the above statement made by the candidate is correct to the best of my knowledge and belief.

Date: 21-08-2017

Dr. Vinay Kumar
Assistant Professor
ECED, TU, Patiala

ACKNOWLEDGEMENT

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. Alpana Aggarwal,** 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. Hem Dutt Joshi**, Assistant Professor and P.G. 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 **Tanvi Mehra**

Date: Roll no. 801561029

ABSTRACT

Recent studies in the field of computer vision and pattern recognition show a great amount of interest in content retrieval from images and videos. This content can be in the form of objects, colours, texture, shape as well as relationship between them. Text in images are important source of information for several advanced applications; such as, video and image retrieval, web image search, multilingual translation, content based automatic annotation of image databases and assisting visually impaired to read labels in map applications . The aim of text localization and detection is to find the regions in the medicine considered text by humans, mark text boundaries (usually by rectangular bounding boxes) and produce the associated characters.

Text Extraction and identification from medicine strips and bottles is a valuable application that can help pharmacies to create their own medicine databases and assist the patients with the medicine information i.e., the salts present in the medicine and the possible substitutes to the medicine. The problem of text extraction is a challenging one due to variety of text variations on medicines such as; font, size, colour, alignment, illumination and reflection. In this dissertation, we put forth an accurate and robust medicine text detection algorithm. With the present technique, images of complete, partial, distorted or occluded medicine strip or bottled medicine can be used to identify the medicine name. The text extraction and identification is performed with the help of edge enhanced Maximally Stable Extremal Regions (MSERs) followed by geometric filtering and Stroke Width Transform to remove the non-text regions. Next, the OCR system uses Stroke Width transformed image and the region of interest in order to recognize and display the text string. To deal with missing and extraneous characters during recognition novel string editing is applied to extract the correct medicine name.

The algorithm is evaluated on a dataset containing both of medicines strips and bottled medicines. Since there is no existing database, we created the database with a wide variety of medicines from different pharmacies using a regular camera under natural lighting conditions. The experimental results exhibit excellent performance for the proposed technique. The system gives an efficiency of approximately 95% on medicine strips and bottled medicines images captured under varied conditions; for example, reflection, bad illumination and skew-ness. The method is capable of detecting highly blurred text in low resolution medicine images as well as rotated text for the bottled medicines.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Sr. No **Figure details** **Page No.**

LIST OF ABBREVIATIONS

AT: Artificial text

BB: Bounding Boxes

BAG: Block Adjacency Graph

CC: Connected Components

CRF: Conditional Random Field

DCT: Discrete Cosine Transform

HOG: Histogram of Oriented Gradient

ICDAR: International Conference on Document Analysis and Recognition

JPEG: Joint Photographic Experts Group

LBT: Local Binary Pattern

LHBP: Local Haar Binary Pattern

MSER: Maximally Stable Extremal Regions

MG: Mean of Gradients

MDF: Mean Difference Features

MATLAB: Matrix Laboratory

OCR: Optical Character Recognition

SWT: Stroke Width Transform

SVM: Support Vector Machine

TIE: Text Information Extraction

TL: Text Localization.

TS: Text Segmentation

CHAPTER 1

INTRODUCTION

Images and videos are the two most important multimedia documents that are being produced on the daily basis by several sources such as broadcast and entertainment industry, medical diagnostic systems, surveillance systems, long distance education programs etc. These days' portable cameras and camcorders are available at a very low price thus making the number of images and videos being captured grow at a very rapid rate. Given such vast quantity of images, it is quite possible that image of our interest is available and can be accessed on the internet. On the contrary, the availability of such huge number of images makes it quite difficult to locate specific images that we require thus making Content Based Information Retrieval (CBIR) system the need of the hour. Recent studies in the field of computer vision and pattern recognition show a great amount of interest in content retrieval from these images. The content present in images can be divided into two categories: perpetual and semantic content. Perpetual content represents features such as colour, intensity, shape and the changes associated with them whereas the semantic content represents objects, events and their relationship. Text, faces, and vehicles in an image are the semantic information. Out of these, text is of greatest interest as:

i) Text can easily and clearly describe the contents of the image compared to the other attributes such as face and vehicles given by Jung, Kim and Jain in [2].

ii) Text in images contains important information which can be utilised for applications such as video and image retrieval, web image search based on its content, content based automatic annotation of image databases and assisting visually impaired to read labels in map applications.

Text information extraction involves detection, localization, enhancement and recognition of text from a particular image. Text localization aims to find the image areas that are considered text by humans, create rectangular boxes around these areas to mark their boundaries and generate a output character sequence related to this content. Scene text localization and recognition is an open problem unlike the document recognition where the existing systems can correctly identify more than 99% of characters. Factors contributing to problem complexity include: Different font styles and sizes, variety of colours, different alignment and orientation (skewed or rotated) of text buried in the images. Text may be partially occluded making the extraction process a very challenging task. Further adding to the problem is the presence of complex background, reflection or blur in the image thus before identifying and retrieving the text, it has to be robustly detected.

1.1 TEXT IN IMAGES

The text in images is categorized in three ways:

1) Scene Text
2) Artificial Text
3) Document Text

Scene Text refers to the text information that was clicked by the camera as the part of the scene, e.g. text on the boards, text on the back of the medicine strips and bottles, text on the T shirts, license plates, or even the road signs. In other words, scene text refers to the text in natural scene images. Figure 1.1(a)-(b) represents the images with scene text. Scene text image is often affected by changes in parameters related to camera for e.g., motion, focus, blur, uneven illumination, distortion, less dynamic range and lack of exposure.

Although reliable information is provided by the embedded text, it is a tough task to detect and extract the text from scene images due to the following reasons: in a single frame, the character size varies from small to big; the image can have varying fonts and styles. The text within the same image can have multiple colours, alignments and orientations. As stated earlier text maybe present in a cluttered and complex background thus making the separation of text and non-text regions an arduous task. Moreover the existing OCR systems work for the text against the plain background (monochrome) and are unable to handle the text against the complex or textured background. Direct recognition of scene text is impossible and the there is still scope for new methods.

Artificial text, as the name suggests, is artificially added during the post processing stage to provide multimedia content description or give any additional information. It finds its use in keyword indexes, title sequence, epilogue, logos and subtitles in videos. As compared to the scene text, artificial text provides more information which makes it better to identify and retrieve thus such text is usually embedded in the videos for quick retrieval.

Figure 1.2 represents the artificial text also referred to as the caption text. Figure 1.1 and 1.2 represent the difference in the scene text and the artificial text in terms of image source, colour, alignment and the background complexity.

Document image is acquired using scanning journal, printed documents, handwritten notes, or book covers etc. *Document text* [46] usually contains text against a monochrome background and few graphic contents. Figure 1.3 represents the research papers that can be considered as document images. Characteristics similar to document images are obtained by book cover scanning, CD packets or documents with multiple colours (Figure 1.4) however they cannot be directly handled using the existing document image analysis techniques.

(a)

(b)

Figure 1.1 Scene text images (a) Images with scew, blur, allignment and illumination [2]. (b) Images of the medicine strips and bottled medicine.

Figure 1.2 Images with artificial (caption) text. Caption is overlaid on the background and the text string maybe polychrome [2].

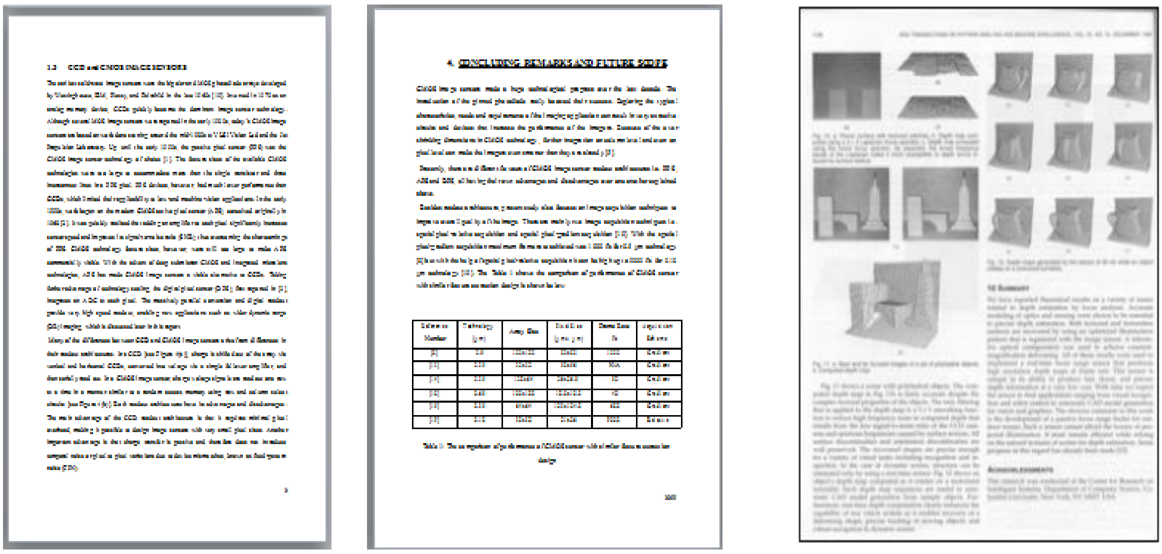


Figure 1.3 Gray scale document images with single column and multiple colums.

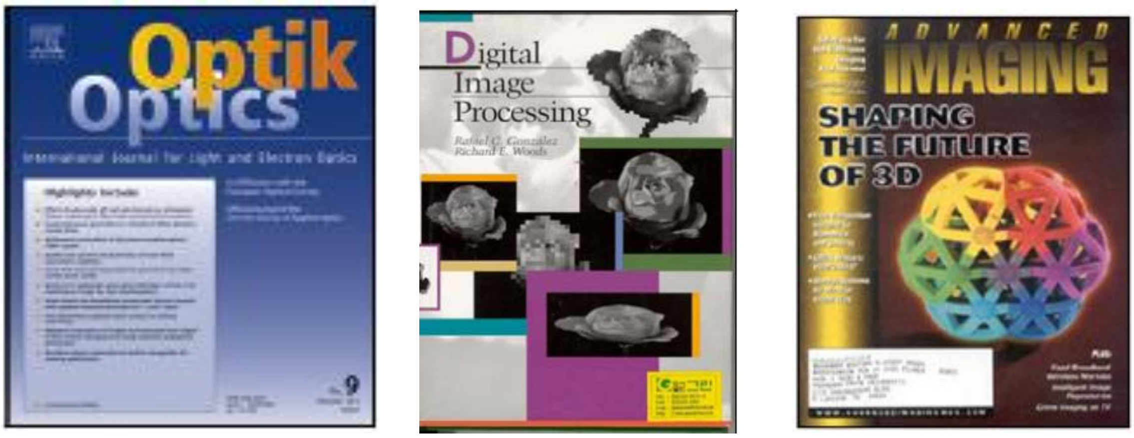


Figure 1.4 Multi colour document images with same or different coloured text lines.

1.2 IMAGE CHARACTRISTICS OF TEXT

Text in images exhibits a number of variations. In this section the characteristics of the text in image are listed that can help in its segmentation from the background i.e. rest of the image scene. These properties have been derived from the extensive research and general image observation.

1) **Size :**

   Text can appear in a variety of sizes as there are varied amount of images that one can work on. Since there exists a range of distance from which the text must be readable in short time, there often is a limit on the minimum size of characters. However there is no such upper bound on the size. Text can be large enough to cover half the height of the frame or more. Figure 1.4 represents text of varying sizes e.g. the size of the text string “ADVANCED” and “IMAGING” is different.

2) **Inter-Character gap:**

   Most caption strings are composed of one word minimum that can go up to maximum few words. Since captions have limited word count the characters have spaces between them i.e. they are well separated.

3) **Geometry:**

   Generally, the text string that is artificial is aligned horizontally. However, if some special effects are used the string may appear as non-planar. In case of scene text direction of text is unknown leading to geometric distortions. Another geometric feature under consideration is the aspect ratio that lies in a specific range for the string.

4) **Colour and Intensity:**

   Colour is a definitive feature for visual information. An important property of text is that its characters have same or similar colour and lie on the same stroke. Most reports concentrate on finding the monochrome strings however some of the video frames may contain polychrome strings.

5) **Edges:**

   As per observation, edges exist between the text boundaries and the background. Thus most text extraction algorithms employ some form of edge detection e.g. canny edge detector. Use of edge based methods has enhanced the extraction process as strong edges exist in most of the scripts.

6) **Rigidity:**

Artificial as well as the scene text strings have the tendency to maintain their shape, size, orientation and font over multiple frames. However, in case of addition of graphic effects the text may have varied properties. This is a good feature which is generally maintained in most of the images and videos.

7) **Motion:**

This is one parameter that is of utmost importance for tracking in videos. The position of the caption string may change from one frame to another but it does show much variation. The caption text can show a uniform motion either horizontally or vertically however the movement of camera decides the orientation of the scene text string.

8) **Compression:**

Usually a compressed format is used to process and store the digital images. For a robust text extraction it is of utmost importance that the text is extracted without the decompression.

An earnest observation suggests that the text in images can exist in several different languages. The type of text may be different in each case e.g. the inter-character distance in Chinese text is different from the English text. Thus there is a need of robust text detection algorithm that works efficiently even if some of the features fail.

1.3 EXTRACTING TEXT FROM MEDICINE STRIPS AND BOTTLES: ISSUES

Text identification in document images is an important area in research for a long time. However text recognition in medicine strips is a problem that requires several different approaches. Medicine strips and bottled medicines (Figure 1.1 (b)) have text in scattered areas rather than ordered columns unlike the document images. In these strips there are fewer separated lines with text being multi-coloured, multi-font, or even the same colour as the background. Further adding to this variety is the presence of the light text against the dark background or the dark text on the light one, this influences the gradient directions. Thus the methods applied to document images based on page models and top down page segmentation are not suitable- there is a stringent need for locally adaptive bottom up methods with least dependence on regularity, font size uniformity, multiple line availability etc. The binarization of grey-scale image is a pre-requisite step for document images; however due to the complex and non-uniform background the medicine strip image maybe noisy and even low resolution. Further, the text regions may be translucent. Clearly, the binarization step is a dicey choice. Moreover the existing OCR systems work for the text against the plain background (monochrome) and are unable to handle the text against the complex or textured background. Direct recognition of text in medicine images is an impossible task.

1.4 PROBLEM STATEMENT: TEXT INFORMATION EXTRACTION MODEL

This section formally defines the Text Information Extraction Model (TIE) for extracting text (name of the medicine) from the medicine images as highlighted by this thesis. Input is a RGB or grey-scale digital image which can be noisy. The picture may be under natural conditions of light. The illumination may be such that the image text suffers from reflection, bad illumination, intensity and contrast across stroke of characters. The orientation of this text is unconstrained. Formally the problem can be stated as the development of algorithm for reliable text detection, localization and extraction of the text (medicine name) from unconstrained image data. Figure 1.5 represents the text extraction model.

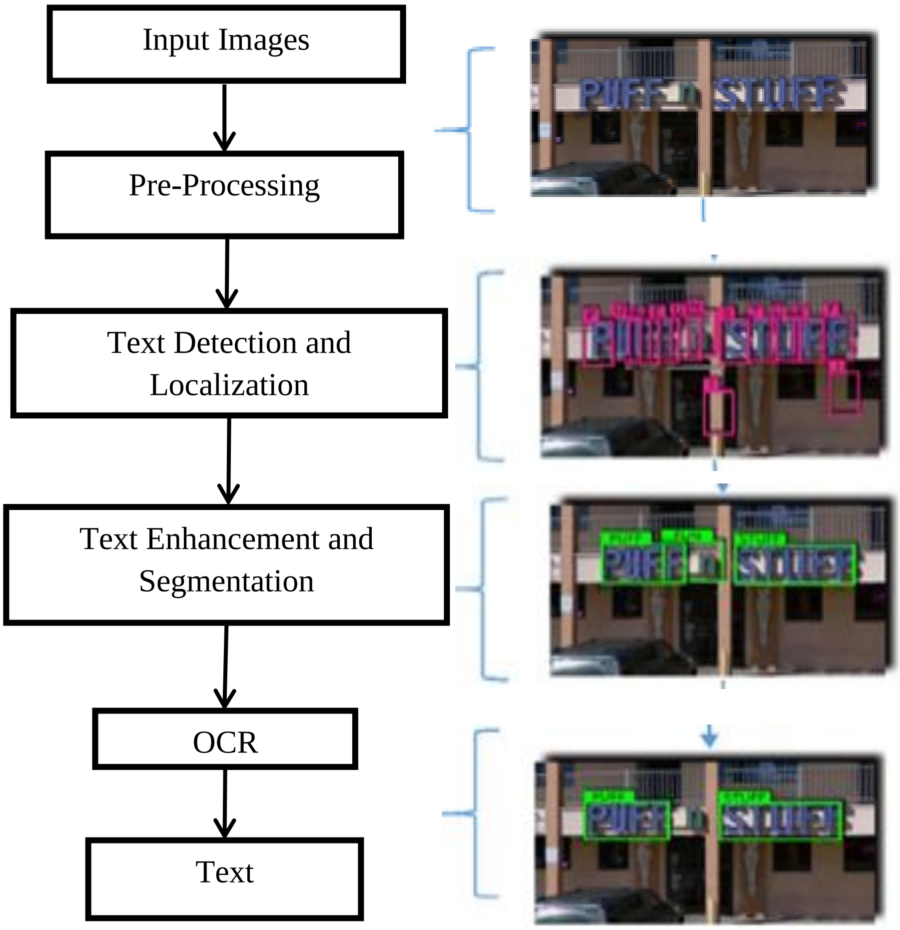


Figure 1.5 Text information extraction model [1].

The problem is broken down to the following sub problems:

1.4.1 Detection

Text detection is the first and the most radical step that indicates whether or not the text exists in the image. This step provides the clarity i.e. weather further processing should be carried out or not by providing binary answers to the question of text existence. A faster detection is a clear advantage. Detection stage is a pre-processing step that seeks to perceive the text occurrence in images. After digitization there are small colour variations that are minimal and not visible to the naked eye. This stage must give preference to the false alarms errors rather than detection rate errors that are missed. The false alarm errors can be rejected in later stages.

1.4.2 Localization

The aim of text localization is to find all the areas in the image that will be considered as text by humans, mark their boundaries usually with rectangular boxes around text. Text Localization generally uses the property and features of text. An array of methods can be used for localization that are discussed in detail in chapter 2.

1.4.3 Binarization

There is a need to segment the text once it has been localized in the image. The outcome is the binary image area(s) with visible text pixels. For this purpose textual features have been have been utilised such as uniformity of colour and background contrast. However before this binarized image is passed through the OCR it needs to be processed. Cleaning the image is the purpose of this processing. It includes noise removal before extraction, production of horizontally oriented text by de-skewing non - horizontally oriented one as expected by the OCR, adjusting boxes to overcome changes in font size.

1.4.4 Algorithm Fusion

Clearly the text offers so many variations hence using a single algorithm i.e. a single approach solution is not feasible. A more robust solution is to use multiple algorithms that are combined in an intelligent way. Such approach offers success on a wide variety of images while reducing the risk of one particular method that works on a certain image to fail on the others.

1.4.5 Performance Evaluation

When using an approach to solve the problem, it is of utmost importance to compare it with the existing methods. It is important to select methods from literature and compare them against an established dataset. An evaluation technique needs to be devised for such comparisons.

1.4.6 Optical Character Recognition and String Editing for Text in Image

Optical Character Recognition (OCR) refers to electronic conversion of images containing text into machine encoded text. For the OCR to work properly text and background should be monochrome. Further the contrast to the background should be high. OCR is complex and unpredictable. It often results in missing and extraneous characters thus in this case OCR is not our final step. It has been taken a step ahead for string editing (matching) in order to cover the OCR mistakes. String editing operation is used to find the optimal correspondence of string using appropriate cost functions. It changes incorrect string to correct string using minimum operations i.e., insertions, deletions and substitutions.

1.5 DESCRIPTION OF RESEARCH

1.5.1 Motivation

A study of literature reveals that no complete text extraction and detection system has been created for medicine strips and bottled medicines. Extraction of text from strips and bottles is an untouched application. This can help the companies to create their medicine database and also assist the patients with medicine information i.e. salts present in medicine and possible substitutes to the medicine. Further no single algorithm can robustly detect the text in these medicine images due to variety present in text as presented in section 1.3. Study of the existing methods in literature for text extraction from natural scenes discussed in detail in chapter-2, reveal that the methods are based on the idea that there is high difference between background and text or even text is composed of a single colour or grey level or that it is a major image component.

Use of such rigid assumptions about the nature of text in images forms a weak heuristic. This study of state of art was the motivation behind the development of framework for text extraction from medicines.

To sum up the major challenges found during the text (name of the medicine) extraction from medicine images are:

1. Recognition of text exclusively in the text areas. The text in strips exists in huge variety of font, colours, orientations, alignments, distortions, complex background, reflection etc.
2. Current OCR systems can work for text against plain single colour background and cannot extract text from complex textured background.

To cope up with these challenges a text detection and extraction algorithm has been developed and to specifically deal with second task a step beyond the OCR system is taken i.e. the string editing. This will help to generate the correct string of medicine name.

1.5.2 Scope of the Dissertation

In Chapter-2, we review the literature of text extraction process that includes detection and localization, text extraction and enhancement, and Optical Character Recognition. A detailed description of all detection and localization methods that includes edge based method, CC-based method, Texture based method, Morphology based and sliding window based methods. Followed by methodologies used for text extraction and Optical Character Recognition.

In Chapter-3, we put forth the text detection algorithm for medicine strips and bottled medicines. The algorithm uses a combination of edge detection algorithm, MSER detector, geometric filtering and SWT algorithm. The last step of algorithm involves the optical character recognition followed by string editing operation for improved results.

In Chapter-4, we display the results that are obtained after applying the proposed the text detection algorithm on the data set of medicine strips and bottles. Performance enhancement for OCR is also provided in this chapter.

Chapter-5, gives the conclusion and future scope for the dissertation providing all observations and applications that can be considered for future.

CHAPTER 2

LITERATURE REVIEW

This chapter is a brief sum up of all the previous work done for text recognition and extraction. Various methods exist for artificial text recognition in images. However the main concentration here lies on the scene text images and the approaches that can be utilized for extraction of this text. According to the study of literature the text extraction process is divided into three sub problems:

1) Text detection and localization
2) Text extraction and enhancement
3) Optical Character Recognition (OCR)

2.1 TEXT DETECTION AND LOCALIZATION

Aim of text detection and localization is to find all the regions in the image which will be considered text by humans and mark their boundaries usually using the rectangular boxes. Although the location is shown by the boxes, removal of background is a pre-requisite. The existing methods can be broken down into five groups: edge based, connected component based, sliding window based, texture based, and morphology based.

2.1.1 Edge based methods

Edges are a definite feature for detection of text. Edge based method [5and 10] pivots the idea of searching those areas that have high contrast between background and text. A morphological operation follows the edge detector to pull out text from the background and reject regions with no text. Images that have strong edges result in good performance however a major pitfall of this method is strong edges may not be obtained under shadow or reflection.

Ou *et al***.** [3] proposed a multi-scale text extraction algorithm that was edge based. This algorithm used three characteristics i.e. strength of edges, orientation and density of text embedded in images leading to the detection and extraction of text in complex images. It worked well with respect to colour, alignment, font and orientation of text and can be used for a number of applications such as licence plate recognition, object identification and mobile robot navigation**. Ren** *et al.* [4] put forth the method that computes edges using Roberts operator. This operator is utilised to process a binary image according to the self-adjusting threshold. Next an erosion operator in morphology is used to eliminate the outliers. Final step involves finding the text region and text extraction by focusing function.

Chen *et al.* [6] utilized a canny edge operator for edge determination in the image. The main focus is the reduction of computational complexity by using a single edge point in the small window for determining scale and orientation. Text edges are enhanced and edges are connected into clusters by morphological operations. Non text clusters are removed by properties such as aspect ratio and height. Two types of Gabor asymmetric filters are used. A suitable scale is further used to enhance the edge information. Structures that do not satisfy specific scales are either eliminated or blurred.

Xiaoqing Liu [7] put forth a framework with 3 sub tasks: i) Text region identification ii) Locating text iii) Extraction of text. First step uses the magnitude of second derivative of intensity in order to calculate the edge strength. This edge strength helps to calculate the edge density. This edge strength helps to calculate the edge density. Variance of orientation is evaluated by four orientations ($0^{\circ}, 45^{\circ}, 90^{\circ}, 135^{\circ}$) where at 0° *and* 90° are horizontal and vertical directions respectively, and at 45°*and* 135° are two diagonal directions. In second step clustering techniques are used for localization. Third step utilizes existing OCR engines which provide good results on documents and images with plain background. Figure 2.1 represents the result of the above mentioned algorithm.

(a)Original image (b) Result after extraction

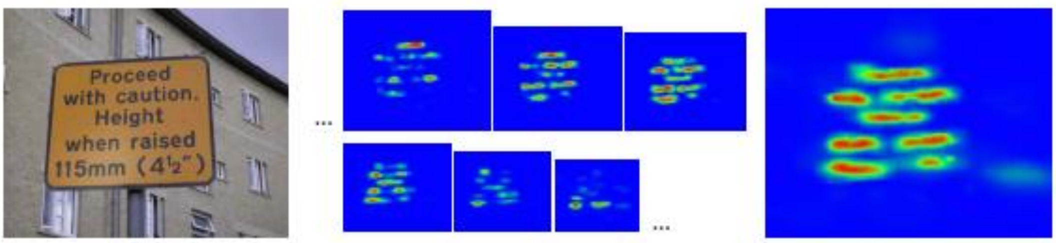
Figure 2.1 Edge detection results [7].

2.1.2 Connected component based methods

Connected Component methods [8, 9 and 11] are used to extract character candidates (set of pixels with similar text properties) from input images. The candidates are refined in order to remove the false positives and grouped into final text. Maximally Stable Extremal Regions (MSER) and Stroke Width Transform (SWT) are popular techniques for connected component analysis. These methods have provided remarkable performance in text detection in natural scenes.

Ohya *et al.* [12] introduced one of the first methods for character localization. They put forth a method in which grey scale image is subjected to local self-adjusting threshold. The outcome is the candidate regions and the regions with high contrast are chosen as characters. **Li** *et al.* [13] replaced grey scale with quantized colour space and hence applied thresholding. They created text blocks by grouping individual characters by basic alignment rules. Both the methods are based on the belief that characters are upright and suffer from no rotation or skew-ness, and that the contrast with the background is high, which is sufficient for road signs and license plate, but not for general localization.

Pan *et al.* [14, 15] use a grey scale image pyramid to create a text recognition map (Figure 2.2) by using a Waldboost classifier based on HOG features. Niblack's binarization algorithm [41] detects the candidate regions and Conditional Random Field (CRF) performs the labelling i.e. weather regions are text or non-text with the text confidence as the main feature. Blocks of text are formed by minimization of gradient graph energy. The method is computationally expensive and localization performance is still in doubt.



(a) (b) (c)

Figure 2.2 Confidence Map. (a) Sign board image. (b) Image pyramid text confidence map. (c) Map for sign board image [14].

(a) (b)

Figure 2.3 Stroke Width Transform [16] (a) Original image in Grey scale (b) Each pixel value replaced by the stroke width

Epshtein *et al.* [16] presented Stroke Width Transform (SWT) i.e., an effective image operator. This operator replaces colour value of each pixel in the image with the stroke width of that pixel. The need for multi-scale computation is eliminated. Important feature of this algorithm is that it's not font or language dependent. Figure 2.3 represents the results of this paper.

Nueman and Matas [17] proposed a technique for text detection and identification in real world images. They exploited MSER which was insensitive to shapes and brightness. The method was tested on two datasets with a detection rate of 72% being achieved on char74K and a f-number of 0.57 is obtained on IDCAR 2003.

Chen *et al.* [18] combined MSER with Canny edges. Canny edges remove the MSER outlier pixels for detection of small letters. For removing the non-text regions left in the image SWT is applied. Finally the letter candidates are clustered into lines and false positives are removed by additional checks.

2.1.3 Texture based methods

Texture based methods [19, 20, and 23] suggest that text in images has special textual properties that are different from the background properties. Usually features are calculated over the region of interest and a trained classifier (via machine learning) is utilised to identify text. Gaussian filtering, Discrete Cosine Transform (DCT), Wavelet decomposition and Local Binary Pattern (LBP) are the most common textual approaches. Detection is possible even when the image is noisy however the shortcoming of this method is its slow speed and sensitivity to orientation and alignment of text.

Zhou *et al*. [21] demonstrated a framework for multilingual text detection. Three different textual features are used: HOG (Histogram of Gradient), MG (Mean of Gradients) and LBP (Local Binary Pattern) in order to find the text region regardless of the language type. Finally, to combine the effect of different features a cascade AdaBoost Classifier is used to decide the text regions**. Pan** *et al.* [22] used a novel technique for location of text in natural images, it was a very fast method that combined training based filtering of regions followed by checking using coarse- to -fine strategy. For filtering, a boosted classifier or polynomial classifier is used and for verification five features such as HOG, LBP, DCT, Gabor and wavelets are utilised.

Angadi and Kodabagi [24] proposed a novel text detection method based on high pass filtering in DCT domain to enhance the foreground. Text regions are identified by feature vectors on uniformity and contrast. This algorithm has a detection rate of 0.966 on 100 low resolution images.

Zhong *et al.* [25] proposed a method to locate high variance text regions in grey scale image using special local variance. Pixel variance is computed using a 1*21 horizontal window Location of horizontal edge by canny detector is a prime step followed by merging of edge components that are small in long lines. Using this edge image pairs are created, opposite direction edges are grouped into lower and upper boundaries of text lines. This method works well for images with horizontal components.

Sin *et al.* [26] use horizontal and vertical direction edge pixels and spectrum (Fourier) to determine regions of text presence in real scene images. Along with the use of frequency features, they assume that only rectangular text regions are present against the background. The edges are detected by rectangular search followed by Hough transform. No clear indication of three stage merger is provided.

2.1.4 Sliding window based methods

Sliding window based methods [27, 28, 29, 30, and 33] work by shifting a window to search for text in a given scene image. Then for each window, weather the location contains text or not is determined via the machine learning techniques. Low level features such as histogram, image gradients or Wavelet Coefficients are used to train the classifier. Since these methods work in multiple scales these are slow. Even though the recall rate is high, the classification is prone to false positives due to large number of candidates. To overcome this issue classifiers like support vector machine and random forest [20, 21] have been suggested however this method remains unsuitable for real time applications due the heavy computations required in window shifting.

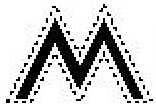
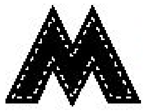
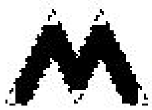
Chen and Yuille [31] combined five features: mean intensity, derivatives, intensity variations, features based on edge linking and histogram along with Ada boost classifier Segmentation is done using a variant of Niblack's algorithm. Apart from being expensive the method needs manual segmentation of several sub windows for training as shown in Figure 2.4 .Also the performance in localization is doubtful. The method [31] is improved by **Pan** *et al.* [32] by the use of HOG and LBP features in the stage of text detection.

Figure 2.4 Positive training data samples used for training the sliding window classifier [31].

2.1.5 Morphology based methods

Morphology is a wide range of image processing operations that process the image on the basis of shape. They are affecting structure, form or shape of an object. MO's are generally applied to binary images and are used for pre or post processing for getting the shape of object/region description. Binary image and structural elements are the two inputs for the morphological operation; these are combined using set operations such as intersection, union, or compliment. In a nutshell, objects are processed in input image based on shape characteristics encoded in structuring element. This method gives promising results for character recognition because it is independent of operations such as translation, rotation scaling, and change in text and light conditions. Most commonly used morphological operations are demonstrated in Table 2.1 below.

Table 2.1 Set of Morphological operations and their mathematical formulations.

Operation	Function	Effect on image	Mathematical formulation
Erosion	Erodes the boundary of foreground pixels; foreground area shrinks making holes in the area become large.		$G(x, y) = I(x, y) \ominus S_{m.n}$
Dilation	Enlarges the boundary of foreground pixels; Foreground pixels enlarge making holes become small.		$G(x, y) = I(x, y) \oplus S_{m.n}$
Opening	Hybrid process; erosion followed by dilation, structured removal of boundary pixels.		$I(x, y) \circ S_{m,n} = (I(x, y) \ominus S_{m,n}) \oplus S_{m,n}$
Closing	Hybrid process; dilation followed by erosion, structured filling in of the boundary pixels		$I(x, y) \bullet S_{m,n} = (I(x, y) \oplus S_{m,n}) \ominus S_{m,n}$

Hasan and Karam [34] put forth a morphological method for text extraction. The RGB space components (Red, Green and Blue) are put together to form an intensity image Y.

$$Y = 0.299\ R + 0.587\ G + 0.114\ B \tag{2.1}$$

This approach works well on colour images however the trouble arises in case of grey scale image. Once the colour conversion is performed, the morphological operator is used for edge identification. After thresholding a binary edge image is obtained. Edges close in space are put to morphological dilation to form candidate regions, while erosion removes smaller components. The method is immune to noise.

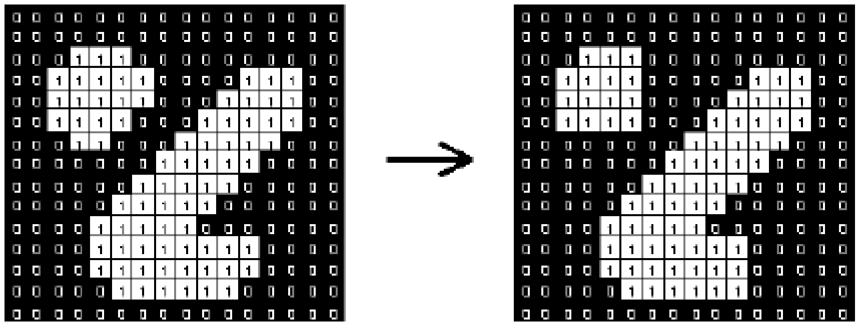


Figure 2.5 Effect of morphological opening using 3*3 square structural element.

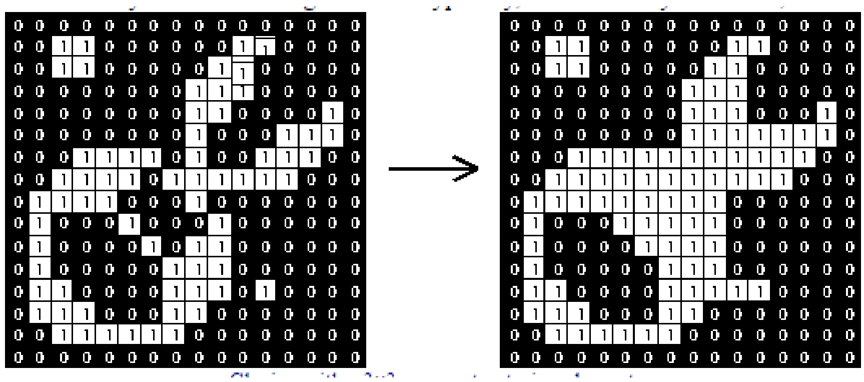


Figure 2.6 Effect of morphological closing using 3*3 square structural element.

2.2 TEXT EXTRACTION AND SEGMENTATION

Low resolution, distortion, blur complex layout and interaction of text and background are some of the ill effects that camera captured images suffer from. Existing OCR work well for document text on plain background. However OCR gives absurd result for text recognition from natural scenes. Thus the text extraction pulls the text from the background and the enhancement focuses on providing a clean image for the OCR that is free from noise and resolution.

Zhou *et al.* [35] put forth improved binarization method for document analysis. First, noise is removed from the given image using local statistics based Wiener filter followed by foreground region estimation. Neighbouring pixel interpolation is used to compute the background value. Finally the pre-processed image is combined with the background surface. Method shows a good performance on uneven illumination. The experiment carried out on printed document is shown in Figure 2.7.

Anand Mishra *et al*. [36] used a sliding window approach to detect locations of characters in the image of word. Next, the most likely word is found out by energy minimization framework. Experiments are carried out on recognising words that are cropped and detection improvement of 15% and 10% is achieved on the street view text WORD and ICDAR data sets.

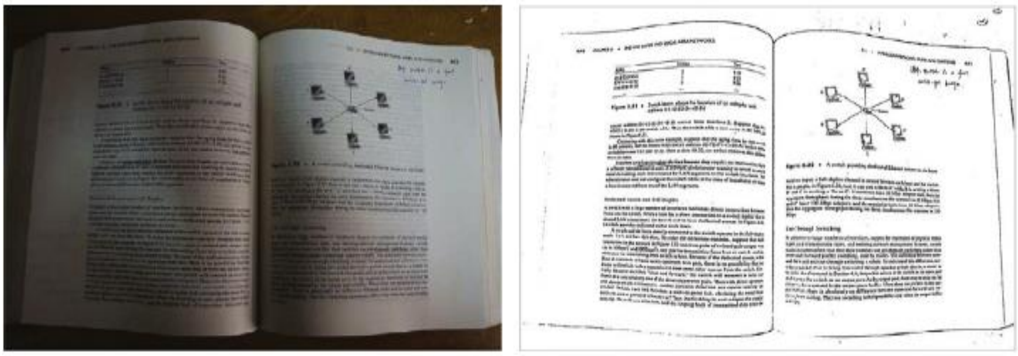


Figure 2.7 Distortion printed document image analysis [36].

Liu and Wang [37] used adaptive threshold for binarization of image. Otsu's threshold is utilised to calculate the best threshold for optimization. Further optimization is achieved by new algorithm for boundary elimination. A comparative analysis of the new and old methods is put forth which indicates that new methods have a better performance. New method keeps the original edge feature and available for wider applications. **Huang** *et al.* [38] proposed an HMM system for extraction from scanned documents. Method was based on detection of foreground and background pixels and morphological analysis for immediate OCR accuracy.

2.3 OPTICAL CHARACTER RECOGNITION

Scene text recognition can benefit from post processing to improve the OCR accuracy just like printed document images (OCR) .This process is a typical one as the scene text is short without much relation to its surrounding text. **Beaufort and Mancas -Thillou** [39] find the sequence of characters that is most probable by weighted Finite State Machine (FSM) from the list of OCR characters. **Donoser et al**.[40] label the individual characters as Maximally Stable Extremal Regions. Further they are classified by using cross relation with trying samples and the most probable characters are selected using standard search engines.

CHAPTER 3

METHODOLOGY

3.1 THE TEXT EXTRACTION AND DETECTION ALGORITHM

The flowchart for text extraction and detection algorithm is depicted in the Figure 3.1. First step involves the selection of either medicine strip or the bottled medicine, for which the user wants to detect the text (name of the medicine). Based on this selection, different algorithms will be applied for edge detection. Subsequently, Maximally Stable Extremal Regions (MSER) are efficiently computed for the image (Section 3.1.1) followed by Laplace of Gaussian (LoG) edge detection for medicine strips and algorithm based edge detection for bottled medicines (Section 3.1.2). To overcome shortcoming of the MSER, we perform the intersection of MSER with LoG edges from the original grey-scale image (Section 3.1.3). Further, the resulting CC's are filtered using geometric parameters such as area, solidity, eccentricity and aspect ratio (Section 3.1.4). As a next step, Stroke Width Transform is applied to remove the non-text regions and to keep the text with uniform stroke (Section 3.1.5). Bounding Boxes are created around the text regions, which are provided as an input to the OCR (Section 3.1.6). The OCR uses Stroke Width Transformed image and the region of interest in order to output the text (Section 3.1.7). Finally, to cope with missing and extraneous characters in the OCR output, a string editing algorithm is proposed (Section 3.1.7).

A brief summary of text detection algorithm:

1) *Detect text candidate regions using MSER:* Text in images exists in consistent colour and high contrast thus the MSER feature descriptor gives satisfactory results for finding the text regions.

2) *Non-text region removal by geometric filtering:* Along with the text regions the algorithm for MSER detects the non-text regions. These stable non-text regions are filtered out by using simple threshold on geometric properties.

3) *Non-text regions removal using stroke width transform:* Stroke width refers to the thickness of lines and curves of text letters. Text regions show little variation in stroke width, whereas larger variations are witnessed in non-text regions.

4) *Merge Text Regions by morphology:* The results of geometric and stroke width filtering involve single characters. These individual characters need to be put together to form the words. This can further help in recognition of text by the OCR.

5) *Recognize Detected Text using OCR and string editing:* `OCR` function is used to recognize the text within the region of interest. Moving a step forward string editing performed that helps to recognise the missing/ extraneous characters in the OCR output.

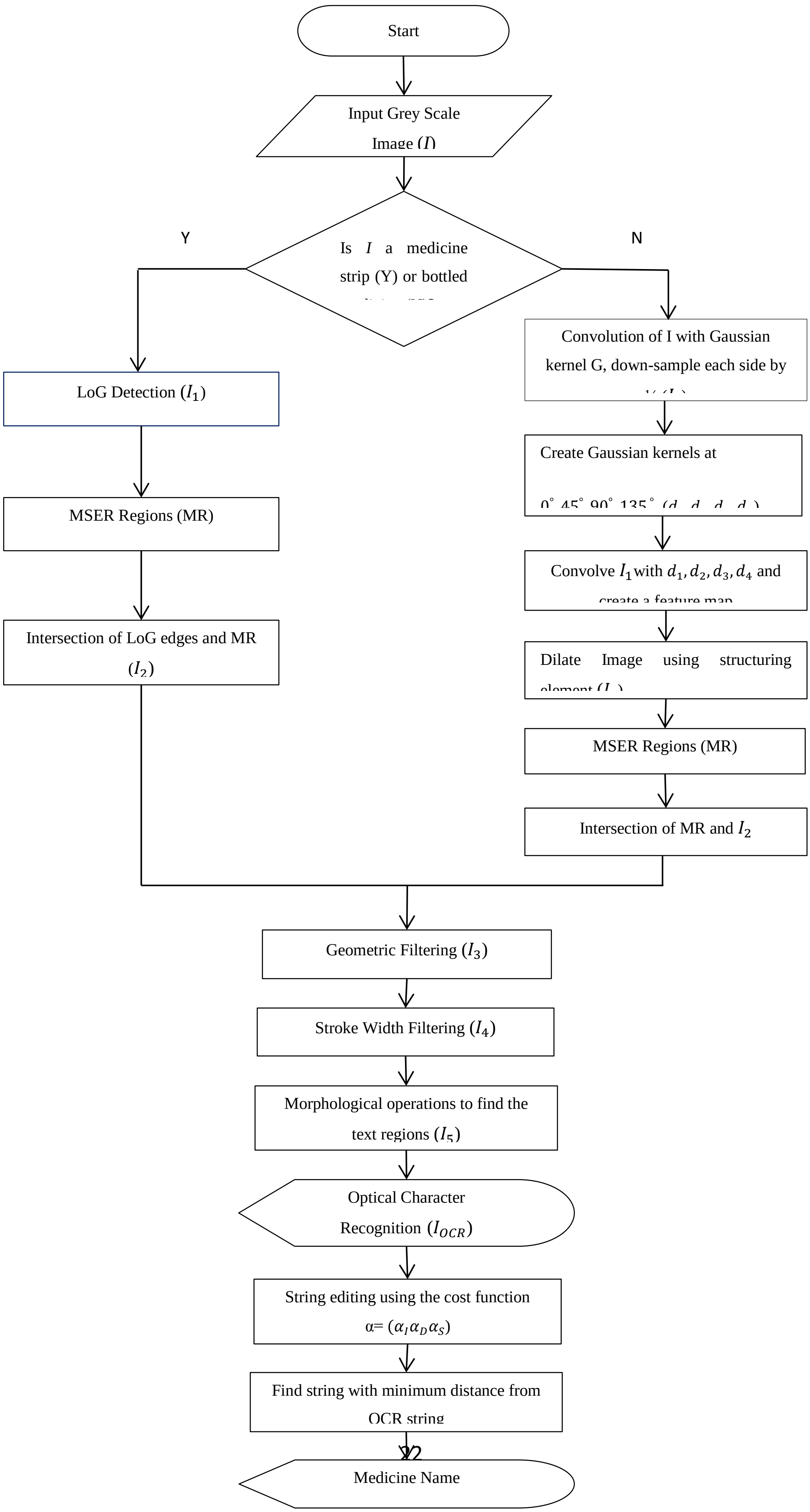
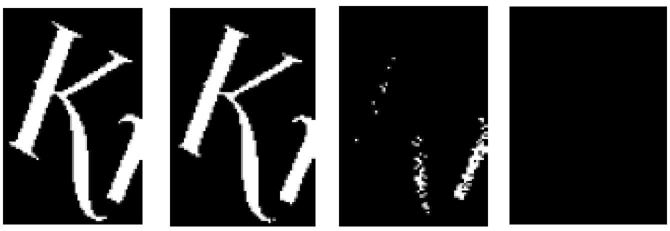


Figure 3.1 Flowchart representation of the text extraction and detection algorithm. $\mathbf{I_1}$, $\mathbf{I_2}$ etc. represent the images obtained after application of intermediate steps and $\mathbf{I_{OCR}}$ is output containing the required text region. The final string i.e., the medicine name is obtained after a string editing operation.

3.1.1 Maximally Stable Extremal Region (MSER)

Maximally stable extremal regions in grey level images are stable connected components. The algorithm considers those regions that undergo minimal change through a huge number of thresholds. The image text has distinctive characteristics relating to frequency, alignment and cohesion similarity. Due to these properties of text, MSER becomes the natural choice for text detection [44]. Consider all possible threholding values of a grey scale image I. Threshold t is the value above which all pixels are black while below are white. As the value of t is varied, a sequence of thresholding images I_t is obtained. The first image is white; as t is increased black spots grow till a completely black image (Figure 3.2) is obtained at a particular threshold. Set of all connected components obtained in the above mentioned process are extremal regions and regions with minimum variation are maximally stable.

(a) Input (b) t=75 (c) t=105 (d) t=135



(e) t=165 (f) t=195 (h) t=225 (i) t=25

Figure 3.2 Results for the image at threshold level t.

The extremal region is mathematically defined by Equation (3.1).

$$R_t = \{x| \; I(x) \; \forall x \in R_t, \forall y \in B(R_t)\} \quad (3.1)$$

where x and y are pixel indices of grey scale image I, t is the threshold value and B (R_t) is set of boundary pixels. The MSER is described by Equation (3.2)

$$\text{r (t)} = |R_{t+\Delta} \setminus R_{t-\Delta}| / \; |R_t| \quad (3.2)$$

where $|.|$ denotes the cardinality, R_t is the extremal region obtained at threshold t and Δ is a stability range parameter. $R_{t+\Delta}$ *and* $R_{t-\Delta}$ are extremal regions found by going upward and downward respectively until a region within t+Δ and t-Δ is found. MSER corresponds to those nodes that have r as the stability value.

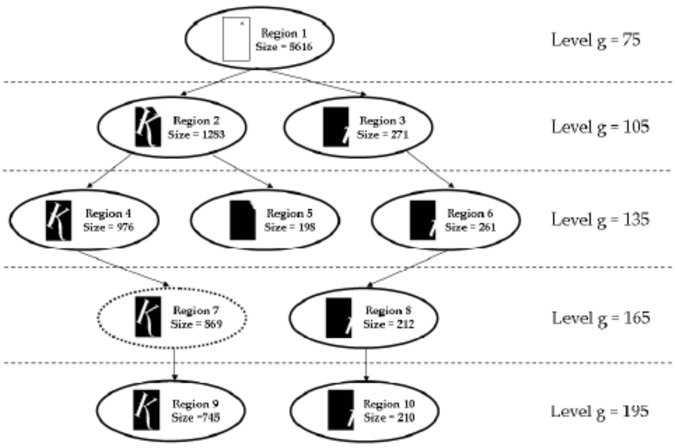


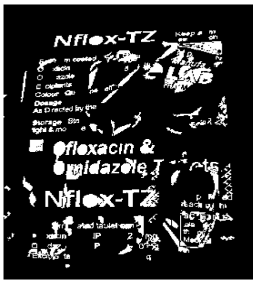
Figure 3.3 Component tree. 7th region with minimal variation is considered as MSER.

The Algorithm description for MSER is as follows:

1) Performing simple luminance, sweep the threshold from black to white in order to sort the pixels by intensity. The sorting preferred in this case is the BINSORT [45]. The range of image values I is small i.e.; {0… 255} leading to the small computational complexity of O(n).
2) After sorting, place the pixels from image either in increasing or decreasing manner and extract the CC's using Union find [45].
3) Find the threshold where extremal regions are "maximally stable".
4) Keep these regions as feature descriptors for further processing.

Figure 3.4 (a)-(d) shows the MSER regions in a medicine strip and a bottled medicine. We have calculated the regions of similar intensity using the MSER region detector.

(a)



(b)

(c)

(d)

Figure 3.4 (a) MSER detected regions for a medicine strip. Colour in the image indicates region that can be considered as text; however some non-text regions have also been detected (b) Binary image obtained after the application of MSER detector (MR). (c) MSER detected regions for a bottled medicine. (d) Binary image for a bottled medicine (MR).

Some of regions (highlighted by different colours) include extra background pixels indicating two major shortcomings of MSER:

1) Several repeating components create problem in character grouping,
2) It's sensitive to image blur.

To cope up with these issues, an intersection of MSER detected regions with edge detected regions is performed.

3.1.2 Edge based detection

Edge based methods search for those areas that have a huge difference between the text and the background. The application of the existing canny edge detection methods [46] to the samples of medicine strip and bottled medicines did not give satisfactory results on the samples. LoG edge detection performs well on medicine strips but not for bottles due to the rotated text. An algorithm for edge detection on medicine bottles is proposed in this section.

3.1.2.1 LoG edge detection for medicine strips

Laplacian is 2D isotropic measure of second order spatial derivative of the image. It finds the rapid intensity regions, hence is very effective for detection of edges. For the Laplacian of image I(x, y) refer to Equation (3.3).

$$L(x, y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \quad (3.3)$$

Since Laplacian definition involves second derivative, convolution kernels (refer Figure 3.5) are used.

0	-1	0
-1	4	-1
0	-1	0

-1	-1	-1
-1	8	-1
-1	-1	-1

Figure 3.5 Discrete approximation to the Laplacian filter

Due to noise sensitivity of the kernels, Gaussian smoothening is applied to image before Laplacian filter. In LoG operation Gaussian smoothening filter is convolved with Laplacian filter followed by convolution of hybrid image to get desired result. Mathematical formulation of 2D LoG function is given by Equation (3.4).

$$LoG(x, y) = \frac{-1}{\pi\sigma^4}\left[1 - \frac{x^2+y^2}{2\sigma^2}\right] e^{\frac{-x^2+y^2}{2\sigma^2}} \quad (3.4)$$

This function is approximated by discrete kernel (figure 3.6)

0	1	1	2	2	2	1	1	0
1	2	4	5	5	5	4	2	1
1	4	5	3	0	3	5	4	1
2	5	3	-12	-24	-12	3	5	2
2	5	0	-24	-40	-24	0	5	2
2	5	3	-12	-24	-12	3	5	2
1	4	5	3	0	3	5	4	1
1	2	4	5	5	5	4	2	1
0	1	1	2	2	2	1	1	0

Figure 3.6 Discrete kernel for LoG operation with Gaussian 1.4

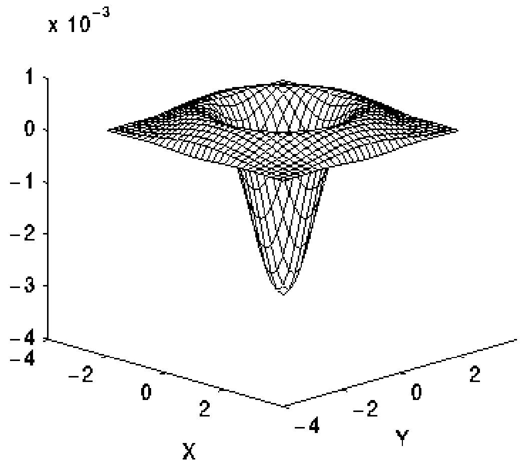


Figure 3.7 2-D LoG function

LoG operator is applied to the medicine strip image; it produces zero response in the areas of constant intensity. Hence it is likely to obtain the text regions along with some false negatives. The Log edge detection for medicine strip is shown in the Figure 3.11.

3.1.2.2 Edge detection algorithm for bottled medicines

Table 3.1 Algorithm description of edge based detection in bottled medicines.

Algorithm Edge detection in bottled medicines

Input: Binary Bottled medicine image BI
Output: Edge detected image for bottled medicines EI
DirectionKernels ← d_0, d_{45}, d_{90}, d_{135};
G ← Gaussian kernel created by fspecial;
Level(1) ← BI;
for i = 2 to 4 **do**
BI ← Convolve (G, BI);
BI ← Down sample each BI by 0.5;
Level(i) ← BI;
end for
for m ← 1 to 4 **do**
for n ←1 to 4 **do**
C(m,n) ← Convolve (Level(m), DirectionKernels(n));
end for
end for
for m = 1 to 4 **do**
total(m) ← C(1,m) + C(2,m) + C(3,m) +C(4,m);
end for
DI ← Dilate(total);
S1 ← total(1,1) and total(1,3);
S2 ← total(1,2) and total(1,4);
S ← S1+S2;
FeatureMap ← DI and S;
TextRegion ← Dilate(FeatureMap) ;
return EI ← TextRegion;

The bottled medicine has slightly rotated (or skewed) text thus LoG edges do not provide appropriate results .The algorithm for edge detection is provided in Table 3.1.

0.0113	0.0838	0.0113
0.0838	0.6193	0.0838
0.0113	0.0838	0.0113

Figure 3.8 Gaussian kernel generated by fspecial in Matlab

Figure 3.9 Four levels of Gaussian pyramid as mentioned in algorithm

-1	-1	-1
2	2	2
-1	-1	-1

-1	-1	2
-1	2	-1
2	-1	-1

-1	2	-1
-1	2	-1
-1	2	-1

2	-1	-1
-1	2	-1
-1	-1	2

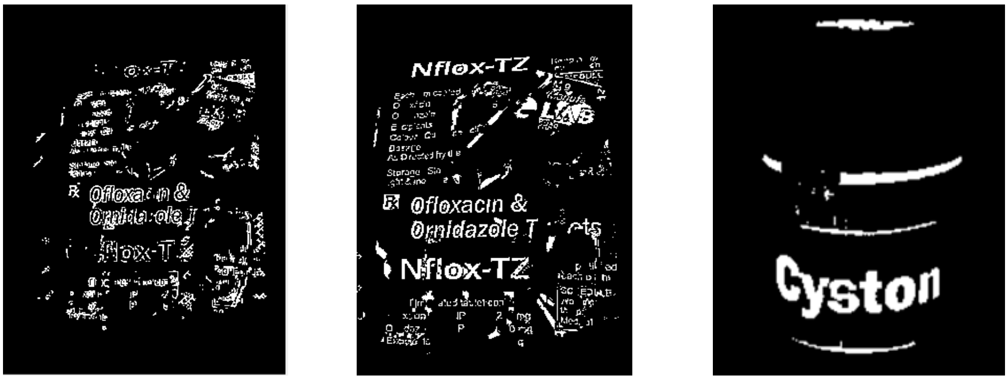
Figure 3.10 The directional kernels representing $0^{\circ}, 45^{\circ}, 90^{\circ}, 135^{\circ}$

(a) (b)

Figure 3.11 Edge detection. (a) LoG edge detection for a medicine strip (I_1) (b) edge detection based on stated algorithm for a bottled medicine (I_1).

3.1.3 Intersection of MSER and LoG edges

Only application of MSER to the image is not enough since it is sensitive to blur and unable to differentiate small letters. To overcome this problem we perform an intersection of MSER and LoG edges as MSER pixel boundaries are removed using LoG edges. For light text and dark background (or vice versa), the gradient direction is adapted (Table 3.2), to generate pixels towards the background refer Figure 3.12 (a). Another advantage of intersection is that the repeating components are lesser in number and the non-text regions are also reduced, refer Figure 3.12 (b)-(c).



(a) (b) (c)

Figure 3.12 Edge enhanced MSER. (a) Edges grown along gradient directions. (b) Intersection of MSER detected regions with LoG regions (I_2). Clearly, the number of non-text regions has reduced as compared to Figure 3(b). (c) Intersection of MSER regions with edge detected regions for a bottled medicine (I_2).

Table 3.2 Brief description of relationship between edge and gradient direction for text

S.no	Text	Background	Edge and Gradient Directions
1.	Dark	Light	Edges grown **against** the gradient direction
2.	Light	Dark	Edges grown **along** the gradient direction

3.1.4 Geometric Filtering

A binary image whose foreground CCs are considered letter candidates is obtained after performing the edge- enhanced MSER detection. In order to further filter out the non-text regions geometric checks are performed on the CC. First of all, taking area as parameter very small and very large objects are rejected. For our sample images, regions with area less than 150 sq. Pixels and greater than 2000 sq. Pixels are rejected.

Aspect ratio is another parameter under consideration; aspect ratio threshold is chosen carefully in order to preserve elongated letters like 'i', 'l' and 'j'. Character candidates with very large aspect ratio and small ratio are rejected.

Simple geometric features cannot discriminate shapes with large distances, thus we chose to combine these features with other complimentary shape descriptors. The shape descriptors we used are Solidity and Eccentricity. Solidity defines the extent to which the shape is convex or concave, given by Equation (3.5).

$$\text{Solidity} = \frac{A_S}{H} \tag{3.5}$$

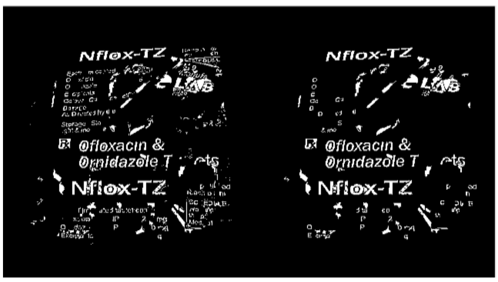
Where A_S the shape of area region and H is is the convex hull of shape. From the experiments we found that the threshold that responds well on the medicine samples is 0.4. Eccentricity is the ratio of the length of major axis to the minor axis. It is calculated using minimum bounding rectangular box.

$$\text{Eccentricity} = \frac{L}{W} \tag{3.6}$$

Where L is the length of the bounding box and W is the width of the bounding box. From the experiments we found that threshold for eccentricity is 0.995. The Figure 3.13 below shows that after the geometric checks the non-text regions have further reduced.

Table 3.3 A brief description of geometric parameters and threshold for medicine samples.

S. No	Geometric Parameter	Threshold used for medicine strip samples
1.	Area	$> 500 \ and < 2000$
2.	Solidity	< 0.4
3.	Eccentricity	< 0.995



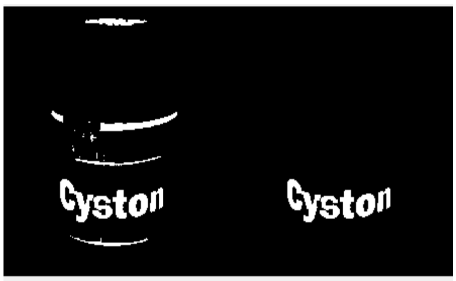


Figure 3.13 Image represents the difference in a medicine strip image and a bottled medicine, before and after region filtering (I_3). The geometric checks have considerably reduced the number of non-text regions, when compared with the previous image.

3.1.5 Stroke Width Transform

After the geometric checks there still are some non- text regions left, for further processing we use Stroke Width Transform (SWT). SWT [16] replaces colour value of each pixel to its most probable stroke value. The idea behind using this transform is that the characters of text have same stroke thickness, thus any region in which stroke width exists with large variation is removed.

The computation of SWT [16] involves sending rays along the perpendicular directions through edge pixels. Only the rays that ended at other edge pixels were considered; however this does not give accurate results for non-parallel stroke edges. SWT is thus determined using the distance transform [23] as in Equation (3.8).

$$I_3(x,y) \in \{\text{fo, bg}\} \tag{3.7}$$

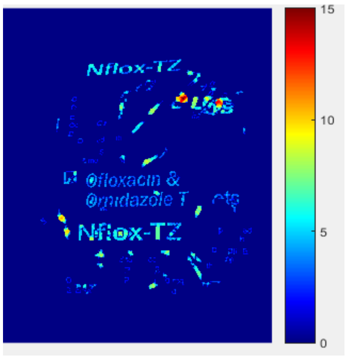
where fo and bg represent the foreground and background pixels respectively.

$$I_{4d}(x,y) = \begin{cases} 0, & I_3(x,y) \in bg \\ \min(\|x - x_0,\ y - y_0\|, & \forall\, I_3(x_0,y_0) \in bg, I_3(x,y) \in fo \end{cases} \tag{3.8}$$

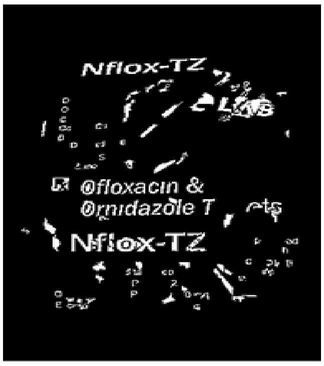
where ||x, y|| is the Euclidean distance. Using this method the stroke width value is given at each pixel of unique CC. Euclidian distance transform is based on distance calculation from foreground to nearest background pixel. The maximum stroke value is computed by Equation (3.9).

$$\text{mStroke= Max } (I_{4d}(x,y)) \tag{3.9}$$

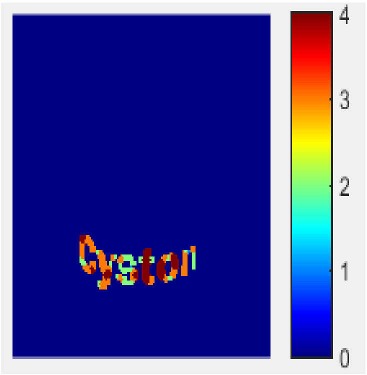
Look up for all the 8 neighbours of the foreground pixel that share the same value lying in the range of 1 to mStroke. The algorithm results in a stroke width transformed image (I_4). Every pixel in transformed image has value of half stroke width. The criteria for rejection that works well for the medicine strips is given as $\frac{std}{mean} > 3$. The result image Figure 3.14 (a)-(d) shows pixels labelled with stroke width which can further used for grouping pixels into letter candidates.



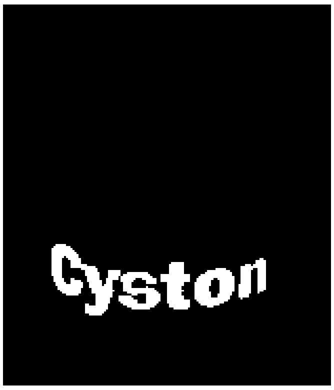
(a)



(b)



(c)



(d)

Figure 3.14 Stroke Width Transform Information. (a) Visualization of text candidate's stroke width for medicine strip. (b) Binary Image after application of Stroke Width Transform (I_4). (c) Visualization of text candidate's stroke width for bottled medicine (d) Binary Image after application of Stroke Width Transform.

3.1.6 Determining Text Regions using Bounding Boxes

In order to calculate the bounding boxes around text regions, the individual characters are merged into a single CC. Morphological operations such as opening and closing are used to remove the outliers. Closing is a morphological operation that enlarges the boundaries of foreground regions and shrinks the background colour holes. Closing needs two data inputs, closing image and structuring element. Structuring elements are shown below in Figure 3.15.

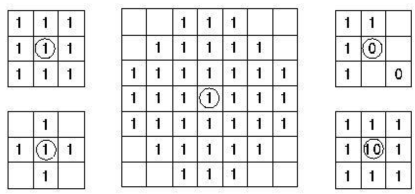


Figure 3.15 Different types of structuring elements.

The mathematical formulation for closing and opening are given by Equation (3.10 & 3.11).

$$I(x, y) \bullet S_{m,n} = (I(x, y) \oplus S_{m,n}) \ominus S_{m,n} \qquad (3.10)$$

$$I(x, y) \circ S_{m,n} = (I(x, y) \ominus S_{m,n}) \oplus S_{m,n} \qquad (3.11)$$

where $S_{m,n}$ is a structuring element, m and n are odd integers > zero. Any structuring element which is rotationally independent can be chosen (Figure 3.16). Once the morphology mask is obtained as in Figure 3.16 (a)-(b), area threshold (5000 sq. pixels are used for medicine samples) is applied to create bounding boxes around the text regions (Figure 3.17(c)).

(a) (b) (c)

Figure 3.16 (a)-(b) Joined characters under mask for strip and bottled medicines, respectively (c) Text (I_5) region (medicine name).

3.1.7 Optical Character Recognition and String Editing

3.1.7.1 Optical Character Recognition

Optical Character Recognition (OCR) means electronic conversion of images containing text into machine encoded text. OCR system requires the input image such that the letters are passed and perceived. For OCR to work properly, the text and the background should be monochrome and the contrast between the two should be high. The two inputs that we provide to OCR, is the image obtained after stroke width transform as shown in Figure 3.15 (b),(d) and the region of interest; i.e., the bounding boxes as shown in Figure 3.17(c). The output of OCR is shown in Figure 3.17(a)-(b).

```
G010 'I '
53 @floxacm 82
⊗;mdazo'le T

N flox-TZ
```

```
°vston
```

(a) (b)

Figure 3.17 OCR output (a)- (b) represent the output of Figure 2(a) and 2(c) respectively. Clearly the output has missing and extraneous charecters(I_{OCR}).

3.1.7.2 String Editing Operation

OCR is complex and unpredictable. The common mistakes made by OCR on the medicine samples and the bottled medicines are missing and extraneous characters as shown in Figure 3.19 (a)-(b). A large of number of strings is detected by OCR. In order to accurately determine the text, we keep the string with word confidence greater than 70%, this OCR text string is compared with the correct string. Due to the missing and extraneous characters it cannot be assumed that the i^{th} character of the OCR string corresponds to the i^{th} character of correct string. *Levenshtein distance* is used for string editing by calculating optimal correspondence of strings using appropriate cost function. It changes the incorrect string to correct one using minimum number of operations; insertions, deletions or substitutions.

Consider a string s_1= $[a_1\ a_2\ a_3\ \ldots\ldots.a_n]$ and s_2= $[b_1\ b_2\ b_3\ \ldots\ldots.b_m]$ where m and n symbols from alphabet A. Three operations that can be applied as:

1) Insertion: Any symbol x $\in A$ can be inserted before a_1, after a_n or between a_i and a_{i+1} (1≤ i ≤ n).
2) Deletion: Symbol a_i can be deleted (1≤ i ≤n).
3) Substitution: The symbol a_i can be replaced by x≠ a_i (1≤ i ≤n).

Let $d_{s1,s2,\alpha}$ be the Levenshtein distance between s_1 and s_2 that denotes the least cost of transforming string s_1 to string s_2 using cost function α. For Levenshtein distance, α is (1,1,1) or (1,1,2). $d_{s1,s2,\alpha}$ is a Levenshtein metric as it satisfies the following properties of a metric.

1) $d_{s1,s2,\alpha} > 0$ if $s_1 \neq s_2$; $d_{s1,s1,\alpha}$= 0;
2) $d_{s1,s2,\alpha}$ = $d_{s2,s1,\alpha}$;
3) $d_{s1,s3,\alpha} \leq\ d_{s1,s2,\alpha} + d_{s2,s3,\alpha}$

The Levenshtein distance between two strings, s_1 *and* s_2 of length i and j respectively is mathematically denoted in Equation (8).

$$d_{s_1,s_2}(i,j) = \begin{cases} max(i,j) & if\ min(i,j) = 0, \\ min\begin{cases} d_{s_1,s_2}(i-1,j)+1 \\ d_{s_1,s_2}(i,j-1)+1 \\ d_{s_1,s_2}(i-1,j-1)+1_{(s_1 \neq s_2)} \end{cases} & otherwise \end{cases} \quad (3.12)$$

where $1_{(s_1 \neq s_2)}$ is the indicator function equal to zero when $s_1 = s_2$ and equal to 1 otherwise. The cost function for edit operation is represented as α = $(\alpha_I\, \alpha_D \alpha_S)$, where $\alpha_I, \alpha_D, \alpha_S$ are non- negative real numbers corresponding to insertion, deletion and substitution respectively. The cost function for edit operation is fixed (1, 1, 1); i.e., unit cost operation. Edit distance algorithm is given by Table 3.4.

Table 3.4 Algorithm description of string editing operation.

Algorithm String Editing Operation

STRING EDIT DISTANCE ($s_1[a_1\ a_2\ a_3\ \ldots\ldots a_n]$, $s_2[b_1\ b_2\ b_3\ \ldots\ldots b_n]$)
int d [i, j] =0
f**or** i ← 1 to $|s_1|$
do d [i, 0] = i
for j ← 1 to $|s_2|$
do d [0, j] = j
for i ← 1 to $|s_1|$
for j ← 1 to $|s_2|$
if $s_1 == s_2$
do d [i, j] ← d [i-1, j-1]
else d [i, j]← *minimum of*
(
d [i-1, j] +1,
d [i, j-1] +1,
d [i-1, j-1]+1
)
return d [$|s_1|$, $|s_2|$]

Clearly, if the edit distance is minimum when strings are matched, the strings bear similarity. OCR strings for medicine samples are matched with the database of strings. The string, with the minimum edit distance is considered the correct string and it is displayed. Thus the string editing operation overcomes the shortcomings of OCR.

Figure 3.18 represents the flowchart for string editing operation with two examples i.e., Nflox-TZ and Cystone. The confusion region and the number of required string editing operations is displayed in boxes.

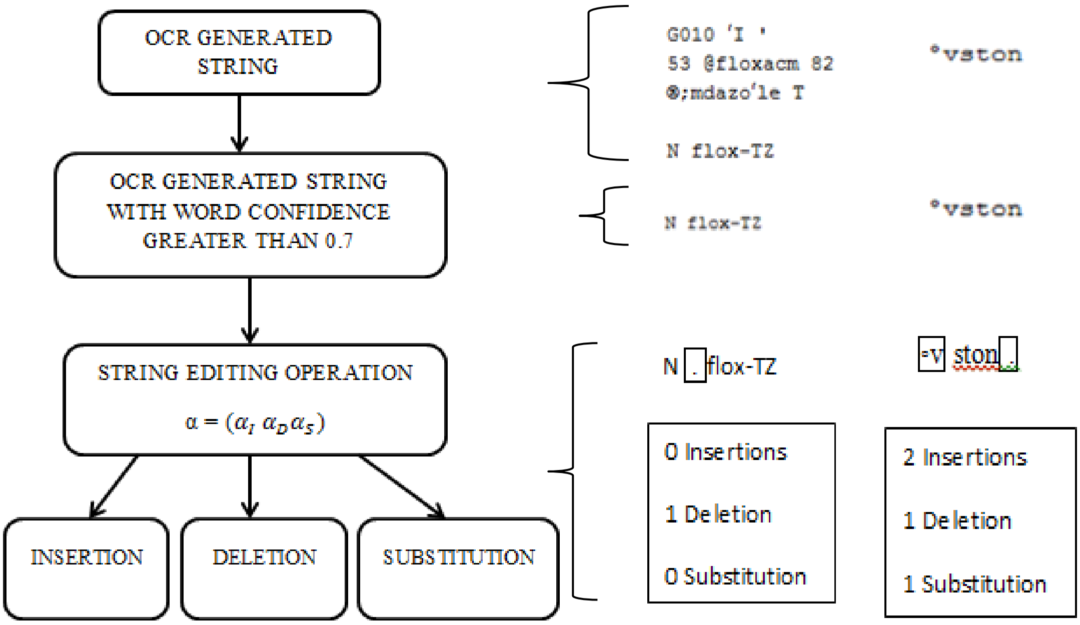


Figure 3.18 A flowchart representation of OCR and string editing operation with examples of two samples i.e., Nflox-TZ and Cystone.

CHAPTER-4

EVALUATION OF TEXT DETECTION ALGORITHM

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

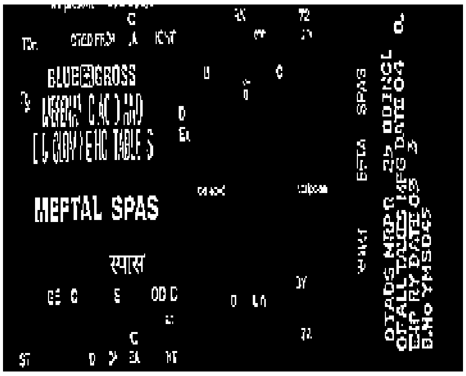
4.1 DATASET AND EXPERIMENTAL RESULTS

To assess our text extraction and detection algorithm we test it on a dataset of medicine samples containing both medicine strips and bottled medicines. We implemented our algorithm using MATLAB R2015a. Since there is no existing database, we created the database with a wide variety of medicines from different pharmacies using a regular camera under natural lighting conditions. The images have been captured at different moments of the day i.e., morning, noon and evening using Motorola G (Gen2) that packs an 8 megapixel primary camera on the rear with size of each image approximately $900 \times x$ ($500 \leq x \leq 1600$) pixels. The captured images have variations of light intensity, colour, alignment of text, camera angles, font, and size. Variations of camera angles indicate that the images have been captured by free hand; i.e., there are no specific angles for capturing the images thus making the technique robust and easy to use for the layman.

The state of art methods use recall rate and precision rate [1, 29 and 30] as the performance evaluation criteria for the OCR; however in this paper we have extended the problem beyond the OCR and performed string editing. The result we obtained has complete medicine name or an erroneous result. Figure below shows samples of the medicine database with variations such as blur-ness, reflection or bad illumination. Text variations in these medicine samples include; light text on dark background (or vice versa), thick text or thin text with elongated characters such as 'i' and 'l' and curved or skewed text as in case of bottled medicines. Each figure consists of original image, image obtained after application of MSER detector, geometric filtering and Stroke Width filtering, region of interest obtained from morphological operation and the OCR detected text.

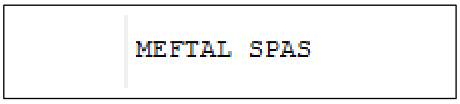
By exploiting the similarity of text, our method can effectively localize a variety of text. The system so developed works very well on medicine strips and bottled medicines, we were able to extract and display the medicine name even in case of reflection, blur, bad illumination and skew-ness as depicted by figures and Table below.

(a)Medicine image



(b) SWT image

(c) Text ROI



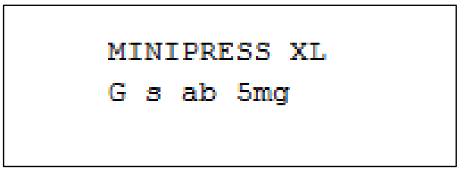
(d) OCR output

(A)

(a) Medicine image

(b) SWT image

(c) Text ROI

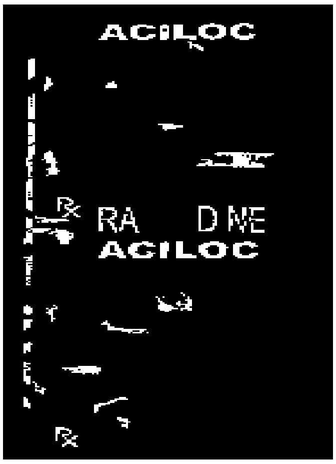


(d) OCR output

(B)

Figure 4.1(A)-(B). Text detection algorithm results.

(a) Medicine image



(b) SWT image

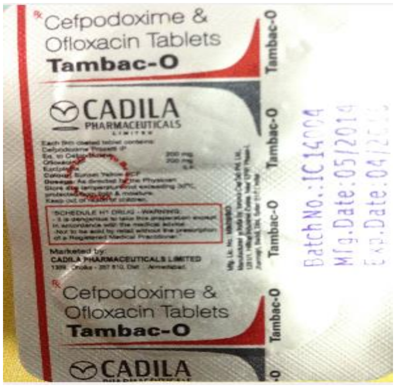
(c) Text ROI

- Ac|Loc

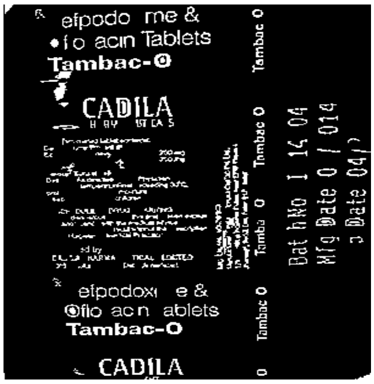
Acu.hoc

(d) OCR Output

(A)



(a) Medicine image



(b) SWT Image

(c) Text ROI

@fIo acn ablets
Tambac-O

CADQA

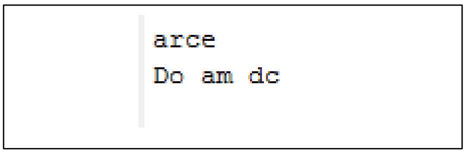
(d) OCR Output

(B)

(a) Medicine image

(b) SWT Image

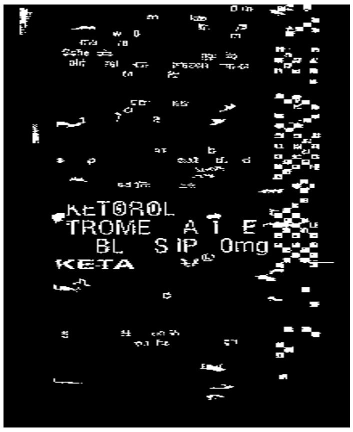
(c) Text ROI



(d) OCR output

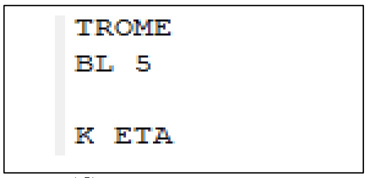
(C)

(a) Medicine Image



(b) SWT Image

(c) Text ROI

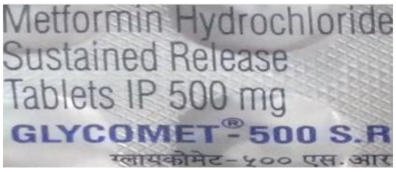


(d) OCR Output

(D)

(a) Medicine Image

(b) SWT Image



(c) Text ROI

```
anedRe a e
Tahes Pfiflflmg

GLYCOMETOSDU s H
```

(d) OCR Output

(E)

(a) Medicine Image

(b) SWT image

(c) Txt ROI

```
Hm»; 2.? m T
BESTO§£'@'V-625
```

(d) OCR Output

(F)

Figure 4.2 (A-F) Text detection results for medicine strip images with reflection.

(a) Medicine Image

(b) SWT Image

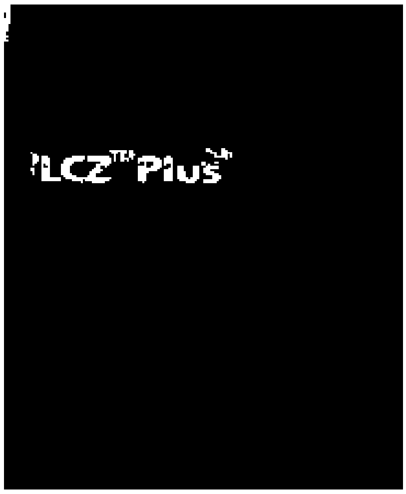
(c) Text ROI

SPOR DEX AF}u7_§O

(d) OCR output

(A)

(a) Medicine Image



(b) SWT Image

(c) Text ROI

Lcz"':

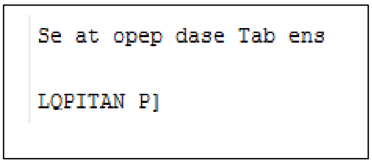
(d) OCR output

(B)

(a) Medicine Image

(b) SWT image

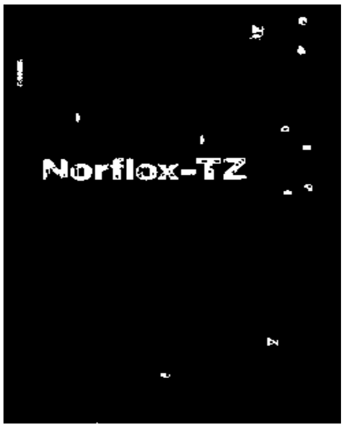
(c) Text ROI



(d) OCR output

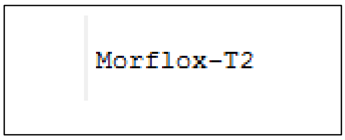
(C)

(a) Medicine Image



(b) SWT Image

(c) Text ROI



(d) OCR output

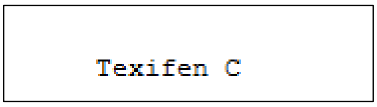
(D)

Figure 4.3(A-D) Text detection results for medicine images with same coloured text as background. In order to detect this text region HSV colour space is used with the saturation channel giving the apt results.

(a) Medicine Image

(b) SWT Image

(c) Text ROI



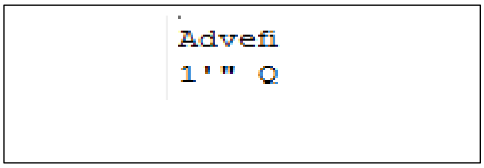
(d) OCR output

(A)

(a) Medicine Image

(b) SWT Image

(c) Text ROI

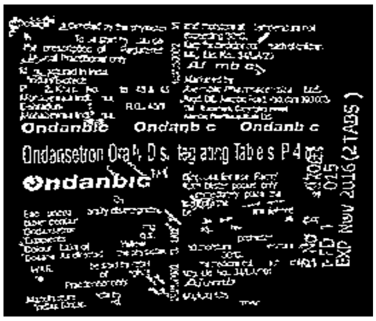


(d) OCR output

(B)

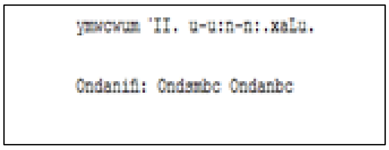
Figure 4.4 (A) - (B) Text detection results for medicine images with light text on dark background.

(a) Medicine Image



(b) SWT Image

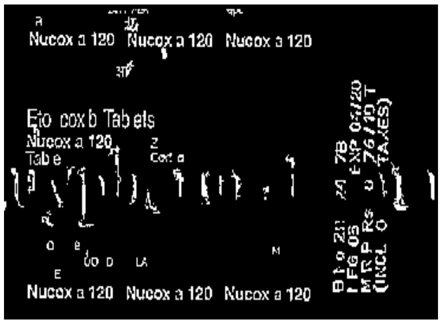
(c) Text ROI



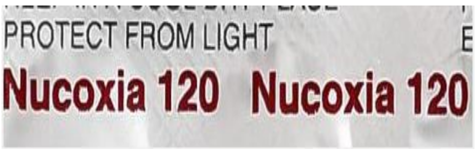
(d) OCR output

(A)

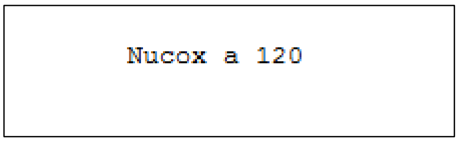
(a) Medicine Image



(b) SWT Image



(c) Text ROI



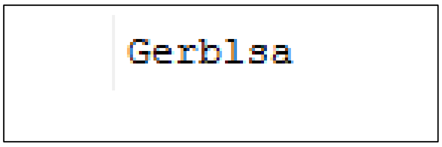
(d) OCR Output

(B)

(a) Medicine Image

(b) SWT Image

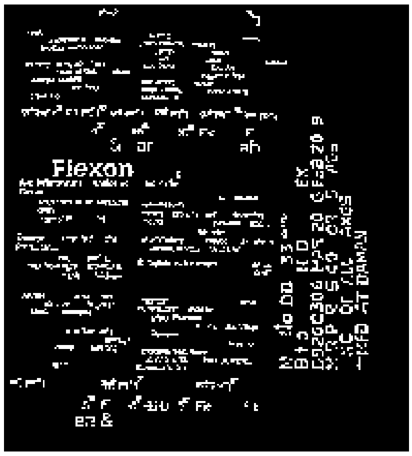
(c) Text ROI



(d) OCR Output

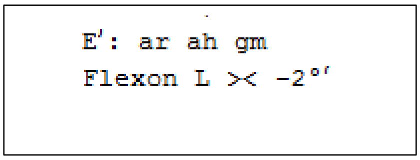
(C)

(a) Medicine Image



(b) SWT Image

(c) Text ROI

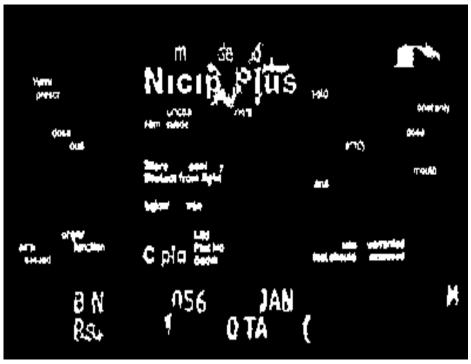


(d) OCR Output

(D)

Figure 4.5 (A-D) Text detection result on medicine strip images with small text and also containing elongated characters like ‘i’ and ‘l’. These letters are often lost in detection.

(a) Medicine Image



(b) SWT Image

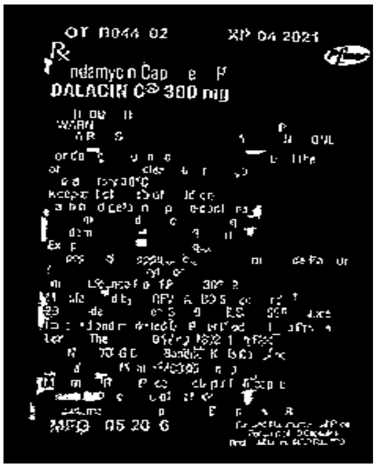
(c) Text ROI

```
"' 'E "A.
NICI P us
:.-.....-r.-2
WU -on
```

(d) OCR output

(A)

(a) Medicine Image



(b) SWT image

(c) Text ROI

```
DALACIH G9 300
```

(d) OCR Output

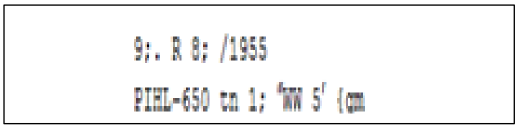
(B)

Figure 4.6 (A)-(B) Text detection results for badly illuminated medicine images.

(a) Medicine Image

(b) SWT Image

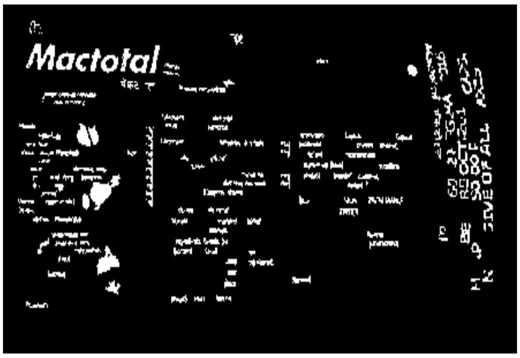
(c) Text ROI



(d) OCR Output

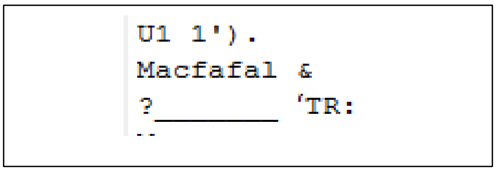
(A)

(a) Medicine Image



(b) SWT Image

(c) Text ROI



(d) OCR Output

(B)

Figure 4.7(A)-(B) Text detection results on medicine strip images with thick text words.

(a) Medicine Image

(b) SWT Image

(c) Text ROI

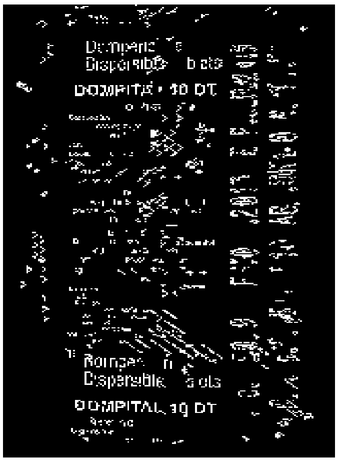
```
G010 'I '
53 @floxacm 82
®;mdazo'le T

N flox-TZ
```

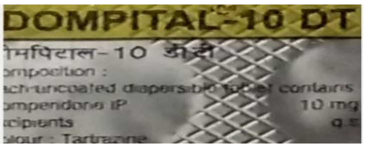
(d) OCR Output

(A)

(a) Medicine Image



(b) SWT Image



(c) Text ROI

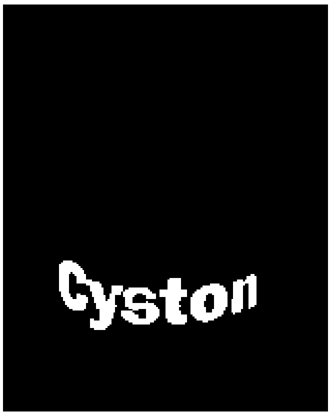
```
DOMPITAL 1g 01
Wm 10
```

(d) OCR Output

(B)

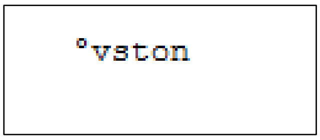
Figure 4.8(A)-(B) Text detection in medicine strips with coloured background.

(a) Medicine Image



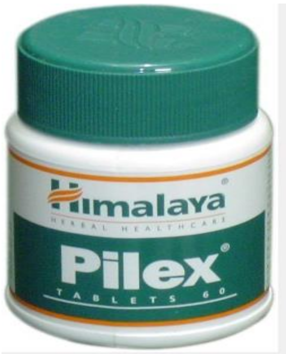
(b) SWT Image

(c) Text ROI



(d) OCR Output

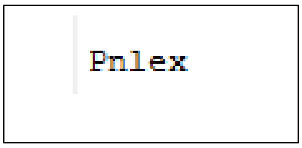
(A)



(a) Medicine Image

(b) SWT Image

(c) Text ROI



(d) OCR Output

(B)

Figure 4.9(A)-(B) Text detection algorithm applied to bottled medicines.

4.2 PERFORMANCE ENHANCEMENT

The text obtained from OCR depicted by (d) of figures above has missing and extraneous characters. As mentioned, we have extended the algorithm beyond the OCR to make up for the flawed characters and hence obtain the correct string. OCR detects multiple strings, to get the correct string, we keep the string for which word confidence is greater than 70% while rejecting the other strings. For example consider the medicine strip in Figure 4.8 (A), PML - 650, the detected OCR string is PIHL-650 (word confidence > 70%) that has extra characters I and H. Table 4.1 below summarizes the string matching (editing) that compares the OCR generated string with the set of strings. String editing calculates the number insertions, deletions and substitutions in order to find the string for which distance from OCR string is minimum. Consider the medicine strip in Figure 4.8 (A) again; OCR generated string 'PIHL-650' when compared to 'PML-650' requires one substitution and one deletion i.e. a total cost of 2. This strings when compared with other strings say, 'HOPE, requires 4 deletions and 4 substitutions i.e. a total cost of 8. So the string with minimum cost is considered the correct detected string. The confusion column in the Table 4.1 shows rectangular boxes around the missing/ extraneous characters. The next three columns show string editing operations; insertion, deletion and substitution. Finally, the last column indicates the correct string i.e. the medicine name. Text extraction and detection algorithm gives an accuracy of approximately 95%.

Table 4.1 String editing operation. OCR string with word confidence >0.7 is compared with other strings and number of insertions (I), deletions (D) and substitutions (S) are calculated. The string with minimum operations i.e., least cost is considered as correct string.

Image	OCR text with word confidence greater than 0.7	Confusion	I	D	S	Correct detected string
4.1 (A)	MEFTAL-SPAS	-	0	0	0	MEFTAL-SPAS
4.1 (B)	MINIPRESS XL	-	0	0	0	MINIPRESS XL
4.2 (A)	-AC|Loc	-AC|Loc	0	1	1	ACILOC
4.2 (B)	Tambac-o	-	0	0	0	Tambac-o
4.2 (C)	Do am dc	Do .am. dc	1	0	1	Dolamide

4.2 (D)	KETA	KETA.	3	0	0	KETANOV
4.2 (E)	GLYCOMETQS DU	GLYCOMET QSDU	0	4	0	GLYCOMET
4.2 (F)	BESTOS¥@V-625	BESTOS¥@V-625	0	0	4	BESTOCLAV-625
4.3 (A)	SPORD X AF	SPOR. DX AF	1	0	0	SPORIDEX AF
4.3 (B)	Lcz	Lcz.	4	0	0	Lcz Plus
4.3 (C)	LOPITAN P	LOPITAN. P	1	0	0	LOPITAN SP
4.3 (D)	Morflox-T2	Morflox-T2	0	0	2	Norflox-TZ
4.4 (A)	Texifen C	Texifen C	1	0	0	Texifen
4.4 (B)	Advefi	Advefi	0	0	2	Advent
4.5 (A)	Ondabc	Ondab. c	1	0	0	Ondabic
4.5 (B)	Nucox a 120	Nucox. a 120	1	0	0	Nucoxia 120
4.5 (C)	Gerblsa	Gerb l sa	0	0	1	Gerbisa
4.5 (D)	Flexon	-	0	0	0	Flexon
4.6 (A)	NICI P US	NICI. P. US	2	0	0	NICIP PLUS
4.6 (B)	DALACIH 300	DALACIH 300	0	0	1	DALACIN 300
4.7 (A)	PIHL- 650	P IH L-650	0	1	1	PML-650
4.7 (B)	Macfafal	Macf af al	0	0	2	Mactotal
4.8 (A)	N flox-TZ	N. flox-TZ	0	1	0	Nflox-TZ
4.8 (B)	DOMPITAL	-	0	0	0	DOMPITAL
4.9 (A)	◦vston	◦v ston.	1	0	2	Cystone
4.9 (B)	Pnlex	P n lex	0	0	1	Pilex

CHAPTER-5

CONCLUSION AND FUTURE SCOPE

Text extraction and identification from medicine strips and bottled medicines in itself is a novel application. The text data on medicine can be buried in image with different fonts, styles, sizes, alignment, and orientation. Further adding to this variety is the presence of reflection, blurriness, rotated text or complex background thus making the text extraction from medicines a very challenging task. This dissertation provides the text extraction and detection algorithm for medicines.

As per the observation, methods based on edge detection use morphology algorithms thus show less efficiency. Consequently, these methods give satisfactory results for complicated backgrounds and are regarded as secondary method to other algorithms. For localization of text regions, CC and texture based methods are more efficient as they lay emphasis on local pixels. However, performance of CC based methods depends on accurate determination of text regions. Thus the issue here is confusion between regions with text and no text. Contrarily texture methods use features fed to classifiers to find text regions. The problem faced here is sensitivity to alignment of text

It can be observed that a single method is not suitable for all the cases. However, a combination of several methods provides solution along with some localization. For text extraction and detection in medicine images i.e., medicine strips and bottled medicines, we utilized connected-component and stroke based methods. In order to recognise the text, it has to be separated from the background. This extracted text (binary image) is fed into the OCR. OCR helps to convert extracted text to plain text.

Our text detection algorithm provides a fast and robust detection for the medicine strip and bottled medicine images. Our method employs MSER as the basic letter candidates. Since MSER is sensitive to image blur we enhance it using edge-based algorithms. Next, SWT is effectively computed for binary CC’s and morphological opening and closing are utilized to find the regions of interest i.e., text. Further SWT and morphological output is fed to the OCR in order to get the text i.e. the medicine name. The text from OCR has missing/extraneous characters thus we applied string editing. With the insertion, deletions and substitution operations we chose to display the string that had minimum distance from the OCR string. The system gives an efficiency of approximately 95% on medicine strips and bottled medicines images captured under various conditions; for example, reflection, bad illumination and skew-ness. The method is capable of detecting highly blurred text in low resolution medicine images as well as rotated text for the bottled medicines.

With the proposed method we obtained promising results on our data set. Medicine industry is a huge. We can further include the medicine samples and bottled medicines with text in different languages. The algorithm can be modified to detect text on various medicines that have it written in different calligraphic styles; as well languages like Chinese require special attention. The algorithm can further be extended to provide applications such as assisting the patients with finding the medicine substitutes and the salts in the medicine. The medicine images do not contain the barcode; another interesting application would be to find the text hidden behind other objects or the watermarked logos of manufacturing companies hidden in the medicine images.

REFERENCES

[1] H. Zhang, K. Zhao, Y. Z. Song, and J. Guo, "Text extraction from natural scene image: A survey," Neurocomputing, 2013, vol. 122, pp.310–323.

[2] K. Jung, K. Jim, and A.K. Jain, "Text Information extraction in images and video: A Survey," Pattern Recognition, vol.5, 2004, pp. 977-997.

[3] W. Ou, J. Zhu, C. Liu, "Text location in natural scene," Journal of Chinese Information Processing, 2004.

[4] Y. Cui, J. Yang, D. Liang, "An edge-based approach for sign text extraction," Image Technology, 2006, vol. 1.

[5] Y. Cui, J. Yang, D. Liang, "An edge-based approach for sign text extraction," Image Technology, 2006, vol. 1.

[6] D. Chen, K. Shearer, and H. Bourlard, "Text Enhancement with Asymmetric Filter for Video OCR," International Conference on Image Analysis and processing, 200, pp. 192-197.

[7] R. Zanibibi and D. Blostein, "Recognition and Retrieval of Mathematical Expressions," International Journal on Document Analysis and Recognition (IJDAR), 2012, vol. 15(4), pp. 331-357.

[8] Z. Liu, S. Sarkar, "Robust outdoor text detection using text intensity and shape features," 19th International Conference on Pattern Recognition, ICPR 2008, IEEE, 2008, pp.1–4.

[9] H. Zhang, C. Liu, C. Yang, X. Ding, K.Q. Wang, "An improved scene text extraction method using conditional random field and optical character recognition," International Conference on Document Analysis and Recognition (ICDAR), IEEE, 2011, pp.708–712.

[10] F. E. Aasami,"Recognition of Ultra Low Resolution, Anti-aliased Text with Small Font Sizes," Ph.D. Thesis, University of Fribourg, Switzerland, 2008.

[11] H. Wang, J. Kangas, "Character-like region verification for extracting text in scene images," Proceedings Sixth International Conference on Document Analysis and Recognition, IEEE, 2001, pp.957–962.

[12] J. Ohya, A. Shio, and S. Akamstu, "Recognizing Characters in scene images," Pattern Analysis and Machine Intelligence IEEE Transections, Feb .1994, vol. 16(4), pp. 214-220.

[13] C. Li, X. Ding, and Y. Wu, "Automatic Text Location in Natural Scene Images," 6^{th} International Conference on Document Analysis and Recognition, 2001, pp. 1069-1073.

[14] Y. -I. Pan, X. Hou, and C. –L. Liu, "A robust system to detect and localize texts in natural scene images," The Eighth IAPR International workshop in Document Analysis systems, Sept. 2008, pp. 35-42.

[15] Y. -I. Pan, X. Hou, and C. –L. Liu, "A hybrid approach to detect and localize texts in natural scene images," Image Processing, IEEE, March 2011, vol.20(3), pp.800-813.

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

[17] L. Neumann and J. Matas, "Real-Time text localization and recognition," in Proc. CVPR 2012, 2012, pp. 3538-3545.

[18] H. Chen, S.S. Tsai, G. Schroth, D. M. Chen, R. Grzeszczuk, and B. Girod, "Robust text detection in natural images with edge-enhanced maximally stable extremal regions," 2011, pp.2609–2612.

[19] X. Chen and A.L. Yuille, "A time-efficient cascade for real-time object detection: With applications for visually impaired," in CPVR –Workshops, 2005, pp.28.

[20] T. Saoi, H. Goto, H. Kobayashi, "Text detection in color scene images based on unsupervised clustering of multi-channel wavelet features," Eighth International Conference on Document Analysis and Recognition(ICDAR'05), 2005, vol.2, pp.690–694.

[21] G. Zhou, Y. Liu, Q. Meng, Y. Zhang, "Detecting multilingual text in natural scene," 1^{st} International Symposium on Access Spaces(ISAS), IEEE, 2011, pp. 116–120.

[22] Y. F. Pan, C. L. Liu, X. Hou, "Fast scene text localization by learning-based filtering and verification," 17^{th} IEEE International Conference on Image Processing (ICIP), IEEE, 2010, pp.2269–2272.

[23] S. M. Hanif, L. Prevost, "Text detection and localization in complex scene images using constrained adaboost algorithm," 10^{th} International Conference on Document Analysis and Recognition, ICDAR'09, IEEE, 2009, pp. 1–5.

[24] H. Chen, S.S. Tsai, G. Schroth, D. M. Chen, R. Grzeszczuk, and B. Girod, "Robust text detection in natural images with edge-enhanced maximally stable extremal regions, 2011, pp.2609–2612.

[25] Y. Zhong, K. Karu, and A.K. Jain, "Locating Text in Complex Color Images," Pattern Recognition, 1995, vol. 28(10), pp. 1523-1535.

[26] B. Sin, S. Kim, and B. Cho, "Locating Characters in Scene Images using Frequency Features," International Conference on Pattern Recognition, 2002, vol. 3, pp 489-492.

[27] H. Cheng, X. Jiang, Y. Sun, and J. Wang, " Color segmentation: advances prospects," Pattern Recognition, 2001, vol. 34(12), pp. 2259-2281

[28] 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," IEEE Transection Pattern Learning Machine Intelligence, 2003, vol. 25(12), pp. 1631-1639.

[29] S. M. Hanif and L. Prevost, "Text detection and localization in complex scene images using constrained adaboost algorithm," ICDAR, 2009, pp.1-5.

[30] J. -J. Lee, P. -H. Lee, S. -W. Lee, A. Yuille, and C. Koch, "Adaboost for text detection in natural scene," in Proc. ICDAR, Beijing, China, 2011, pp. 429–434.

[31] X. Chen and A.L. Yuille, "Detecting and reading text in natural scenes," In Proc. CVPR, 2006, pp. 553-560.

[32] Y. –F. Pan, X. Hou, and C. –L. Liu, "A robust system to detect and localize texts in natural scene images," The Eighth IAPR International Workshop on Document Analysis Systems, Sept. 2008, pp. 35-42.

[33] S. Tian, Y. Pan, C. Huang, S. Lu, K. Yu and C.L. Tan, "A unified text detection system in natural scene images," In Proc. ICCV 2015, 2015

[34] Y. M. Y. Hasan and L. J. Karam, "Morphological Text Extraction from Images," IEEE Transactions on Image Processing, 2000, vol. 9(11), pp. 1978-1983.

[35] S. Zhou, C. Liu, Z. Cui, and S. Gong, " An improved adaptive document image binarization method," 2nd International Congress on Image and Signal Processing, 2009, pp. 1-5.

[36] A. Mishra, K. Alahari, and C.V. Jawahar, "Top-down and bottom-up cues for scene text recognition," CVPR'12, 2012, pp. 2687-2964.

[37] J. Liu and C. Wang, "An Algorithm for image binarization based on adaptive threshold," Chinese Disease and Control Conference, CCDC'09, IEEE, 2009, PP. 3958-3962.

[38] S. Huang, M.A. Sid- Ahmed, M. Ahmadi, and I. EI-Feghi, "A binarization method for scanned documents based on hidden Markov model," IEEE International Symposium on Circuits and Systems, ISCAS 2006, IEEE, 2006, p.4.

[39] R. Beaufort and C. Mancas-Thillou, "A weighted finite-state framework for correcting errors in natural scene ocr," Document Analysis and Recognition, ICDAR 2007, Ninth International conference on, Sept. 2007, vol. 2, pp. 889-893.

[40] M. Donoser, H. Bischof, and S. Wagner, “Using web search engines to improve text recognition,” Pattern Recognition, ICPR 2008, , 19th International conference on, Dec. 2008, pp. 1-4.

[41] W. Niblack, “ An introduction to digital image processing.” Strandberg Publishing Company, Birkereod, Denmark,1985.

[42] L. Neumann and J. Matas, “Real-Time text localization and recognition,” in Proc. CVPR 2012, 2012, pp. 3538-3545.

[43] Y. -F. Pan, X. Hou, C. -L. Liu, “A hybrid approach to detect and localize text in natural scene images,” IEEE Trans. Image Processing, 2011, vol. 20(3), pp. 800-813.

[44] K. Mikolajeczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J.Matas, F.Schaffalitzky, T. Kadir, and L. Van Gool, “A comparison of affine region detectors,” International Journal Computer Vision, 2005, vol. 65, pp. 43-72.

[45] J. Matas, O. Chum, M. Urban, and T, Pajdla,“Robust wide baseline stereo from maximally stable extremal regions,” in British Machine Vision Conference, 2002, vol. 1, pp. 384-393.

[46] N.Yadav and V. Kumar,“ A novel technique for automatic retrieval of embedded text from books,” Optik- International Journal for Light and Electron Optics, Oct. 2016, vol. 127(20), pp. 9538-9550.

LIST OF PUBLICATIONS

1) "Design and Implementation of Text Extraction and Detection Algorithm for Bottles and Strip Medicines" Under preparation.

erg

ORIGINALITY REPORT

%17	%12	%16	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Lecture Notes in Computer Science, 2012. Publication	%1
2	citeseerx.ist.psu.edu Internet Source	%1
3	Sahare, Parul, and Sanjay B. Dhok. "Review of Text Extraction Algorithms for Scene-text and Document Images", IETE Technical Review, 2016. Publication	%1
4	Hojin Cho, Myungchul Sung, Bongjin Jun. "Canny Text Detector: Fast and Robust Scene Text Localization Algorithm", 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016 Publication	%1
5	www.ijcaonline.org Internet Source	<%1
6	Rodolfo P. dos Santos. "Text Line Segmentation Based on Morphology and Histogram Projection", 2009 10th International Conference on Document Analysis and Recognition, 07/2009	<%1

Publication

7	www.iis.ee.ic.ac.uk Internet Source	<%1
8	Opitz, Michael, Markus Diem, Stefan Fiel, Florian Kleber, and Robert Sablatnig. "End-to-End Text Recognition Using Local Ternary Patterns, MSER and Deep Convolutional Nets", 2014 11th IAPR International Workshop on Document Analysis Systems, 2014. Publication	<%1
9	docs.com Internet Source	<%1
10	Sun, Lei, and Qiang Huo. "An Improved Component Tree Based Approach to User-Intention Guided Text Extraction from Natural Scene Images", 2013 12th International Conference on Document Analysis and Recognition, 2013. Publication	<%1
11	imlab.jp Internet Source	<%1
12	Tehsin, Samabia Masood, Asif Kausar, Sum. "Fuzzy-based segmentation for variable font-sized text extraction from images/videos. (Research Articl", Mathematical Problems in Engineering, Annual 2014 Issue Publication	<%1

13	NARTKER, THOMAS A.. "BENCHMARKING DIA SYSTEMS", Handbook Of Character Recognition and Document Image Analysis, 1997. Publication	<%1
14	web.stanford.edu Internet Source	<%1
15	www.iapr-tc11.org Internet Source	<%1
16	Seok, Jae-Hyun, and Jin Hyung Kim. "Scene text recognition using a Hough forest implicit shape model and semi-Markov conditional random fields", Pattern Recognition, 2015. Publication	<%1
17	makcura.com Internet Source	<%1
18	Yadav, Niharika, and Vinay Kumar. "A novel technique for automatic retrieval of embedded text from books", Optik - International Journal for Light and Electron Optics, 2016. Publication	<%1
19	Xu-Cheng Yin, , Xuwang Yin, Kaizhu Huang, and Hong-Wei Hao. "Robust Text Detection in Natural Scene Images", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014. Publication	<%1

20	hal.univ-lorraine.fr Internet Source	<%1
21	Yan, J.. "Chinese text location under complex background using Gabor filter and SVM", Neurocomputing, 201110 Publication	<%1
22	www.academypublisher.com Internet Source	<%1
23	H. Bischof. "Efficient Maximally Stable Extremal Region (MSER) Tracking", 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1 (CVPR 06), 2006 Publication	<%1
24	BASAVANNA, M., P. SHIVAKUMARA, S. K. SRIVATSA, and G. HEMANTHA KUMAR. "MULTI-ORIENTED TEXT DETECTION IN SCENE IMAGES", International Journal of Pattern Recognition and Artificial Intelligence, 2012. Publication	<%1
25	Jameson, Jinjuli, and Siti Norul Huda Sheikh Abdullah. "Extraction of arbitrary text in natural scene image based on stroke width transform", 2014 14th International Conference on Intelligent Systems Design and Applications, 2014. Publication	<%1

26 www.bmva.org
Internet Source
<%1

27 pdfs.semanticscholar.org
Internet Source
<%1

28 Wei, Yuanwang, Zhijiang Zhang, Wei Shen, Dan Zeng, Mei Fang, and Shifu Zhou. "Text detection in scene images based on exhaustive segmentation", Signal Processing Image Communication, 2017.
Publication
<%1

29 www.sersc.org
Internet Source
<%1

30 prr.hec.gov.pk
Internet Source
<%1

31 Al Alusi , Ali Ibrahim. "A Model for Handwriting Recognition Using Wavelet Function and Genetic Algorithm = نموذج لتمييز الكتابة اليدوية بإستخدام الدالة الموجية و خوارزمية التوريث", Middle East University, .
Publication
<%1

32 Sethi, Ganesh K., and Rajesh K. Bawa. "Text Information Extraction: in Context of Indian Languages", Advanced Materials Research, 2012.
Publication
<%1

33 opus4.kobv.de
Internet Source
<%1

34 Rumale, A.S., A.J. Patankar, and M.H. Kolekar. "Image text detection using directional edge kernels", National Conference on Signal and Image Processing Applications, 2009. <%1
Publication

35 cvww2011.icg.tugraz.at <%1
Internet Source

36 www.av.it.pt <%1
Internet Source

37 lib.tkk.fi <%1
Internet Source

38 Zhu, Anna, Guoyou Wang, and Yangbo Dong. "Detecting natural scenes text via auto image partition, two-stage grouping and two-layer classification", Pattern Recognition Letters, 2015. <%1
Publication

39 Diamantatos, Paraskevas, Ergina Kavallieratou, and Pilar Gomez-Gil. "Binarization: A Tool for Text Localization", 2014 14th International Conference on Frontiers in Handwriting Recognition, 2014. <%1
Publication

40 www.coursehero.com <%1
Internet Source

41 Advances in Computer Vision and Pattern <%1

Recognition, 2014.
Publication

42 web.donga.ac.kr
Internet Source
<%1

43 Amandeep Kaur, Renu Dhir, Gurpreet Singh Lehal. "A survey on camera-captured scene text detection and extraction: towards Gurmukhi script", International Journal of Multimedia Information Retrieval, 2017
Publication
<%1

44 www.scribd.com
Internet Source
<%1

45 thoth.inrialpes.fr
Internet Source
<%1

46 Wang, Runmin, Nong Sang, Ruolin Wang, and Xiaoqin Kuang. "A hybrid approach for text detection in natural scenes", MIPPR 2013 Pattern Recognition and Computer Vision, 2013.
Publication
<%1

47 www.rroij.com
Internet Source
<%1

48 Jose, Rosmy C., and Diana Davis. "Extraction of text from videos based on layered method", 2015 IEEE International Conference on Engineering and Technology (ICETECH), 2015.
Publication
<%1

49	Yusufu, Tuoerhongjiang, Yiqing Wang, and Xiangzhong Fang. "A Video Text Detection and Tracking System", 2013 IEEE International Symposium on Multimedia, 2013. Publication	<%1
50	Wang, Chuan, and Yimin Zhou. "A New Approach for Text Location Based on SUSAN and SVM", 2013 International Conference on Information Science and Cloud Computing Companion, 2013. Publication	<%1
51	dspace.uah.es Internet Source	<%1
52	thesis.lib.ncu.edu.tw Internet Source	<%1
53	www.ijrat.org Internet Source	<%1
54	journal.iis.sinica.edu.tw Internet Source	<%1
55	homepages.dcc.ufmg.br Internet Source	<%1
56	www.cs.uwaterloo.ca Internet Source	<%1
57	www.ict.griffith.edu.au Internet Source	<%1
	proton.dmkhost.com	

58 Internet Source <%1

59 KRISDA KHANKASIKAM. "Restoration of Degraded Historical Document Image: An Adaptive Multilayer-Information Binarization Technique", Journal of Information Science & Engineering, 2014.
Publication <%1

60 weber.itn.liu.se
Internet Source <%1

61 acikerisim.deu.edu.tr
Internet Source <%1

62 Yi, C., and Y. Tian. "Localizing Text in Scene Images by Boundary Clustering, Stroke Segmentation, and String Fragment Classification", IEEE Transactions on Image Processing, 2012.
Publication <%1

63 cycabtk.gforge.inria.fr
Internet Source <%1

64 Liu, Xinhao, and Naoki Chiba. "Transparent Text Detection and Background Recovery", 2013 2nd IAPR Asian Conference on Pattern Recognition, 2013.
Publication <%1

65 Lukas Neumann. "A Method for Text Localization and Recognition in Real-World Images", Lecture Notes in Computer Science, <%1

2011
Publication

66 research.ijcaonline.org
Internet Source
<%1

67 Jung, K.. "Text information extraction in images and video: a survey", Pattern Recognition, 200405
Publication
<%1

68 ijltet.org
Internet Source
<%1

69 Zhou, Gang, Yuehu Liu, Quan Meng, and Yuanlin Zhang. "Detecting multilingual text in natural scene", 2011 1st International Symposium on Access Spaces (ISAS), 2011.
Publication
<%1

70 SUN, QIAO-YU, and YUE LU. "TEXT LOCATION IN SCENE IMAGES USING VISUAL ATTENTION MODEL", International Journal of Pattern Recognition and Artificial Intelligence, 2012.
Publication
<%1

71 Yu, Chong, Yonghong Song, and Yuanlin Zhang. "Scene text localization using edge analysis and feature pool", Neurocomputing, 2016.
Publication
<%1

72 Shin, Do Kyoung, Jong Min Lee, Yong Min Kim, Young Shik Moon, and Ki Tae Park.
<%1

"Text detection in video sequence using 1-D DCT", The 18th IEEE International Symposium on Consumer Electronics (ISCE 2014), 2014.
Publication

73 Shah, Sanjay, Chintan Modi, and Manisha Patel. "Novel approach for text extraction from natural images using ISEF edge detection", 2011 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC), 2011.
Publication
<%1

74 www.thinkmind.org
Internet Source
<%1

75 data.conferenceworld.in
Internet Source
<%1

76 Zhong, Guoqiang, and Mohamed Cheriet. "Image patches analysis for text block identification", 2012 11th International Conference on Information Science Signal Processing and their Applications (ISSPA), 2012.
Publication
<%1

77 research.aston.ac.uk
Internet Source
<%1

78 repository.um.edu.my
Internet Source
<%1

79 "Pattern Recognition", Springer Nature, 2014

Publication

<%1

80 cmm.ensmp.fr
Internet Source
<%1

81 Daehyun Kim. "Figure Text Extraction in Biomedical Literature", PLoS ONE, 01/13/2011
Publication
<%1

82 V. Monga. "Perceptual Image Hashing Via Feature Points: Performance Evaluation and Tradeoffs", IEEE Transactions on Image Processing, 11/2006
Publication
<%1

83 Sam S. Tsai. "Mobile visual search using image and text features", 2011 Conference Record of the Forty Fifth Asilomar Conference on Signals Systems and Computers (ASILOMAR), 11/2011
Publication
<%1

84 ethesys.library.ttu.edu.tw
Internet Source
<%1

85 Haqiri, Tayyebe, and Federico Poloni. "Methods for verified stabilizing solutions to continuous-time algebraic Riccati equations", Journal of Computational and Applied Mathematics, 2017.
Publication
<%1

Advances in Intelligent Systems and

86 Computing, 2016.
Publication <1%

87 Kwon, Jae-Hyun, Tae-Yong Park, Yun-Tae Kim, Yang-Ho Cho, Yeong-Ho Ha, and Gabriel G. Marcu. "", Color Imaging X Processing Hardcopy and Applications, 2005.
Publication <1%

88 Lecture Notes in Computer Science, 2011.
Publication <1%

89 Yang, Haojin, Bernhard Quehl, and Harald Sack. "A framework for improved video text detection and recognition", Multimedia Tools and Applications, 2014.
Publication <1%

90 Seeri, Shivanand V., Shantala Giraddi, and B. M. Prashant. "A novel approach for Kannada text extraction", International Conference on Pattern Recognition Informatics and Medical Engineering (PRIME-2012), 2012.
Publication <1%

91 Liang, K.. "A splitting moving mesh method for reaction-diffusion equations of quenching type", Journal of Computational Physics, 20060701
Publication <1%

92 "Robust Seed Localization and Growing with Deep Convolutional Features for Scene Text Detection", Proceedings of the 5th ACM on <1%

International Conference on Multimedia Retrieval - ICMR 15, 2015.
Publication

93 Roy, Sangheeta, Palaiahnakote Shivakumara, Hamid A. Jalab, Rabha W. Ibrahim, Umapada Pal, and Tong Lu. "Fractional poisson enhancement model for text detection and recognition in video frames", Pattern Recognition, 2016. <%1
Publication

EXCLUDE QUOTES	OFF	EXCLUDE MATCHES	OFF
EXCLUDE BIBLIOGRAPHY	OFF		