St. Mary's University

Faculty of Informatics

Department of Computer Science

Customers Identity Card Data Detection and Recognition Using Image Processing

Thesis Submitted to the School of Graduate Studies of St. Mary's University in Partial Fulfilment of the Requirements for the Degree of Master of Science in Computer Science

By

Chala Tamirat Haile

January, 2023

ACCEPTANCE

Customers Identity Card Data Detection and Recognition Using Image Processing

By

Chala Tamirat

Accepted by the Faculty of Informatics, St. Mary's University, in partial fulfillment of the requirements for the degree of Master of Science in Computer Science

Thesis Examination Committee:

____

Internal Examiner

____

External Examiner

____

Dean, Faculty of Informatics

DECLARATION

I, the undersigned, declare that this thesis work is my original work, has not been presented for a degree in this or any other universities, and all sources of materials used for the thesis work have been duly acknowledged.

Chala Tamirat

Signature

Addis Ababa

Ethiopia

This thesis has been submitted for examination with my approval as advisor.

Million Meshesha (Ph.D)

Signature

Addis Ababa

Ethiopia

ACKNOWLEDGEMENTS

First of all, I praise the almighty GOD for making everything possible. Thank you for giving me the strength, courage, and opportunity to complete this research. I would like to thank my advisor Dr. MILLION MESHESHA for his wonderful constructive comments, guidance, and great support for the successful accomplishment of this research. I would also like to thank him for not giving up on me when the research took longer than expected. I would also like to thank my family, friends, and Ethio-Telecom staff for encouraging and pushing me to finish the research.

Abstract

Many business sectors require the information contained in the ID card to perform the registration process. Previously customer data was inputted manually. Therefore, we need a system that processes automatically. Based on that problem, the Image Processing technique can be used as an alternative solution to the manual input. This process starts by extracting information from ID cards. Then, it will be pre-processed to obtain the necessary part of the image. This research follows the experimental research approach in which independent variables are manipulated or introduced, and all other variables are carefully controlled for the experimenter to measure the dependent variable. To conduct an extensive experiment first image data is captured from customers' identity cards and prepared using image pre-processing.

The main objective of this study is to detect and identify Amharic text from customers'' identity cards by applying effective page segmentation that can recognize text and non-text blocks from ID cards. Effective page layout segmentation is performed to detect and identify object information captured from the ID cards to achieve this goal. first image pre-processing techniques skew, and a perspective correction method is implemented to make collected document images ready for processing. Then, binarization methods are used to solve lightning issues. Based on the experiment Sauvola's method worked better and faster. The second process is segmentation. This is done by applying page layout segmentation techniques, morphological dilation, and connected component (CC) to separate graphics from the text area and segment text line areas. For document images containing a small amount of noise, the system's performance without skew correction shows 90.87% precision and 98.40% recall. After the proposed skew and perspective rectification were applied a 93.6 % precision and 100% recall were registered.

This study tried to detect and identify Amharic text from the ID cards of customers. Customer ID cards have different physical and logical layouts such as complicated graphics, logos, pictures, etc. The proposed study adopts google tesseract OCR for Amharic ID card document recognition; However, the recognition accuracy depends on the quality of ID cards. The study focuses on determining and identifying sample attributes. Therefore, to determine the overall layout of every scanned ID card, extracting a sense of the format and content of every scanned ID card needs further research to be conducted.

Key Words: **Amharic text information extraction, Accuracy, Identity card, OCR, Page layout segmentation, Precision, Recall**

List of Acronyms

ETC	Ethiopian telecommunication corporation
E-CAF	Electronic customer Acquisition form
ICR	Intelligent character recognition
ID	Identification
IWR	Intelligent word recognition
LSTM	Long Short-Term Memory
MSER	Maximal Stable Extremal Region
MSE	Mean Square Error
OCR	Optical Character Recognition
SVM	Support Vector Machine
RLSA	Run Length Smoothing Algorithm
PSNR	Peak signal-to-noise ratio
VAS	Value added service

Contents

List of Figures

List of Tables

List of Algorithm

CHAPTER ONE

INTRODUCTION

1.1 Background

Technology development is growing quickly in the last decade and having an impact on the daily activities of humans. Many manual works automated and percentage of human intervention reduced and changed by applying a computer [1]. In terms of visual information, digital image processing plays an increasingly important role in many aspects of our daily life, as well as in a wide variety of disciplines and fields in science and technology, with applications such as television, photography, robotics, remote sensing, medical diagnosis and industrial inspection [2]. It deals with manipulation of digital images through a digital computer. Image processing is the application of signal processing techniques to the domain of Images two-dimensional signals such as photographs or video [3].

Image processing does typically involve filtering or enhancing an image using various types of algorithms in addition to other techniques to extract information from the images. It also means "Analysing and manipulating images with a computer" [3]. Image processing is performed in three steps: First, import images with an optical device like a scanner or a camera or directly through digital processing. Second, manipulate or analyse the images in some way. This step can include image improvement and data summary, or the images are analysed to find rules that are not seen by the human eyes [4]. Lastly, output the result of image processing. The result might be the image changed by some way or it might be a report based on analysis or result of the images.

Digital image processing is a very popular and rapidly growing area of application under computer science and engineering [3]. Its growth leads by technological innovations in the fields of digital imaging, computer processing and mass storage devices. Some of the important applications of image processing in the field of science and technology include computer vision, remote sensing, feature extraction, face detection, forecasting, optical character recognition, fingerprint detection, medical image processing, and morphological imaging [4]. In various business areas, especially those with the registration process, customer data were inputted into database manually, which means the customer fill the form and the staff enters the data written on the paper into computer. One of the most essential data entity used for registration activities is Identity Card [5].

The modern state system of Europe between 1400 and 1600 introduced the process of identifying the citizens in some forms [6]. The basic function of a national identity Card is to link a stream of data with a person [7]. Identification is defined as the act of identifying, the state of being identified or something that identifies one [7].The verb identify is linked to the noun identity, such as in the case of the term “identity card” which can be used to identify someone belonging to a particular group [6].

The ID card information identification depends on image processing, which has the following three Parts: a) ID card image pre-treatment, b) information segmentation, c) character recognition. First of all, there is a need to get the interested parts of ID card image according to priori knowledge. This is followed by line segmentation and column split. Finally, character images are segmented and character recognition is performed to convert into its equivalent textual form. The image segmentation and character recognition are the key parts of the recognition system [8].

1.2 Ethio-telecom

Ethiopian Telecommunications Corporation (ETC) is introduced in 1894, seventeen years after the invention of telephone technology in the world[9]. ETC is government owned sole telecommunication service provider in Ethiopia. The Ethiopian government has decided to transform the telecommunication infrastructure and services to excellent standard bearing in mind them as a key lever to the development of Ethiopia. Thus, Ethio Telecom (ET) is born out of ETC on December 2, 2011, in order to bring about a paradigm shift in the development of the telecom sector to support the steady growth of our country [9].

Currently Ethio Telecom provide telecom service in the whole country on voice, internet, data, channels and value added services (VAS) with comprehensive plans in place to meet the requirements set out by the Ministry of Communications & Information Technology (MCIT) and peoples of Ethiopia [9].

Since the services are given in entire Ethiopia detecting and capturing data from each ID card is a challenging task because of the wide variety of ID cards with many text appearances, due to variations in font, thickness, colour, size and texture. Thus, image analysis technique will be used to extract the words that used in subsequent recognition steps. At first scanned ID card image used as an input into the computer system and then the target text region separated from the picture region. The text region used for separation of lines and words based on the vertical and horizontal projections of image intensity, respectively [10].

1.3 Statement of the Problem

Identification card (ID card) is the main reference in obtaining information of a customers. Many business sectors require the information contained in the ID card to perform the registration process. Data of customers who buy items usually asked to get their identity card for registration. Detail information of customer’s identity can be obtained from their ID Card. Previously in case of Ethio-telecom, customers’ data inputted manually. The customer filling agent handles both paper and soft copies of agreement form records, check the accuracy of each subscriber’s data and scans and uploads the hardcopies of the customer acquisition form, So that it is not efficient process because it needs a lot of time to input data one by one [11].

Ethio-telecom provides nationwide telecommunications services for their customers in Ethiopia. To get this service customer must subscribe or register to their nearest telecom shop. While giving services telecom operators takes a lot of procedure to have adequate amount of information about their customers, requires customers to submit some documents that they could use to verify their identity and get relevant details about them. The customer expected to submit a digital copy of his identity card to identify if it is fake, to get information like name, address and enter it into a database.

So, while registering and adding new customers they are doing it manually. That means customer-filling agent spent hours typing, checking and re-typing information from multilingual

ID cards into data systems, and it is repetitive task. They collect each customer's Identity card then, check and put it on their system when there is any transaction with their subscribers. This process is time consuming and disposed to error.

In Ethiopia, there is no identity card issued at the national level, except for passports. All the regional governments, which speaks different languages such as Afan-Oromo, Amharic, Somali and Tigrigna, have their own identification cards, including Kebele cards. The identity card is used for different purposes, such as for bank account opening, telephone subscribing as well as in many real-life transactions and security checking. Many organizations during their ongoing procedures wants to have adequate amount of information about their customers, requires customers to submit some documents that they could use to verify their identity and get relevant details about them. However, this process is usually time consuming and prone to errors since it is done manually at many places. The customer expected to submit a digital copy of the documents, which a manual reviewer will review, identify if it is fake, extract information like name, address and enter it into a database.

Different research has been done at different parts of the world to suggest a solution in detecting and segmenting useful information from Identity cards. In 2005, Wang et al [12] used Gabor-filters for character recognition with low image quality and for Chinese-readable characters. The experiment shows that the proposed method has excellent performance on both low-quality machine – printed character recognition and cursive handwritten character recognition. In 2011, Vikas et al [13] developed document segmentation using histogram analysis and presented a primary work for segmentation of lines, words and characters of Devnagari script. He nearly achieved 100% successful segmentation in line and word segmentation, but suggested character level segmentation needs more effort, as it is complicated for Devnagari script.

In 2015 Ryan et al [14] conducted research about character recognition on people using Zhange-suen algorithm divided into 2 algorithms: 3x3 algorithm and pixel-by-pixel algorithm. In 2016 Valiente et al [15], using Optical Character Recognition to detect ID card combined with cloud technology. Most of above-mentioned researchers used image-processing technology combining with Machine Learning to detect citizen ID card. The appropriate selection of image processing and machine learning techniques can improve accuracy of prediction.

To the best of the researcher's knowledge, there is no work done on Identity Card Data detection in Ethiopia. However, many Optical Character Recognition (OCR) applications have been widely implemented and used in various multilingual historical and modern documents including books, newspapers, magazines, and cultural and religious archives written with different scripts. Since no research has been done for Ethiopia on customers Identity Card data Detection and recognition it will motivates us and paves a way for other research. Therefore, the purpose of this research is to design and develop automatic Amharic text detection and recognition system from identity card of customers.

To this end, the following research questions are investigated and answered:

- Which image-processing algorithm is more suitable to detect texts from ID card?
- What are the important characteristics of Amharic language to model the automatic extraction of text from ID cards?
- To what extent the proposed prototype works in ID card text detection and recognition?

1.4 Objective of the study

1.4.1 General Objective

The general objective of this study is to design automatic Amharic text detection and recognition system for identity card of customers.

1.4.2 Specific Objectives

To achieve the general objective of the study, the following specific objectives are formulated.

- To review previously proposed related works on ID card detection and recognition so as to identify suitable methods and algorithms.
- To capture image of ID card for training and testing.
- To prepare dataset by applying different image enhancement and restoration techniques.
- To identify texts on ID card and recognize for verification using image processing algorithms.

- To model and design a system that detects and recognizes Customer ID cards taking into consideration different types of ID formats and size.
- To evaluate the performance of the developed system on detection of customers ID Card.

1.5 Methodology of the study

To achieve the above-mentioned objective of the study, the below methods and procedures are used.

1.5.1 Research design

This research follows the experimental research approach to which independent variables are manipulated or introduced, and all other variables are carefully controlled for the experimenter to measure the dependent variable and make conclusions about how the variables are related. The goal of this experimental research conducted in a controlled setting is to construct an optimal model for detection of text from customers' identity card and recognition for simplifying checking the validity of customer ID card[16].

To conduct an extensive experiment first image data is captured from customers' identity card and prepared using image preprocessing. This is followed by image analysis and implementation and finally evaluation of the prototype to measure its performance.

1.5.2 Dataset collection and preparation

Different Ethiopian national ID cards with various font styles, sizes and types are collected based on customer's traffic from ethio telecom like students ID card, kebele ID card, and license and government employees ID card with Amharic. First collection of customers ID card is done, then, uses in the training data. After collecting appropriate data, pre-processing should be performed in order to make image that is used in forward tasks.

Pre-processing is a processing that is performed on raw data to ready it for another processing procedure. Hence, pre-processing is the initial step which transforms the data into a format that is more easily and effectively processed. In order to achieve higher recognition rates, it is important to have an effective pre-processing stage, therefore using effective pre-processing algorithms. Pre-processing techniques are needed on colour, binary document or grey-level images containing text and/or graphics. Following are some of the techniques for character recognition systems[17]:

- Image Enhancement Technique: To remove noise or correct the contrast in the image.
- Thresholding Technique: To remove the background containing any scenes watermarks etc.
- Page Segmentation Technique: To separate graphics from text.
- Character Segmentation: To separate characters from each other.
- Morphological Pre-processing: To enhance the characters in cases.

Image capturing devices used for this research is portable document scanner (model - DY1030) with DPI 280 and 3 Mega pixels.

1.5.3 Implementation Tool

Depending on different studies done on ID card text detection, they have used MATLAB for image pre-processing, segmentation, feature extraction and text recognition. Because it has a lot of built-in methods for image processing, analysis and recognition. Hence, this research uses MATLAB for image segmentation, image feature extraction, developing a prototype and to implement the proposed system for recognition.

1.5.4 Evaluation Procedure

The performance of the proposed approach is tested at several stages. First, the effectiveness of the proposed segmentation technique was measured using direct mapping, which determines the performance by finding the correspondences between detected entities and ground truth. The output of recognition engines often contains number of mistakes such as misspelled words or spurious characters. In order to obtain a measure, which is independent of the text size, the number of mistakes is usually normalized to the length of the expected contents (ground truth text).

Evaluation output of recognition of the model developed in this research is conducted by manually comparing the identified string with the actual one in a tabular format. The table layout allows visualization of the performance of the model.

Degree of Accuracy: -Character accuracy is defined by the number of actual characters with their places divided by the total of actual characters. Areas of concern are missed characters, extra characters, and misplaced characters.

1.6 Significance of the Study

Much research has been done on identity card text recognition around the world. However, in case of Ethiopia where different type of identity cards exists with multilingual languages, it needs a lot of work to improve the manual data capturing trends. Digitizing information has several advantages a business can gain on several grounds. Businesses can track their processes better, can provide better customer service, improve the productivity of their employees and reduce costs.

Some reasons why ethio telecom has to consider to automatically extract data from their customers ID cards are it automates processes. They can automatically take the customer ID card images, extract text from them, extract the values of different fields, and make error corrections and make ready for process. It also increases efficiency, for example once a customer is registered and when he come back to subscribe for other services the salesperson can simply check and compare detail information of customer with corresponding Id card. In addition to it increases customer satisfaction and help them improve their customer service experience as it makes delivery of service faster and smoother. Finally, yet importantly, they have better storage after all post-processing of data, having automated the process of entering their customer data into a database, now they can retrieve information of their customer anytime.

Furthermore, it can be used as an input for future works which might be aimed to develop a full-sized Automatic Amharic Text Detection and Recognition system from identity card of customers.

1.7 Scope and Limitation of the study

The main purpose in conducting this study was to design a system that detects and recognizes data from customer ID cards and that takes into consideration different types of ID layouts and size formatted in Amharic script. The study uses ID cards customer use around ethio-telecom in their daily transaction in Addis Ababa, the font is computer written, and handwritten documents recognition are out of the scope of this study and only focuses on text extraction and recognition. The system looks for the foreground fields in customer ID card such as first name, last name, and sex, and nationality, date of birth, signature and picture (Photo).

This research follows different approaches for segmentation, extraction, and recognition of customer ID cards and also layout analysis is essentially performed for detection of Region of Interest (ROI), such as text, image region. Page layout analysis is to analyse and determine locations, structures of different information areas on the input image such as text and image [18].

CHAPTER TWO
LITERATURE REVIEW

2.1 Customers Identity Card

A customer is a person or company who purchases goods and services [19]. A customer becomes a consumer when he or she uses the goods or services; i.e. where there is some consumption [20]. While customers purchase and doing transaction, some business sectors require and ask the information contained in the ID card to perform the registration process. It enables them to provide these services with better integrity and control. The identification card is merely to verify whom a person claims to be and to specify what things are and what they are not. The specification of similarities and differences is basic to define who a person is, both for himself/herself and for the institution. One of the most important data used in Ethiopia for various business areas and other government activities especially those with the registration process is Identity Card. Currently the country uses a manual ID card system. This study devises an approach for automating manual ID card system.

2.1.1 Identity Cards Issued in Addis Ababa

In Ethiopia, all the regional governments, which use different languages such as Afan Oromo, Amharic, Somali, Tigrigna and English, have their own identification cards, including Kebele cards. Residence identity cards in Ethiopia are issued by kebeles (which is an administrative sub-district)to people 18 years and older but there is "no uniform identification document [21].

However, for the region of Addis Ababa, ID-cards are issued by a woreda (an administrative district) and the word ‘kebele’ has been replaced by ‘woreda’ on identity cards which were issued in the region of Addis Ababa. Previously a person could have such cards issued in the so-called kebeles; however, due to a structural change a few years ago, kebeles were merged into a woreda. Attributes contained on the card includes the individual's full name, woreda, kebele, House Number, mother's full name, telephone number, date of birth, place of birth, sex, ethnicity, occupation, a sealed photo; the name of a person to be called in case of an accident, the date of issue, the name of the person who issued the card, and a signature.

The ID card also contains the following notice on the back of the card if you found this ID card, please return it to the above-mentioned address or to the concerned city administration. The resident should always hold this ID card When the resident vacates[s] from the Kebele he/she should return the ID card. This ID shall be valid for 2 years starting from the issue date [21].

2.2 Overview digital image processing

An image is a way of transferring information, and the image contains lots of useful information. Understanding the image and extracting information from the image to accomplish some works is an important area of application in digital image technology, and the first step in understanding the image is the image segmentation. In practice, it is often not interested in all parts of the image, but only for certain areas, which have the same characteristics [22].

Digital Image Processing means processing digital image by means of a digital computer.we can also say that it is a use of computer algorithms order to get enhanced image either to extract some useful information[23]. The captured image for designing an application must pass through different steps for designing an application, such as image pre-processing, segmentation, feature extraction and identification. The following steps are followed in digital image processing[22].

2.2.1 Image Acquisition

In image processing, image acquisition is an action of retrieving image from an external source for further processing. It's always the foundation step in the workflow since no process is available before obtaining an image.Image acquisition plays a major role in image processing, since if the images are not acquired properly the various image processing techniques may not be much effective, even with the presence of various enhancement techniques[24]. The general aim of Image acquisition is to transform an optical image (Real World Data) into an array of numerical data, which could be later manipulated on a computer, before any image processing can commence an image must be captured by camera and converted into a manageable entity. The image acquisition process consists of three steps: First, optical system which focuses on the energy; second energy reflected on the object of interest and finally a sensor which measure the amount of energy [25].

Image Acquisition is achieved by suitable camera. We can use different cameras for different application. If we need an x-ray image, we use a camera (film) that is sensitive to x-ray. If we want infra-red image, we can use camera which are sensitive to infrared radiation. For normal images (family pictures etc.) a camera is used which is sensitive to visual spectrum. Image Acquisition is the first step in any image processing system as shown below in figure 2-1 [25].

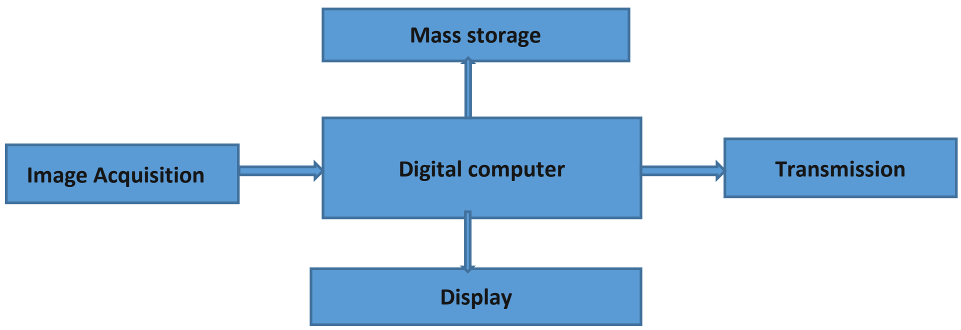


Figure 2-1 image processing [25]

2.2.2 Image Pre-processing

Image pre-processing methods are intended for image improvement for the needs of next processing of the image. The main goal of pre-processing is noise suppression (usually the origin of the noise is digitizing and transmission), removal of distortion given by the scanning device, eventually suppress or highlight other attribute, which are important for following processing, segmentation and edge detection.

The aim of image pre-processing is the improvement of image data by enhancing some features while suppressing some unwanted distortions. Enhancing the features depends on specific applications. Image data recorded by sensors on a satellite, consist of errors related to geometry and brightness values of the pixels. In image pre-processing, these errors are corrected using appropriate mathematical models which are either definite or statistical models. Image pre-processing also includes primitive operations to reduce noise, contrast enhancement, image smoothing and sharpening, and advanced operations such as image segmentation [24].

According to the size of the pixel neighbourhood that is used for the calculation of a new pixel brightness the image pre-processing techniques can be categorized as,

- Pixel brightness transformations.

- Geometric transformations.
- Pre-processing techniques that use a local neighbourhood of the processed pixel.
- Image restoration that requires knowledge about the entire image.

2.2.3 Image Enhancement

Image enhancement (IE) is one of the challenging issues in image processing. The objective of Image enhancement is to process an image so that result is more suitable than original image for specific application. Digital image enhancement techniques provide a lot of choices for improving the visual quality of images. The images captured from some conventional digital cameras may lack in contrast and Brightness, because of the limitations of imaging subsystems and illumination conditions while capturing an image. Here it enhances some features that are concealed or highlight certain features of interest, for subsequent analysis of an image [26].

Some of the areas in which IE has wide application are noted below [26]:
1. In atmospheric sciences, image enhancement is used to reduce the effects of haze, fog, and turbulent weather for meteorological observations. Image enhancement helps in detecting shape and structure of remote objects in environment sensing. Satellite images undergo image restoration and enhancement to remove noise.
2. In forensics, Image enhancement is used for identification, evidence gathering and surveillance. Images obtained from fingerprint detection, security videos analysis and crime scene investigations are enhanced to use in identification of culprits and protection of victims.
3. Astrophotography faces challenges due to light and noise pollution that can be minimized by IE. For real time sharpening and contrast enhancement several cameras have in-built IE functions. However, numerous software, allow editing such images to provide better results.

We can categorize the image enhancement techniques into two. They are spatial domain method and frequency domain method. Spatial domain method deals with the modification or aggregation of pixels that forms the image and frequency domain method enhances the image in a linear manner by positioning the invariant operator

Image enhancement techniques

Some of the widely used techniques for image enhancement are discussed as follows.

A. Histogram equalization

Histogram equalization is a very common technique for enhancing the images. Suppose we have an image which is predominantly dark. Then its histogram would be skewed towards the lower end of the grey scale and all the image detail is compressed into the dark end of the histogram. If it could `stretch out' the grey levels at the dark end to produce a more uniformly distributed histogram then the image would become much clearer. Histogram equalization stretches the histogram across the entire spectrum of pixels (0 – 255). It increases the contrast of images for the finality of human inspection and can be applied to normalize illumination variations in image understanding problems [26].

B. Local Enhancement

Previous methods of histogram equalizations and histogram matching are global. So, local enhancement is used. Define square or rectangular neighbourhood (mask) and move the centre from pixel to pixel. For each neighbourhood, calculate histogram of the points n the neighbourhood obtain histogram equalization /specification function. Map grey level of pixel centre in neighbourhood. It can use new pixel values and previous histogram to calculate next histogram[26].

C. Contrast Stretching

To expand the range of brightness values in an image the contrast enhancement techniques are used, so that the image can be efficiently displayed in a manner desired by the analyst. The level of contrast in an image may vary due to poor illumination or improper setting in the acquisition sensor device. Therefore, there is a need to manipulate the contrast of an image in order to compensate for difficulties in image acquisition[26].The idea behind contrast stretching is to increase the dynamic range of the grey levels in the image being processed. The idea is to modify the dynamic range of the grey-levels in the images.

D. Thresholding Transformations

Thresholding transformations are particularly useful for segmentation in which we want to isolate an object of interest from a background. Image threshold is the process of separating the information (objects) of an image from its background, hence, thresholding is usually applied to grey-level or colour document scanned images. Thresholding can be categorized into two main categories: global and local. Global thresholding methods choose one threshold value for the entire document image, which is often based on the estimation of the background level from the intensity histogram of the image; hence, it is considered a point processing operation. Global thresholding methods are used to automatically reduce a grey-level image to a binary image.

2.2.4 Image Restoration

The process of recovering degraded or corrupted image by removing the noise or blur, to improve the Appearance of the image is called image restoration. The degraded image is the convolution of the original image, degraded Function, and additive noise. Restoration of the image is done with the help of prior knowledge of the noise or the disturbance that causes the degradation in the image.

It can be done in two domains: spatial domain and frequency domain. In spatial domain the filtering action for restoring the images is done directly on the operating pixels of the digital image and in frequency domain the filtering action is done by mapping the spatial domain into the frequency domain, by Fourier transform. After the filtering, the image is remapped by inverse Fourier transform into spatial domain, to obtain the restored image[24].

2.2.5 Binarization

Image binarization is an important process for document image analysis. The inherently bi-level nature of text documents has led to many of the document analysis algorithms being designed for use on bi-level images. If the image binarization is improperly done, then the follow-on steps cannot proceed appropriately[27].

Many studies of binarization have been completed. These surveys focus on the binarization algorithm. The algorithms have evolved from global thresholding to local adaptive thresholding to allow for variations in the image background and today range from relatively simple algorithms, to some that are rather complex. Some algorithms are a hybrid of local and global by using background shading estimation. The choice of 'best' binarization algorithm usually depends on the other constraints of the problem. The images can have stains or bleed through, or a uniformly noisy background. The text can suffer additional degradations. Algorithms that work best on one image may not be best for another.

The binarization algorithms that were evaluated include Otsu, Niblack, Sauvola, Gatos and a background estimation and subtraction algorithm the authors of this paper submitted to the DIBCO 2009 contest.5 Descriptions of these binarization algorithms follow [27]. **Otsu** is an often used global thresholding method. It is based on treating the gray level intensities presenting the image as values to be clustered into two sets, one foreground (black) and one background (white). To accomplish this the algorithm minimizes the weighted sum of within-class variances of the foreground and background pixels to establish an optimum threshold. This is equivalent to maximizing the between-class scatter. From this a scalar number, *K*, is returned. This is then used to binarize the image through.

$$I_{bin}(x,y) = \begin{cases} 1, & \text{if } I_{gray}(x,y) \leq K \\ 0, & \text{if } I_{gray}(x,y) > K \end{cases} . \tag{2.1}$$

Niblack is a local adaptive thresholding algorithm. The threshold for each pixel is determined by examining the average of the pixels in a neighborhood, $m(x, y)$, and the standard deviation, $\sigma(x, y)$, in that same neighborhood. The threshold for Niblack is then chosen as[25]

$$T(x,y) = m(x,y) + k \cdot \sigma(x,y). \tag{2.2}$$

Instead of a global threshold *K* as in Equation 2.1 each pixel is subjected to the threshold process separately. The most common value for the constant *k* is -0.2, which is what was used in this paper. While Niblack is one of the more widely cited local adaptive binarization algorithms, in low contrast regions it is prone to producing ghosting speckle. Variations on it are therefore often implemented, or a post-processing filter is applied. The implementation in the Gamera toolkit looks at the absolute intensity of the pixel in question and has an upper and lower bound beyond which the adaptive evaluation is not used. This prevents much of the ghosting speckle. That variation was implemented for these tests with bounds at 20 and 150[27].

Sauvola is a similar local adaptive thresholding algorithm. It has fewer of the side effects associated with Niblack, and thus was also considered in this study. The threshold for Sauvola is also determined by combining the local average and standard deviation of the pixels,

$$T(x,y) = m(x,y) + \left\{1 + k\left(\frac{\sigma(x,y)}{R} - 1\right)\right\}. \tag{2.3}$$

Values of $k = 0.5$ and $R = 128$ are used in this study.

Gatos et al.3 developed a binarization algorithm that is particularly designed to work on documents with uneven backgrounds resulting from bleed through and stains. The algorithm has four main parts. The first is application of a Wiener filter. The second step is to apply the Sauvola threshold to get a rough estimate of foreground and background pixels. Next an estimate of the background is made for the pixels determined in step 2 to belong to the foreground $S(x, y)$ by

$$B(x,y) = \frac{\sum_{i_x=x-dx}^{x+dx}\sum_{i_y=y-dy}^{y+dy}(I(i_x,i_y)(1-S(i_x,i_y)))}{\sum_{i_x=x-dx}^{x+dx}\sum_{i_y=y-dy}^{y+dy}(1-S(i_x,i_y))}. \tag{2.4}$$

The final thresholding is accomplished by comparing the difference from the background image, $B(x, y)$, and the preprocessed gray level image, $I\,filt(x, y)$, through

$$B(x,y) - I_{filt}(x,y) > d(B(x,y)). \tag{2.5}$$

The threshold d is a function of the image background and takes on variable values that are smaller in darker regions. This is achieved through the function

$$d(B(x,y)) = q\delta\left(\frac{(1-p_2)}{1+exp(\frac{-4B(x,y)}{b(1-p_1)} + \frac{2(1+p_1)}{(1-p_1)})} + p_2\right). \tag{2.6}$$

The parameter b is the average background surface value B(x, y) over the text areas. Parameter of q = 0.6,p1 = 0.5, and p2 = 0.8 are suggested[27].

2.2.6 Normalization

Normalization is a pre-processing stage of any type problem statement. Especially normalization takes important role in the field of soft computing, cloud computing etc. for manipulation of data like scale down or scale up the range of data before it becomes used for further stage. There are so many normalization techniques namely Min-Max normalization, Z-score normalization and Decimal scaling normalization[28].

Normalization is scaling technique or a mapping technique or a pre-processing stage. Where, we can find new range from an existing one range. It can be helpful for the prediction or forecasting purpose a lot .as we know there are so many ways to predict or forecast but all can vary with each other a lot. So to maintain the large variation of prediction and forecasting the Normalization technique is required to make them closer [28].

Image normalization techniques have been used as a preprocessing step in many applications such as pattern recognition and classification image retrieval, and image water- marking Image normalization is used in such applications with the purpose of representing objects, patterns (or the entire image) regardless of changes in their orientation, size, or position. The use of image normalization as a preprocessing step might limit the range of variations of images/patterns, as it effectively decouples the problem of image deformations from the main task of the application, for example, retrieval, recognition or classification, etc [29].

2.2.7 Skew Detection

Skew detection and correction of documents is a problematic step in document image analysis. Many methods have been proposed by researchers for estimating the angle at which a document image is rotated (document skew) in binary image documents. Some researchers tried to evaluate the most frequently skew detection techniques cited in the literature which are (i) Projection Profile Analysis (PP), (ii) Hough Transform (HT) and (iii) Nearest Neighbor (NN). This study points out the weaknesses and the strengths of each method and compares the performance of these methods in term of speed and accuracy.

The evaluation result shows that in term of speed, the NN technique achieves the fastest time. However, NN performs poorly for the accuracy estimation. PP gives the best angle estimation even though it takes the longest time to execute. Hence, this finding can be used as the basis evaluation review for image analysis researchers in improving the existing technique of skew detection and recommend algorithm with a better accuracy in a shorter time.

Massive information has been stored in documents, for a better storage and intelligent processing is important to convert the paper-based documents into an electronic format. One of the problems in this field is that the document may be rotated on a flatbed scanner at an arbitrary angle. This means that the document images are skewed, which led to degrade the image quality and causes many problems in analysis the image, extract the content and also reduce the performance of Optical Character Recognition[30].

Skew detection is an important step in the document image analysis and understanding for optical character recognition system, many researchers proposed an algorithm to estimate the skew angle which is an orientation angle from the horizontal or vertical direction. After that the text rotated into opposite direction to remove the skew (must be a zero degree). The skew in the document images can be classified into three different types:

A. Global skew: assume that the document have common degree angle orientation, this occurs because the process of scanned

B. Multiple skew: the document have different skew angle

C. Non-uniform text line skew: when single line on the document has different angles, line took a wavy shape.

Several methods have been developed by researchers for skew angle detection as described below:

A. Projection Profile Analysis: It is the one of the popular skew estimation technique which was initially proposed by Postl [30].Projection is a Process that converts a binary image into one-dimensional array (projection profile). Horizontal projection profile Represented by lines is equal to number of line on document, where each line in projection profile has a value that represents a number of black pixels in the corresponding row of the image. For the documents with skewed angle zero, the horizontal projection profile has valleys that commensurate with the space between the lines and the height of maximum peak equal to heights text lines in document images. Therefore, Postl proposed a method that calculates the variation in projection profile at a number of different angles the skew Angle is equal to angle that have the most variation[30].

B. Hough Transform: Hough Transform is known widely technique in computer vision and image analysis. It is a feature extraction technique that used to isolate features of a particular shape like line or curve within digital images by specifying the desired features for the shape in parametric form. The Hough transform was introduced first by Richard Duda and Peter Hart in 1972 as a simple linear transform to detect a straight line. In image space, any straight line can be described in a slope-intercept model as in equation (1).

$$y = mx + b \tag{2.7}$$

Where *m* parameter is the slope of the line, and *b* is the intercept (*y*-intercept). The main idea in Hough transform is that to represent the characteristics of straight lines in terms of its parameters according to the slope-intercept model not as discrete image points (x1, y1), (x2, y2), etc.. Therefore, the straight line $y = mx + b$ can be represented as a point (b, m) in the parameter space[30].

C. Nearest Neighbor: Nearest Neighbor method is based on finding the connected components of document, then finds the histogram of the direction vectors for all nearest neighbors for all components and computes the first nearest neighbor of each component. The angle between centroids of nearest neighbor components is founded and accumulated in the histogram. To find document skewed angle, the dominant peak is founded.

2.2.8 Underline Detection

Detecting text in constrained, controlled environments can typically be accomplished by using heuristic-based approaches, like exploiting gradient information or the fact that text is typically grouped into paragraphs and characters appear on a straight line. The text detection methods are generally connected components or texture based. The connected components method segments the character's into smaller components. Arrange connected components to text/background by using their geometrical properties. The connected components based method is difficult to run when the text contain noise. The connected components method is most efficient for extracting the text from camera captured images. Texture based images are used here to extract textual properties of the images. Neural networks can be used to differentiate text from the document images.

Give an image Maximal Stable Extremal Region (MSER) tries to find connected (and nearly connected) regions. Here need to take the parameters such as the illumination, orientation, size etc. In grey scale images, the local variations are analysed to locate the regions[31]. The combination of texture based method and Connected Components extract the wavelets features of image and then classify using neural network. The texture-based methods perform well in noisy image. The process is usually slow.

Text detection procedures are categories into two [31]: connected comment and region based detections.

1) *Connected Components (CC):* Using Maximally Stable Extremal Region Algorithm (MSER) the images are extracted. The extracted components are then classified by using SVM/Ada Boost classifier.

2) *Region Based (RB):* Consider text has different texture features compared to the background. In pixel values the neighboring gray values are considered as features. They are classified by using SVM to differentiate the text/non-text content.

Text detection process is the major step in text line detection task. It is used in Optical Character Recognition (OCR) and preprocessing algorithms. Text line detection is difficult in curved text and blurred text. The English language text in the images is captured. The main text detection challenges are divided into three groups [29]:

1) Variety of natural images: The characters in natural images are in different font style, different size, unique color, and unique font alignment.

2) Problems in backgrounds: The background includes grasses bricks, rocks, are leads to more complexity for identifying text.

3) Interference factors: The main inference factors are blurring.

2.2.9 Segmentation

Segmentation is an operation that seeks to decompose an image into sub-images of individual symbols [32]. Reliable character segmentation and recognition depend up on both document quality, registered image quality, and is one of the hotspots in image processing and computer vision. It is also an important basis for image recognition. It is based on certain criteria to divide an input image into a number of the same nature of the category in order to extract region of interest as per design requirement of a given application. Moreover, it is the basis for image analysis and understanding of image feature extraction and recognition [33].

There are many commonly used image segmentation algorithms. They are divided into region based segmentation, edge based segmentation and segmentation based on clustering.

2.2.9.1 Region-based Segmentation

Region-based segmentation comprises separating an image into regions with similar features. Each region is a group of pixels, which the algorithm locates via a seed point. A region in an image can be defined by its border (edge) or its interior, and the two representations are equal.

The most commonly used region-based segmentation algorithms are discussed as follows.

A. Threshold Segmentation

Threshold segmentation is the simplest method of image segmentation. It is the most common segmentation algorithm, which directly divides the image based on the grey value of different targets. Threshold segmentation can be divided into local thresholding method and global thresholding method. The global threshold method divides the image into two regions of the target and the background by a single threshold. The local threshold method needs to select multiple segmentation thresholds and divides the image into multiple target regions and backgrounds by multiple thresholds. The most commonly used threshold segmentation algorithm is the largest interclass variance method, which selects a globally optimal threshold by maximizing the variance between classes [34].

The size of an image is M*N (the total number of pixels is M*N), and a threshold is set to divide the image into two parts: background and target. The average gray value of the image is. N1 is the number of pixels in the background and N2 is the number of pixels targeted. Background pixel percentage w1=N1/ (M*N), and the target pixel ratio w2=N2/ (M*N).
The background average gray value:

$$\mu 1 = \mu / w1 \tag{2.8}$$

The average grey value of the target:

$$\mu 2 = \mu / w2 \tag{2.9}$$

Inter-class variance of background and target:

$$g = w1 * (\mu - \mu 1)^2 + w2 * (\mu - \mu 2)^2 \tag{2.10}$$

In addition to this, there are entropy-based threshold segmentation method, minimum error method, and co-occurrence matrix method, moment preserving method, simple statistical method, probability relaxation method, fuzzy set method and threshold methods combined with other methods[34]. The advantage of the threshold method is that the calculation is simple, and the operation speed is faster. When the target and the background have high contrast, the segmentation effect can be obtained. The disadvantage is that it is difficult to obtain accurate results for image segmentation problems where there is no significant grey scale difference or a large overlap of the grey scale values in the image. Since it only considers the grey information of the image without considering the spatial information of the image, it is sensitive to noise and grayscale unevenness, leading it often combined with other methods.

B. Region Growing Segmentation

The region growing method is a typical serial region segmentation algorithm, and its basic idea is to have similar properties of the pixels together to form a region. The method requires first selecting a seed pixel, and then merging the similar pixels around the seed pixel into the region where the seed pixel is located. Figure 2.2 shows an example of a known seed point for region growing. Figure 2.2 (a) shows the need to split the image.

There are known two seed pixels (marked as grey squares) which are prepared for regional growth. The criterion used here is that if the absolute value of the grey value difference between the pixel and the seed pixel is less than a certain threshold T, the pixel is included in the region where the seed pixel is located. Figure 2-2 (b) shows the regional growth results at T = 3, and the whole plot is well divided into two regions. Figure 2-2 (c) shows the results of the region growth at T = 6 and the whole plot is in one area. Thus, the choice of threshold is very important.

$$\underset{\text{(a)}}{\begin{bmatrix} 1 & 0 & 4 & 7 & 5 \\ 1 & 0 & 5 & 7 & 7 \\ 0 & 1 & 5 & 5 & 5 \\ 2 & 0 & 5 & 6 & 5 \\ 2 & 2 & 5 & 6 & 4 \end{bmatrix}} \qquad \underset{\text{(b)}}{\begin{bmatrix} 1 & 1 & 5 & 5 & 5 \\ 1 & 1 & 5 & 5 & 5 \\ 1 & 1 & 5 & 5 & 5 \\ 1 & 1 & 5 & 5 & 5 \\ 1 & 1 & 5 & 5 & 5 \end{bmatrix}} \qquad \underset{\text{(c)}}{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}}$$

Figure 2-2 An example of region growing segmentation [34]

Given the sensitivity of region-growing segmentation algorithms to user-supplied parameters, this letter proposes an objective function for measurement of the quality of the resulting segmentation. The function aims at maximizing intrasegment homogeneity and intersegment heterogeneity. It has two components: a measure of intrasegment homogeneity and one of intersegment heterogeneity. For Intrasegment variance of the regions produced by a segmentation algorithm. It is calculated by the formula[32]

$$v = \frac{\sum_{i=1}^{n} a_i \cdot v_i}{\sum_{i=1}^{n} a_i} \qquad (2.9)$$

In the above equation, *vi* is the variance and *ai* is the area of region *i*. The intrasegment variance v is a weighted average, where the weights are the areas of each region. This approach puts more weight on the larger regions, avoiding possible instabilities caused by smaller regions.

The advantage of regional growth is that it usually separates the connected regions with the same characteristics and provides good boundary information and segmentation results. The idea of regional growth is simple and requires only a few seed points to complete. And the growth criteria in the growing process can be freely specified. Finally, it can pick multiple criteria at the same time. The disadvantage is that the computational cost is large. Also, the noise and grayscale unevenness can lead to voids and over-division. The last is the shadow effect on the image is often not very good.

2.2.9.2 Edge Detection Segmentation

The edge of the object is in the form of discontinuous local features of the image, that is, the most significant part of the image changes in local brightness, such as grey value of the mutation, colour mutation, texture changes and so on [34]. The use of discontinuities to detect the edge, to achieve the purpose of image segmentation. There is always a grey edge between two adjacent regions with different grey values in the image, and there is a case where the grey value is not continuous. This discontinuity can often be detected using derivative operations, and derivatives can be calculated using differential operators.

Parallel edge detection is often done by means of a spatial domain differential operator to perform image segmentation by convoluting its template and image. Parallel edge detection is generally used as a method of image pre-processing. The proposed method uses two levels of parallelism for performing the convolution and thresholding operations. If the input image is a colour image, then it is converted to grey as the proposed algorithm handles only grey images. Let size of the image and mask be (m x n) and (k x k) respectively. The image is divided into (m – k +1) * (n – k +1) sub windows or overlapping blocks.

$$cp = \sum_{i=1}^{k*k} i * K \qquad (2.11)$$

Where cp is the impulse response at centre pixel; $_{Ki}$ is the mask value in i^{th} location; Ii is Intensity value of sub window at i^{th} location. The widely used first-order differential operators are Prewitt operator, Roberts operator and Sobel operator [34]. The second-order differential operator has nonlinear operators such as Laplacian, Kirsch operator and Wallis operator.

A. Sobel Operator

The Sobel operator is mainly used for edge detection, and it is technically a discrete differential operator used to calculate the approximation of the gradient of the image luminance function [34]. The Sobel operator is a typical edge detection operator based on the first derivative. As a result of the operator in the introduction of a similar local average operation, so the noise has a smooth effect, and can effectively eliminate the impact of noise.

The influence of the Sobel operator on the position of the pixel is weighted, which is better than the Prewitt operator and the Roberts operator. The Sobel operator consists of two sets of 3x3 matrices, which are transverse and longitudinal templates, and are plotted with the image plane, respectively, to obtain the difference between the horizontal and the longitudinal difference. In actual use, the following two templates are used to detect the edges of the image.

$$Gx = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \qquad Gy = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

(a) Horizonatal egde (b)Vertical edge

Figure 2-3 Detecting edge; [34]**.**

The horizontal and vertical gradient approximations of each pixel of the image can be combined to calculate the size of the gradient using the following formula [34]; where G is

$$G = \sqrt[2]{G_x^2 + G_y^2} \tag{2.11}$$

The gradient can then be calculated using the following formula:

$$\Theta = \arctan(\frac{G_y}{G_x}) \tag{2.12}$$

2.2.9.3 Segmentation based on clustering

Clustering is in accordance with certain requirements and laws of the classification of things in the process. The feature space clustering method is used to segment the pixels in the image space with the corresponding feature space points. According to their aggregation in the feature space, the feature space is segmented, and then they are mapped back to the original image space to get the segmentation result. K-means is one of the most used clustering algorithms. The basic idea of K-means is to gather the samples into different clusters according to the distance. The closer the two points are, the closer they are to get the compact and independent clusters as clustering targets[34].

The algorithm for K-means clustering based segmentation is presented as follows [34].

Step 1: Randomly select K initial clustering centres.
Step 2: Calculate the distance from each sample to each cluster centre and return each sample to the nearest clustering centre.
Step 3: For each cluster, with the mean of all samples as the cluster of new clustering centres.
Step 4: Repeat steps (2) to (3) until the cluster centre no longer changes or reaches the set number of iterations.

The advantage of K-Means clustering algorithm is that the algorithm is fast and simple, and it is highly efficient and scalable for large data sets. In addition, its time complexity is close to linear, and suitable for mining large-scale data sets. The disadvantage of K-means is that its clustering number K has no explicit selection criteria and is difficult to estimate. Secondly, it can be seen from the K-means algorithm that every-iteration of the algorithm traverses all the samples, so the time of the algorithm is very expensive. Finally, the K-means algorithm is a distance-based partitioning method. It is only applicable to the data set, which is convex and not suitable for clustering non-convex clusters [34].

2.2.10 Text Detection and Recognition

2.2.10.1 Text Detection

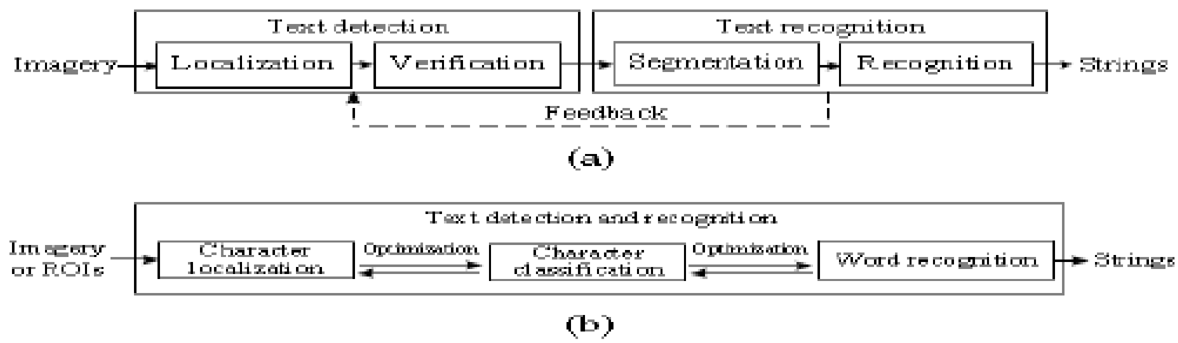
The process of detecting and recognizing text is divided into text detection stage and recognition stage. Text detection deals with finding text area from input image, whereas recognition deals with converting obtained text into computer understandable characters and words. Methods used for this purpose are categorized as stepwise methods and integrated methods. Stepwise methods have separate stages of detection and recognition, and they proceed through detection, classification, segmentation and recognition. Integrated methods have information sharing amongst detection and recognition stages and these methods aim at recognizing words from text available [35]

A. Stepwise methods

Stepwise methods follow stages of text detection and localization, classification, segmentation and recognition and remove background part from text in each stage. Since they are independent of lexicon size, they can be used to recognize text from images independent of volume of text. Elagouni et al. [35] used stepwise approach for text recognition in videos using convolutional neural networks based classifier. Text detection, tracking, character segmentation, recognition and correction are important processing steps used in this approach. Neural networks based horizontal text detection is performed, followed by statistical intensity based shortest path algorithm for character segmentation.

Convolutional neural classification is used for recognition along with language model. Also follows stepwise methods to detect texts of arbitrary orientation from natural images, such as component extraction, component analysis, and candidate linking and chain analysis. Connected components are extracted in component extraction stage using edge detection followed by SWT. Component analysis stage eliminates non-text candidates and remaining text candidates are paired based on their adjacency in the candidate linking stage. Chain level classifier analyses chains formed at this stage in chain analysis stage [35]

B. Integrated methods



Integrated methods focus on detecting particular words from images. Integrated methods often avoid segmentation stage or replace it with word recognition or matching stage. These methods

Figure 2-4 Text image detection and recognition, (a) separately, (b) together

are used for applications with small size of lexicon recognizing fixed set of words. Wang and Belongie [36] used integrated method for word spotting from natural scenes in. It uses character recognition and word configuration stages in this system. It crops region around the text from an image and uses it as input image along with lexicon for word recognition.

It uses Histogram of Oriented Gradients (HOG) features with nearest neighbour classifier for character detection [36]. Word recognition represents each word in lexicon in the form of chain of connected characters and matches it with output of character recognition stage to obtain nearest word for set of characters. It performs end-to-end text scene localization and recognition by keeping multiple segmentations of single characters until last stage of processing where character contexts in text line are known.

This system detects character as external regions, i.e., regions whose outer boundary pixels have higher values than the region. System uses threshold, adjacency and colour space projections as three parameters for every detected character and stores their multiple segmentations from which optimal values are selected based on contexts of characters in text line. Any single parameter does not guarantee efficient results, which causes proposal of storing multiple segmentations of three parameters. For sequence selection process, text regions are divided into text lines using exhaustive search method, followed by rejection of low confidence regions in text lines and construction of directed graph by assigning scores to each node and edge from which correct sequence of characters is selected [35].

2.2.10.2 Text Recognition

Text recognition stage converts images of text into string of characters or words. It is important to convert images of text into words, as word is an elementary entity used by human for his visual recognition. Different approaches of recognition are character recognition and word recognition. Character recognition methods divide text image into multiple cut-outs of single characters. Separation between adjacent characters is very important for these methods.

Character recognition approach using Optical Character Recognition module (OCR) is used in [where initially images are segmented into k classes followed by binary text image hypothesis generation which passes through connected components analysis and grey scale consistency constraint module before getting fed to OCR. Support Vector Machine (SVM) based classifier is used for character recognition in. SVM gives good support for multi-class classification, which is tested in on Indic language Kannada which has total 578 characters formed by altering 34 base consonants using 16 vowels. Many of these characters falls in similar classes which has made use of SVM based layered classification approach easier.

Word recognition uses character recognition outputs along with language models or lexicons to recognize words from text image. It can be used in case of degraded characters. For applications with limited number of word possibilities in input images, word recognition is better approach than character recognition. Wachenfeld, et al. used graphical approach for word recognition. During segmentation stage, when each character is segmented, a hypothesis graph is formed to represent every segmentation and each path of graph is an ordered Segmentation sequence which leads to formation of words from characters[35].

Table 2.1 compares approaches and features used for text detection during image processing stages.

Processing Stage	Approach for processing	Features
Text Detection and Localization	Connected components Analysis	• Graph based method • High speed • Uses color or edge Features • Not efficient for noisy images
	Region Based Methods	• Window-based approach • Less speed • Use texture Features or morphological operations • Efficient for noisy images also
	Supervised Approach	• Supervised classifier has a training phase • Classifier knows features of the text before classification starts

Table 2-1: Comparison of methods used for Text Detection[35]

In the same way, Table 2.2 presents a comparison of approaches and features employed for text image segmentation and recognition.

Classification		• Features like color, size and projection profile are used
	Unsupervised approach	• Unsupervised classifiers do not have a training phase • Classifier learns from Features extracted in the previous classification • Wavelet, stroke width, contrast etc. are used as features
Segmentation	**Binarization**	• Converts color or gray-scale image into black and white image • uses simple or adaptive thresholding simple algorithm • Not suitable for connected characters or degraded text
	Character segmentation	• Converts text into multiple sets of single characters • Uses properties of characters for segmentation • Complex algorithm • Suitable for degraded text as well as connected characters
Text Recognition	**Character Recognition**	• Divide the text into cut-outs of single characters • Independent of a lexicon • Used when a number of words to be recognized are not limited
	Word Recognition	• Identifies words from text images • Recognizes a small number of words provided by lexicon • Suitable only for recognizing a limited number of words

Table 2-2: Comparison of approaches used for processing stages of text recognition[35]

2.3 OPTICAL CHARACTER RECOGNITION

Optical character recognition, usually abbreviated to OCR, is the mechanical or electronic translation of scanned images of handwritten, typewritten, or printed text into machine-encoded text. Imagine we have got a paper document - for example, magazine article or PDF contract your partner sent to you by email. Obviously, a scanner is not enough to make this information available for editing, say in Microsoft Word. All a scanner can do is create an image or a snapshot of the document that is nothing more than a collection of black and white or colour dots, known as a raster image[19].

In order to extract and repurpose data from scanned documents, camera images or image-only PDFs, you need OCR software that would single out letters on the image, put them into words and then - words into sentences, thus enabling you to access and edit the content of the original document. First, the program analyses the structure of document image. It divides the page into elements such as blocks of texts, tables, images, etc. The lines are divided into words and then - into characters. Once the characters have been singled out, the program compares them with a set of pattern images.

It advances numerous hypotheses about what this character is. Basing on these hypotheses the program analyses different variants of breaking of lines into words and words into characters. After processing huge number of such probabilistic hypotheses, the program finally takes the decision, presenting user the recognized text**.**

2.3.1 Methods of Optical Character Recognition

The main methods of the character recognition can be divided into the following groups by the used algorithm: Pattern systems, Structural systems, Feature systems and Neuronal network systems. Each of the mentioned systems has both advantages and disadvantages, which are namely the following: Pattern algorithms are stable to small defects of the image and have sufficiently high recognition velocity. However even minor distortions of the image, which lead to the characters distortion, may influence negatively on the result of recognition and Structural algorithms are very sensitive to the image defects. Besides, in contrast to the pattern and feature systems, effective automated learn procedures for structural systems are not implemented yet.

Feature systems loose important information while calculating the character features and consequently make errors on objects classification referring them to the wrong classes. Although neuronal networks can recognize different fonts taking into consideration their defects and distortions, nevertheless they require complicated multi-layer structure and need a long training using sets of samples. This is not always practicable in industrial environment and at the same time, the economic forces are of great importance here[19].

Types of recognition

- Optical character recognition (OCR) – targets typewritten text, one glyph or character at a time.
- Optical word recognition – targets typewritten text, one word at a time (for languages that use a space as a word divider). (Usually just called "OCR")
- Intelligent character recognition (ICR) – also targets handwritten print script or cursive text one glyph or character at a time, usually involving machine learning.
- Intelligent word recognition (IWR) – also targets handwritten print script or cursive text, one word at a time. This is especially useful for languages where glyphs are not separated in cursive script.

OCR is generally an "offline" process, which analyzes a static document. Handwriting movement analysis can be used as input to handwritten recognition. Instead of merely using the shapes of glyphs and words, this technique can capture motions, such as the order in which segments are drawn, the direction, and the pattern of putting the pen down and lifting it. This additional information can make the end-to-end process more accurate. This technology is also known as "on-line character recognition", "dynamic character recognition", "real-time character recognition", and "intelligent character recognition".

2.3.2 OCR Algorithm

There are two basic types of core OCR algorithms, which may produce a ranked list of candidate characters [19]

2.3.2.1. Matrix matching

Matrix matching involves comparing an image to a stored glyph on a pixel-by-pixel basis; it is also known as "pattern matching" or "pattern recognition" [37].This relies on the input glyph being correctly isolated from the rest of the image, and on the stored glyph being in a similar font and at the same scale. This technique works best with typewritten text and does not work well when new fonts are encountered. This is the technique used in the early physical photocell-based OCR implemented, rather directly.

2.3.2.2. Feature extraction

Feature extraction decomposes glyphs into "features" like lines, closed loops, line direction, and line intersections[38].These are compared with an abstract vector-like representation of a character, which might reduce to one or more glyph prototypes. General techniques of feature detection in computer vision are applicable to this type of OCR, which is commonly seen in "intelligent" handwriting recognition and indeed most modern OCR software. Classifiers such as the k-nearest neighbour’s algorithm are used to compare image features with stored glyph features and choose the nearest match. Software such as Cuneiform and Tesseract use a two-pass approach to character recognition [39].

The first pass is recognition; an attempt is made to recognize each word in turn. Each word that is satisfactory is passed to an adaptive classifier as training data. The second pass is known as "adaptive recognition" and uses the letter shapes recognized with high confidence on the first pass to better recognize the remaining letters on the second pass. This is advantageous for unusual fonts or low-quality scans where the font is distorted (e.g. blurred or faded) [39].

2.4 Amharic Writing System

2.4.1 The Ethiopic Script

At about the beginning of South-Arabian Semitic migrated from Habashat on the Arabian coast across to Africa and founded a kingdom there with its capital at Axum[40]. The immigrants called themselves Ge’ez, which means the ‘emigrants’. Besides, according to Feren   the name

Ge’ez comes from the South Arabian people, the Agazyān, who crossed the Red Sea and settled in North Ethiopia and along the Red Sea coast. The Ge'ez or Ethiopic script possibly developed from the Sabaean/Minean script [40]. The earliest known inscriptions in the Ge'ez script date to the 5th century BC. At first, the script represented only consonants. Vowel indication started to appear in 4th century AD during the reign of king Ezana, though might have developed at the earlier date.

Notable features of Ge'ez writing system are the following [41].

- Type of writing system: abugida
- Writing direction left to right in horizontal lines.
- Each symbol represents a syllable consisting of a consonant plus a vowel. The basic signs are modified in a number of different ways to indicate the various vowels.
- There is no standard way of transliterating the Ge'ez script into the Latin alphabet.

Ge'ez, the classical language of Ethiopia is still used as a liturgical language by Ethiopian Christians and the Beta Israel Jewish community of Ethiopia. Amharic is the national working language of Ethiopia, has about 27 million speakers [41]. It is spoken mainly in North Central Ethiopia. There are also Amharic speakers in a few other countries, particularly in Egypt, Israel and Sweden.

2.4.2 The Amharic Writing

Amharic is written using a writing system called fidel - ፊደል ("alphabet", "letter", or "character") adapted from Ge'ez, the liturgical language of the Ethiopian Orthodox Church. In Amharic writing system, there are Amharic characters, numeric and punctuation marks. Computer using Unicode represented Amharic characters. Unicode provides a unique number for every character, no matter what the platform, no matter what the program, no matter what the language. Ethiopic characters (fidel - ፊደል) have more than 380 Unicode representations including punctuations and special characters (U+1200- U+137F) [42].

There are 33 basic Amharic characters, each of which has seven forms called orders depending on which vowel is to be pronounced in the syllable. The seven orders were representing seven vowel sounds. Therefore, these 33 basic characters with their seven forms give 7*33 syllable patterns (syllographs), or fidels. Amharic writing system has of thirty-three core characters. The thirty-three characters occur in one basic form and six other forms known as orders, as shown in Table 2-3 below. These orders are derived from the basic forms by regular modification. The seven orders of the Ethiopic represent the different sounds of a consonant- vowel combination known as syllabic.

Geéz	Kaéb	Salis	Rabé	Hamis	Sadis	Sabé
ሀ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ
ለ	ሉ	ሊ	ላ	ሌ	ል	ሎ
ሐ	ሑ	ሒ	ሓ	ሔ	ሕ	ሖ
መ	ሙ	ሚ	ማ	ሜ	ም	ሞ
ሠ	ሡ	ሢ	ሣ	ሤ	ሥ	ሦ
ረ	ሩ	ሪ	ራ	ሬ	ር	ሮ
ሰ	ሱ	ሲ	ሳ	ሴ	ስ	ሶ
ሸ	ሹ	ሺ	ሻ	ሼ	ሽ	ሾ
ቀ	ቁ	ቂ	ቃ	ቄ	ቅ	ቆ
በ	ቡ	ቢ	ባ	ቤ	ብ	ቦ
ቨ	ቩ	ቪ	ቫ	ቬ	ቭ	ቮ
ተ	ቱ	ቲ	ታ	ቴ	ት	ቶ

Table 2-3: Amharic Characters with their seven orders[42]

2.4.3 Amharic Numeration System

As shown in table 2.4 below Amharic numeration system consists of basic single symbols for one to ten, for multiple of ten (twenty to ninety), hundred and thousand. These numerals are derived from the Greek numerals with some modifications. Each symbol has a horizontal stroke below and above. There is no symbol for representing zero value in Amharic number system, and it is not a place value system, thus arithmetic computation using this system is very difficult. As a result, in most printed document Hindu- Arabic numerals are used.

Ethiopic	Arabic	Ethiopic	Arabic	Ethiopic	Arabic
1	፩	11	፲፩	30	፴
2	፪	12	፲ ፪	40	፵
3	፫	13	፲፫	50	፶
4	፬	14	፲፬	60	፷
5	፭	15	፲፭	70	፸
6	፮	16	፲፮	80	፹
7	፯	17	፲፯	90	፺
8	፰	18	፲፰	100	፻
9	፱	19	፲፱	1000	፻፻
10	፲	20	፳		

Table 2-4: Amharic Numeric System [43]

2.4.4 Challenges of Amharic writing

Character recognition from document images that are printed in Ethiopic script is a challenging task. To develop a successful character recognition system for Amharic script, the following issues are addressed.

2.4.4.1. Large number of characters in the script

The total number of characters in Amharic script is more than three hundred. Existence of such a large number of Amharic characters in the writing system is a great challenge in the development of Amharic character recognizer. Memory and computational requirements are very intensive. We need to design a mechanism to compress the dimension of character representation to come up with computationally efficient recognizers [44].

2.4.4.2 Visual similarity of most characters in the script

There are a number of very similar characters in Amharic script that are sometimes difficult for humans to distinguish them easily. Robust discriminant features needs to be extracted for classification of each of the character into their proper category or class[44].

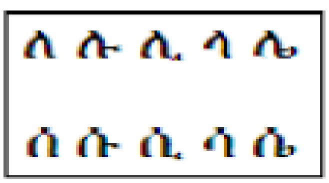


Figure 2-5 Samples of visually similar characters in Amharic writing system

2.4.4.3 Language related issues

African indigenous languages pose many additional challenges. Some of these are: (i) lack of standard representation for the fonts and encoding, (ii) lack of support from operating systems, browsers and keyboard, and (iii) lack of language processing routines. These issues add to the complexity of the design and implementation of an optical character recognition system.

2.4.4.4 Degradation of documents

Document images from printed documents, such as books, magazines, newspapers, etc. are extremely poor in quality. Popular artefacts in printed document images include [43]

(a) Excessive dusty noise,

(b) Large ink-blobs joining disjoint characters or components,

(c) Vertical cuts due to folding of the paper,

(d) Cuts at arbitrary direction due to paper quality or foreign material,

(e) Degradation of printed text due to the poor quality of paper and ink,

(f) Floating ink from facing pages etc.

This is the main issue where most character recognition research even for Latin and non-Latin scripts also fails. Carefully design on an appropriate representational scheme and classification method is needed so as to accommodate the effect of degradation[44].

2.5 Related Works

Many researchers developed ID card data detection for different languages like English, Chinese, Vietnam and Spanish. Some of these works are presented as follows:

Wira Satyawan et al [15] compares the result of character recognition of ID card name and identity number in ID card using two different tesseract models. The first model uses the train data manually that created from five ID card as data set and training on tesseract 3.05 with the support of software QT-box version 1.08 and using tesseract version 4.0 for OCR, which in that version has implemented neural network model that is LSTM.

Image processing techniques in this research consist of pre-processing, text area extraction, and segmentation. OCR proposed for character recognition. This research combines grayscale pre-processing techniques with binary image processing techniques such as Sobel, morphological transformation, and OTSU. Text area extraction uses a kernel that identifies the text area of the NIK and name on the ID card citizen. The experiments with training data made using tesseract 4.0 show that accuracy of detection reach between 90 - 100 % using the proposed technique. They also created another model with training data created using qt-box as benchmarks.

In addition, Kurama [15] discussed how any organization can use deep learning to automate ID card information extraction, data entry and reviews procedures to achieve greater efficiency and cut costs. They used different deep learning approaches and object detection algorithms such as Faster-RCNN, Mask-RCNN, YOLO, SSD, and Retina Net for this problem, and compared the results. Here the models are trained on characters, which are then recognized as objects in the images. They identified texts from images using object detection.

On another study Rodolfo Valiente et al [45] recommended a process for text recognition of generic identification documents over cloud computing. It efficiently combines MSER, a locally adaptive threshold method for text segmentation and a rectification correction using the Hough transform algorithm. Some of the operations described are too intensive to be run in a mobile phone; therefore, cloud computing is used for processing. By using MSER Algorithms, experimental results confirm that the process enhances the result of OCR, allowing it to obtain better accuracy of words recognition. As a mobile device and cloud computing used, the result is highly functional and interesting. Nevertheless, the recognition process has some inaccurate results for images with complex backgrounds and different.

On other hand, Kirubel Abebe et al [22] applied a robust method for Ethiopian vehicle plate recognition using sift features. The Proposed work contains modules such as vehicle plate detection (VPD) and optical character recognition for Ethiopia license plate number. Also used sift features for features extractions in order to generate sift descriptor of extracted character for each character of plate number and the template class pre-set descriptors. Tested the method with many plate images captured in different environments and even class distribution along with each class. The experimental results show that the proposed algorithm can detect the extract of vehicle plate (86.4%) and recognize it in different environments with 84.67% rate [46].

Martin and Villa Verde [47] have developed easy methods to deal with Spanish ID card images in order to extract useful data from them, they applied these methods in two very different situations :first a still card, captured with a scanner (getting high quality images) and second, a moving card, captured with a video camera(getting lower quality).These two systems have different application the former could be a complete data extraction system, useful for data acquisition application, whereas the latter is a system aimed for access control purposes (using

ID-card as a login document).this last system could be combined with a biometric verification system to get a very safe system.

Most of the previous studies are done for Chinese [12], Vietnam[18] ,Spanish[47] and other non-English languages to extract information from identity card and data entry to achieve greater efficiency. However, when we come to our country, with their own scripts there is no research work done specially on identity card which is written in Amharic language and future developments regard the extension of the proposed method in whole process of the recognition of customer's Identity card. More by collecting different ID model in order to test and improve the accuracy of the method.

Therefore, this study intended to fill this gap in making the recognition in automated way in case of Ethio-Telecom while customer buys different products or services and discusses how this organization can use image processing to extract information from identity card and data entry to achieve greater efficiency.

CHAPTER THREE

DESIGN AND METHODS

3.1 Overview

To solve all pinpointed difficulties, it is possible to obtain necessary information from customer's ID Card. Image processing technique used as an alternative solution of manual input process. This process starts by extracting information in ID card image. Then, it has been pre-processed to obtain the necessary part of image. Furthermore, Optical Character Recognition (OCR) will perform in order to recognize text in images. This research is intended to automate and extract useful information from ID card of customers.

3.2 Proposed Architecture

Figure 3-1 shows the architecture for the proposed approach. The proposed approach for Customers Identity Card Data Detection and Recognition consist of a series of tasks such as scanning the ID card for image acquisition, which is followed by image preprocessing for binarization, Skew Detection and Correction as well as Underline Detection and Removal.

The segmentation stage detects the scanned ID image files, categorizes them into homogeneous blocks such as photo, textual region and signature. After that, extracted text regions faded to tesseract OCR to recognize and extract the text information's. After all steps are finished all extracted information is stored on the database for text image detection and recognition.

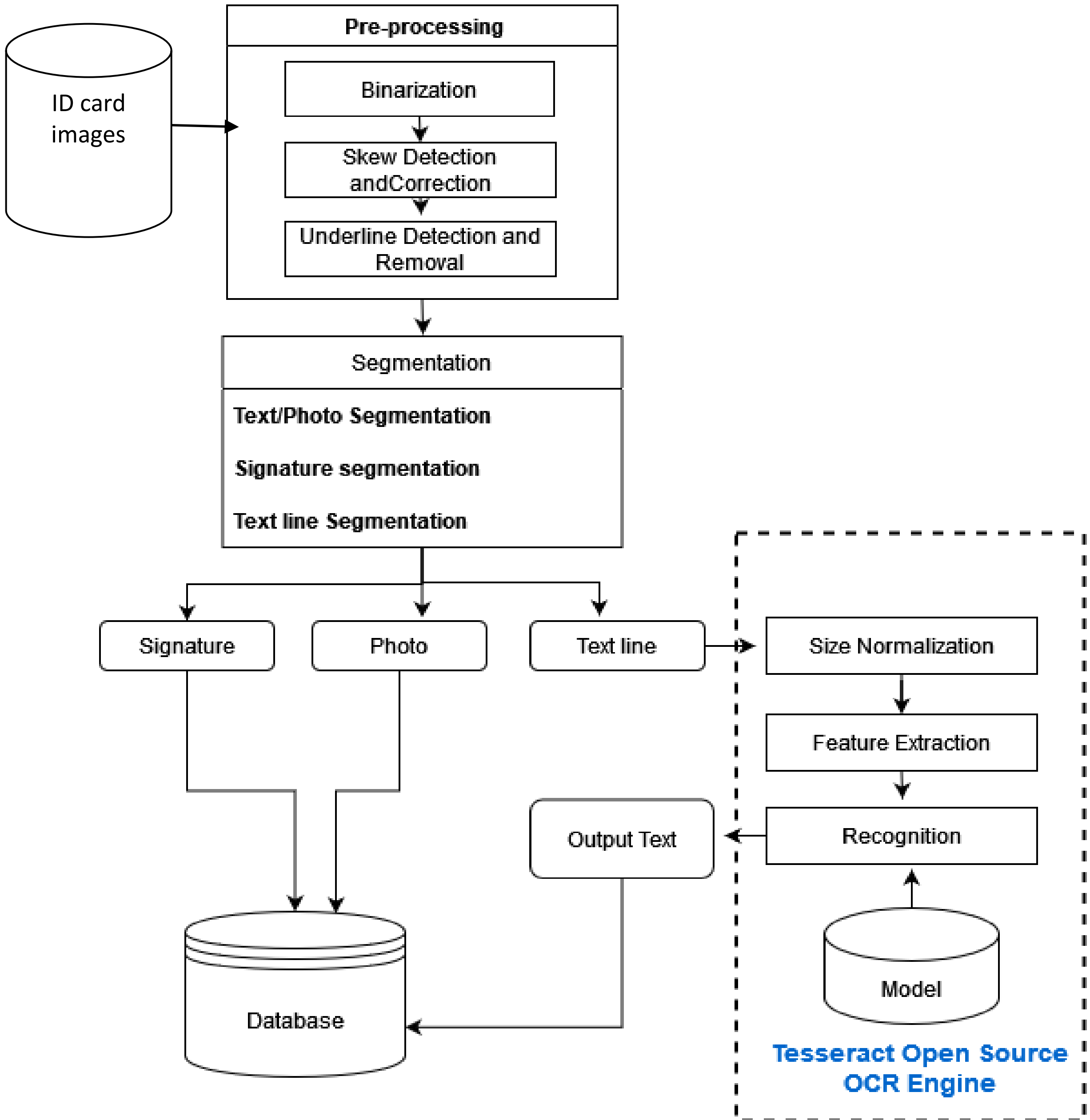


Figure 3 -1 Architecture of the proposed approach

3.3 Pre-processing

This is the most important and crucial step that is very helpful to enhance the performance of segmentation and extraction steps. Major pre-processing tasks while working with scanned ID card document images are binarization, Skew Detection and Correction as well as Underline Detection and Removal.

3.3.1 Binarization

The Binarization method converts the grey scale image (0 up to 256 grey levels) in to black and white image (0 or 1). The result of page layout segmentation highly depends upon the binarization. The high quality binarized image can give more accuracy.

The common method used to binarize is known as thresholding. In thresholding, the grayscale or colour images are represented as binary images by picking a threshold value. There are two categories of thresholding. In the case of **Global thresholding,** a threshold value is selected for the entire document image which is frequently based on an estimation of the background intensity level with that of the image using an intensity histogram. On the other hand, **Local or adaptive thresholding** uses different values for each pixel depending on the information at different pixel point. Local thresholding is commonly used in works that involve images that are of varying level of intensities, such as pictures from satellites cameras or scanned medical images.

Real-life documents especially ID card images are designed deliberately with stylistic, colorful, and complex backgrounds, causing difficulties in character extraction methods. While global thresholding techniques can extract objects from simple, uniform backgrounds at high speed, local thresholding methods can eliminate varying backgrounds at a price of long processing time [4][25][20]. From the local adaptive thresholding algorithms Sauvola's Method, Bernsen's method, Chow and Kaneko's method, Eikvil's method, Mardia and Hainsworth's method, Niblack's method, Taxt's method, Yanowitz and Bruckstein's method, White and Rohrer's Dynamic Threshold Algorithm, White and Rhorer's Integrated Function Algorithm, Parker's method, and Trier and Taxt's method the Sauvola's, Niblack's and Bernsen's methods along with post processing step are proposed by literatures were the fastest and best.

3.3.1.1 Otsu Thresholding Method

Among many global thresholding techniques, the Otsu threshold selection is ranked as the best and the fastest global thresholding method as mentioned by many scholars[48]. The thresholding operation corresponds to partitioning the pixels of an image into two classes, $C0$ and $C1$, at a threshold t. The Otsu method solves the problem of finding the optimal threshold t

* by minimizing the error of classifying a background pixel as a foreground one or vice versa. Without losing generality, we define objects as dark characters against lighter backgrounds. For an image with gray level ranges within $G = \{0, 1, . . . , L - 1\}$, the object and background can be represented by two classes, as $C_0 = \{0, 1, . . . , t\}$ and $C_1 = \{t + 1, t + 2, . . . , L - 1\}$[27].

Otsu's thresholding method is a global binarization method which involves iterating through all the possible threshold values and calculating a measure of spread for the pixel levels on each side of the threshold, i.e. the pixels that either treated as foreground or background [25]. The aim is to find the threshold value where the sum of foreground and background spreads is at its minimum. This method gives satisfactory results when the number of pixels in each class is close to each other. The Otsu method still remains the reference for comparing different thresholding and binarization methods in general. Camera flash has uneven lighting conditions due to this global thresholding method for binarization is not ideal.

3.3.1.2 Niblack's Local Adaptive Thresholding Method

Niblack's local adaptive thresholding is one of the most popular local threshold binarization method which is proposed by W. Niblack. This method involves calculating for each image pixel the mean and the standard deviation of the gray level value of the neighboring pixels that are found in a window of a predefined size. This size influences the quality of the output and it is recommended to be small enough to conserve local details and large enough to suppress noise [27].

The algorithm of Niblak's method is calculated as:

$$T(x,y) = m(x,y) + w \times s(x,y) \tag{3.1}$$

Where, T is the threshold, m(x,y) is the mean, s(x,y) is the standard deviation of all pixels in the window and w is a constant, which determines how much of the total print object edge is retained, and has a value between 0 and 1.

3.3.1.3 Sauvola Local Thresholding Method

Sauvola's algorithm claims to improve Niblack's method by computing the threshold using the dynamic range of image gray-value standard deviation[27]. Niblack proposed that a threshold for each pixel be calculated based on the local mean and local standard deviation. Sauvola's variant of Niblack's method is implemented by dividing the grayscale image into N ×N adjacent and non-overlapping blocks and processing on each block separately. In Sauvola's binarization method, the threshold $t(x, y)$ is computed using the mean $m(x, y)$ and standard deviation $s(x, y)$ of the pixel intensities in a $w \times w$ window centered on the pixel (x, y) [27]:

$$t(x,y) = m(x,y)\left[1 + k\left(\frac{s(x,y)}{R} - 1\right)\right] \tag{3.2}$$

Where, R is the maximum value of the standard deviation (R = 128 for a grayscale document), and k is a parameter which takes positive values in the range [0.2, 0.5]. The local mean $m(x, y)$ and standard deviation $s(x, y)$ adapt the value of the threshold according to the contrast in the local neighborhood of the pixel.

3.3.1.4 Bernse's Local Thresholding Method

The Bernsen's local thresholding method computes the local minimum and maximum for a neighborhood around each pixel $f(x, y) \in [0, L-1]$, and uses the median of the two as the threshold for the pixel in consideration [49].

$F\max(x, y)$ and $F\min(x, y)$ are the maximal and minimal values in a local neighborhood centered at pixel (x, y).To avoid "ghost" phenomena, the local variance $c(x, y)$ can be computed as:

$$g(x, y) = (F_{\max}(x, y) + F_{\min}(x, y))/2, \tag{3.3}$$

$$b(x, y) = \begin{cases} 1 & \text{if } f(x, y) < g(x, y), \\ 0 & \text{otherwise.} \end{cases} \tag{3.4}$$

3.3.2 Skew Detection and Correction (De-Skewing)

A text line is a group of characters, symbols, and words that are adjacent relatively close to each other, and through which a straight line can be drawn (usually with horizontal or vertical orientation) Deviation of the baseline of the text from horizontal direction is called skew.

The dominant orientation of the text lines in a document page determines the skew angle of that page. Skew is result of improper ID card feeding into the scanner. During the ID card scanning process, the whole ID card or a portion of it can be fed through the scanner. Some parts may not be fed straight into the scanner, however, causing skewing of the bitmapped images of these pages. So, document skew often occurs during ID card scanning process[30].

An ID card originally has zero skew, where horizontally or vertically printed text lines are parallel to the respective edges of the paper, however when an ID card is manually scanned or photocopied, non-zero skew may be introduced. This effect visually appears as a slope of the text lines with respect to the x-axis, and it mainly concerns the orientation of the text lines; since such analysis steps as OCR, retrieval and page layout analysis most often depend on an input page with zero skew, it is important to perform skew estimation and correction before these steps. Also, since a reader expects a page displayed on a computer screen to be upright in normal reading orientation, skew correction is normally done before displaying scanned pages. Skew detection is one of the primary tasks to be solved in document image analysis system, and it is necessary for analysing a document before further processing. Therefore, the document image may also need de-skewing. After skew detection, the image is usually rotated to zero skew angles. An example of skewed document image is displayed in Figure 3-2 below.

Figure 3-2 Skewed ID card (left) and Skew corrected ID card(right)

There are three kinds of skew in the images document they are: Single Skew, Multiple Skew, Non-uniform text line Skew [28]. In single skew, whole parts of documents are skewed to single angle. Most of the document images have this kind of skew. A lot of work has been done in this field and lot of research continues to be happening. In the case of **Multiple Skew,** scanned documents have several parts; each may be rotated to totally dissimilar angles. Sensing such variety of skewness needs lot of efforts. Multiple Skew issues are present, but it has not got a much consideration from researchers[50]. Documents having many directions within the single line are named as non-uniform text line skew. In non-uniform text line skew there are many orientations in every line of the document.

Many methods have been proposed by many researchers for the detection of skew in binary image documents. The majority of them are based on Projection file, Fourier transform, Hough transform. Main advantage of Hough transform is its accuracy and simplicity. Even though Hough transform is slow in speed due to its good accuracy the proposed system used it to detect and correct skew. In case of single skew Hough transform method used to deskew images back to normal.

3.3.3 Underline Detection and Removal

Just like other scanned documents, scanned ID cards also uses underlines usually for the identification of personal information. They are drawn horizontally below the specific text. Scanned ID cards images also contain lines other than underlines like over line, vertical line and some other different lines that are used for different purpose.

Underlines in scanned ID cards images can occur untouched with the text line they are emphasizing or touched with the lower parts of some characters in the text lines. They can also be found fragmented (or disconnected) and slightly curved.

In this study, introduced and applied that is based on *Hough Transform* and the algorithm is modified to detect, plot and remove underlines and other possible lines from the images. The Hough Transform method was introduced, in its most elementary form, by P.V.C. Hough in 1962, in the form of a patent. It is a technique, which can be used to isolate features of a particular shape within an image. Its intended application was in particle physics, for detection of lines and arcs in photographs obtained at cloud chambers. Many elaborations and refinements of this method have been investigated. Hough transform is a transform used to detect straight lines.

To apply the transform, first some pre-processing including edge detection is desirable. It is very helpful to detect not only the underlines, but also every line in the image weather it is on the underline or over line can be removed using this method. In Hough transform, a line in the image space can be expressed with two variables. For example: *Cartesian coordinate system*, parameters: (m, b) and *Polar coordinate system*: parameters: (r, ɵ) are considered as a line as shown in figure 3-3.

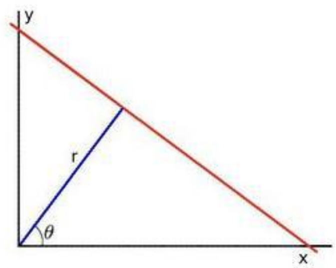


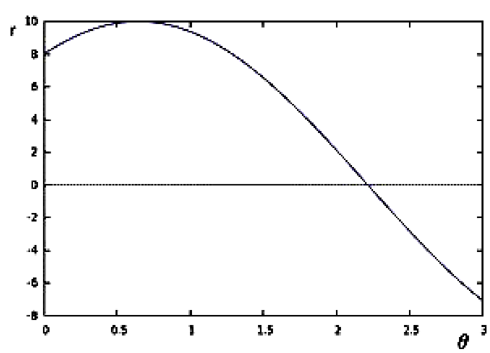
Figure 3-3 Polar coordinate System

For Hough Transforms, we express lines in the *Polar system*. Hence, a line equation can be written as:

$$\boldsymbol{y} = \left(-\frac{\cos\theta}{\sin\theta}\right)\boldsymbol{x} + \left(-\frac{r}{\sin\theta}\right)$$

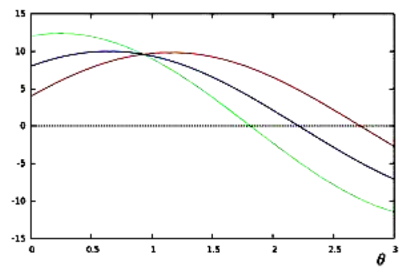
By arranging the terms, it gives: $r = xcos(\theta) + y\sin(\theta)$ and generally:

For each point (x(x_0,y_0**)**, we can define the family of lines that goes through that point as: and it means that each pair $(r\theta, \theta\ )$ represents each line that passes by (x_0,y_0**)**.The other point here is for a given line (x_0,y_0**)**, we plot the family of lines that goes through it, and we get a sinusoid. For instance, for $x_{0=}$ 8 and $y_{0=}$ 6, we get the following plot of a $(\theta - r)$ plane.



a $(\theta - r)$ plane

We can do the same operation above for all the points in an image. If the curves of different points intersect in the plane $(\theta - r)$ that means that both points belong to a same line. The following plotting example shows the plot for two more points. The line can be detected by finding the number of intersections between curves and when more curves intersects, that means the line represented by that intersection have more points and we can define a threshold of the minimum number of intersections needed to detect a line.



a$(\theta - r)$ plane with more than one point and the intersections represent the line

This is what the Hough Line Transform does. It keeps track of the intersection between curves of every point in the image. If the number of intersections is above some *threshold*, then it declares it as a line.

3.4 Page Segmentation

Page segmentation is the next step to follow the prepressing of a scanned ID card converted to images and binaries. It is performed to separate text, photo, and signature blocks. Then, the next image processing is applied over the textual area to recognize textual data. Therefore, page segmentation is an important stage of the proposed approach because the remaining stages heavily depends on this stage. Thus, in this study, page segmentation techniques are applied to extract text regions from non-text regions signature regions.

This study explored five segmentation techniques namely: watershed transforms, run length smoothing, connected component labelling, whitespace analysis and morphological dilution. These techniques are experimented in different types of ID Cards.

3.4.1 Morphology Dilution approach

The morphology approach quantitatively describes the shape of objects in an image and has recently attracted much attention[41]. The mathematical morphology describes such operations by combinations of basic set operations between an image and a small object called a structuring element. It is very attractive for this purpose because it efficiently deals with geometrical features such as size, shape, contrast, or connectivity that can be considered as segmentation-oriented features. One of the advantages of using morphological approach is its low computational cost. Filters based on opening and closing by partial reconstruction can efficiently achieve the simplification for segmentation. The size of structuring element is progressively decreased to allow the introduction of more local information to improve the segmentation[41]. In this paper, considering these advantages of morphological approach, it is used for segmenting the document PDF images.

Dilation is one of the most basic morphological operations. It is used to connect characters in words, words in a text line, and text lines in a column by adding pixels to the boundaries of objects in an image. The number of pixels added to the objects in an image depends on the size and shape of the structuring element used to process the image. In the morphological dilation operation, the state of any given pixel in the output image is determined by applying a rule to the corresponding pixel and its neighbours in the input image. The dilation rule used to process the pixels is; the value of output pixel is the maximum value of all pixels in the input pixel's neighbourhood. For instance, in a binary image, if any of the neighbourhood pixels values are 1, the output pixel is set to 1 and if both neighbourhood values are 0, the output pixel is set to 0 [51] .

The dilation function applies the appropriate rule to the pixels in the neighbourhood and assigns a value to the corresponding pixel in the output image by using structuring element.In figure 3-4, the morphological dilation function sets the value of the output pixel to 1 because one of the elements in the neighbourhood defined by the structuring element is on. Structuring element is an essential part of the dilation operation which is used to probe the input image. It is a matrix consisting of only 0's and 1's that can have any arbitrary shape and size. It can be vertical, horizontal, cross-shaped, multi directional. Based on its shape, structuring element determine to what direction it increases the pixel value of an image[41].

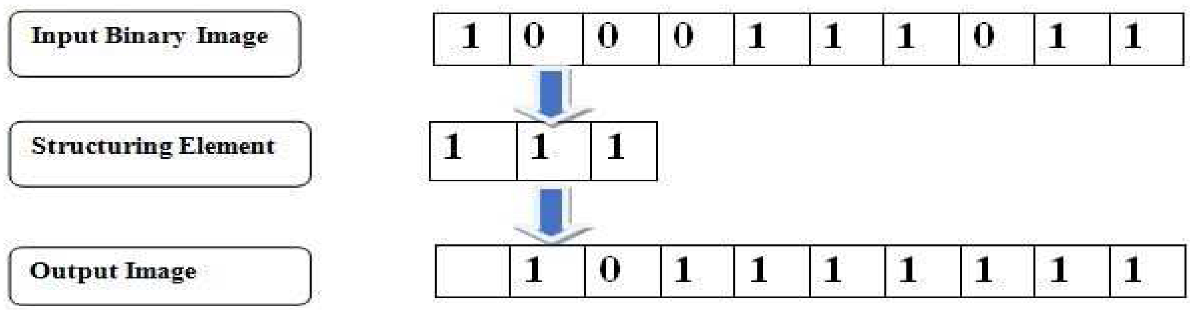


Figure 3-4 Morphology Dilution approach

3.4.2 Connected Component Analysis

Connected component (CC) labelling is used in image processing to detect connected regions in binary images. It is an algorithmic application of graph theory, where subsets of connected components within an image are uniquely labelled based on a given heuristic. Connected components scans an image and groups its pixels into components based on pixel connectivity. All pixels in a connected component share similar pixel intensity values and are in some way connected with each other. Once all groups have been determined, each pixel is labelled with a grey level or a colour (colour labelling) according to the component it was assigned to.

Connectivity of pixels divided in to 4 and 8 connectivity in order to find the CC of the given image depending on its purpose. The difference between 4 and 8 CC connectivity labelling is how the algorithm defines connected pixels. For example, for the pixel P, 4 connectivity only checks the four neighbours, called direct-neighbours i.e., right, left, up and down neighbours of P whereas 8 connectivity is known as indirect-neighbours checks all the surrounding pixels around P including diagonal pixels. The labelled pixels represent pixels that are considered as connected to the central pixel in both approaches[52].

Once all groups have been determined, each pixel is labelled with a grey-level or colour labelling according to the component it was assigned to. Extracting and labelling of various disjoint and connected components in an image is central to many automated image analysis applications such as OCR systems.

There are two types of connected component labelling algorithm; one pass and two pass. The one pass version goes through each pixel only once and for each pixel in an image, all the neighbour pixels are tested for connectivity to label connected components and the two pass scans the image two times. The first pass goes through each pixel and checks each pixel and using these pixel labels, it assigns a label to the current pixel and the second pass cleans up any mess it might have created. Two pass labelling takes high processing time and memory space than one pass. Algorithm 2.1 shows the steps followed in one pass connected component analysis.

1. Connected-component matrix is initialized to size of image matrix.
2. A marker is initialized and incremented for every detected object in the image.
3. A counter is initialized to count the number of objects.
4. A row-major scan is started for the entire image.
5. If an object pixel is detected, then following steps are repeated until (Index ! =0)
    5.1. Set the corresponding pixel to 0 in Image.
    5.2. A vector (Index) is updated with all the neighboring pixels of the currently set Pixels.
    5.3. Unique pixels are retained and already marked pixels are removed.
    5.4. Set the pixels indicated by Index to 1 in the connected-component matrix.
6. Increment the marker for another object in the image

Algorithm 3-1: One pass connected component labeling algorithm [52]

Two pass labelling scans the image two times as it has been mentioned earlier and algorithm 3.1 presents the two-pass connected component labelling algorithm [52]

First Pass:

1. Iterate through each element of the data by column, then by row (Raster Scanning)
2. If the element is not the background
    2.1. Get the neighboring elements of the current element
    2.2. If there are no neighbors, uniquely label the current element and continue
    2.3. Otherwise, find the neighbor with the smallest label and assign it to the current element
    2.4. Store the equivalence between neighboring labels

Second Pass:

1. Iterate through each element of the data by column, then by row
2. If the element is not the background

2.1. Relabel the element with the lowest equivalent label

Algorithm 3-2 Two pass connected component labeling algorithm[52]

3.4.3 Watershed Algorithm Based on Connected Components

Watershed algorithm based on connected components is one of the algorithms used to segment Amharic PDF document in this study. This algorithm gives the same segmentation results as the traditional watershed algorithm. At the same time, it has an advantage of lower complexity, simple data structure and short execution time. It connects each pixel to its lowest neighbour pixel and all pixels connected to same lowest neighbour pixel, make a segment [53]. In the analysis of the objects in images it is essential that we can distinguish between the objects of interest and the rest. This latter group is also referred to as the background. The techniques that are used to find the objects of interest are usually referred to as segmentation techniques – segmenting the foreground from background [54].

The basic concept of connected components-based watershed algorithm is shown in Figure 3-5. The original 6 x 6 image has three local minimum values indicated by grey boxes (3-5a). If a pixel is not a local minimum, then it is connected to its lowest neighbour as shown by arrows in (3-5b), where m indicates a local minimum. All components directed towards the same local minimum make a segment and are given the same label value (3-5c) [55].

7	4	8	12	11	3
7	7	8	12	11	7
13	13	15	16	16	13
19	19	18	17	15	7
20	18	17	16	15	5

(a) The original image

→	m	←	←	→	m
↗	↑	↖	←	↗	↑
↑	↑	↖	↖	↗	↑
↑	↑	↑	→	↘	↓
→	→	→	→	→	m

(b) Each pixel connect to lowest minimum

0	0	0	0	1	1
0	0	0	0	1	1
0	0	0	0	1	1
0	0	0	2	2	2
2	2	2	2	2	2

(c) The image with labels

Figure 3-5 Example of connected components-based watershed algorithm[55]

3.4.4 Run Length Smoothing

The Run Length Smoothing Algorithm (RLSA) is a method that can be used for Block segmentation and text discrimination. The method developed for the Document Analysis System consists of two steps. First, a segmentation procedure subdivides the area of a document into regions (blocks), each of which should contain only one type of data (text, graphic, halftone image, etc.). Next, some basic features of these blocks are calculated[56].

The basic RLSA is applied to a binary sequence in which white pixels are represented by 0's and black pixels by 1's. The algorithm transforms a binary sequence x into an output sequence y according to the following rules[56]:

1. if the number of adjacent 0's is less than or equal to a predefined threshold C, then change 0's in x to 1's in y.
2. 1's in x are unchanged in y.

3.4.5 Line Segmentation

To correctly distinguish probable line segments from scanned ID card image, this study used Y histogram projection. Since document image was pre-processed and free from skew by removing skew, line segmentation was done. Procedures for line segmentation are as follows[55].

- *First read pre-processed images,*
- *then count the black pixel in each row,*
- *Then construct the horizontal histogram for the image.*
- *Using the Histogram, find the rows containing white pixel.*
- *Using the start and end line of white pixel, mark the bounding box for text.*

The procedure starts from top of image to down up to the end of the image.

3.5 Text Recognition using OCR engine

OCR is a technology that allows converting scanned images of text into plain text. This study explored three open-source OCR engines; namely, Tesseract OCR, FineReader and FreeOCR. These OCR engines are experimented in different combinations on real-life Amharic document images.

3.5.1 Tesseract OCR

Tesseract is an open source optical character recognition (OCR) engine[57]. HP originally was originally started it as a project [44]. Later it was modified, improved, and taken over by Google and later released as open source in year 2005. It is very portable as compared to others and supports various platforms. Its focus is more towards providing less rejection and improved accuracy. Currently only command base version is available but there are many projects with UI built on top of it which could be forked. As of now Tesseract version 3.02 is released and available for use, which provides support for around 139 language including Amharic. Figure 3-6 presents python code used to recognize Amharic text image using Tesseract.

```
for file in file_list:
    name=os.path.splitext(file)[0]
    # selecting image file type
    if file.endswith(".jpg"):
        txt=ocr(file) # calling the ocr function
                                os.remove(name+".jpg")
        file = open(directory+"\\"+name+".txt",'a+', encoding="utf-8")
                                #os.remove(file)
            file.write(str(txt))
print (success)
```

Figure 3-6 python code used to recognize Amharic text image using Tesseract

3.5.2 FineReader OCR

A global company, called ABBYY, as advanced OCR engine, produces FineReader commercially. The performance of FineReader has been enhanced by ABBYY for many years. FineReader 12 supports 190 languages including Amharic script using dictionary support[58]. It supports multi-font types, multi-size and multi-resolution images.

3.5.3 FreeOCR

FreeOCR is a free Optical Character Recognition engine that supports different popular image file formats, such as JPG,PNG,TIFF,PDF and GIF. FreeOCR outputs plain text and can export directly to Microsoft Word format [57].

3.6 Performance Evaluation

In order to evaluate proposed method and the dataset obtained from the planned and executed experiment, direct mapping is used. It determines the performance of segmentation and text recognition by finding the correspondences between detected entities and ground truth.

The performance of recognized data can be analysed in several aspects. We can measure the algorithm Accuracy, computational memory, speed and quality of obtained data. The most common measure to calculate the efficiency of segmentation algorithm is degree of Accuracy. It is defined as the ratio of total number of bits required to store uncompressed data and total number of bits required to store compressed data. It is defined as the number of actual characters with their places divided by the total of actual characters.

3.6.1 Accuracy

Accuracy is a metric that generally describes how the model performs across all classes. It is useful when all classes are of equal importance. It is calculated as the ratio between the number of correct predictions to the total number of predictions[22]

$$Accuracy = \frac{True_{positive} + True_{negative}}{True_{positive} + True_{negative} + False_{positive} + False_{negative}} \qquad (3.4)$$

3.6.2 Precision

The precision is calculated as the ratio between the number of *Positive* samples correctly classified to the total number of samples classified as *Positive* (either correctly or incorrectly). The precision measures the model's accuracy in classifying a sample as positive[21]

$$Precision = \frac{True_{positive}}{True_{positive} + False_{positive}} \qquad (3.5)$$

3.6.3 Recall

The recall is calculated as the ratio between the number of *Positive* samples correctly classified as *Positive* to the total number of *Positive* samples. The recall measures the model's ability to detect *Positive* samples. The higher the recall, the more positive samples detected [21].

$$Recall = \frac{True_{positive}}{True_{positive} + False_{negative}} \qquad (3.6)$$

CHAPTER FOUR

EXPERIMENTATION AND EVALUATION

The main purpose of this study is design Automatic Amharic Text Detection and Recognition system from identity card of customers by applying an effective page segmentation technique that can identify text and non-text blocks and applying text recognition using OCR. The proposed techniques are integrated with Google developed OCR systems tesseract to recognize Amharic and English texts from textual objects.

For the experimentation purpose, HP Intel(R) Core (TM) I7 CPU 3110M @ 2.4GHz (2 CPUs), 8GB RAM and Windows 10 Ultimate operating system were used. MATLAB™ image processing toolbox R2021 is used for developing prototype and integration.

The systems assume that the user points the E-CAF camera directly at (or even manually marks) the object that is captured, with blurring, scaling and perspective bias being the major issues that is addressed for this study excrement purpose.

4.1 Dataset Preparation

The dataset collected contain real life ID cards document images with different level of noise, skewness and perspective. The Amharic and English scanned ID card that have an embedded Photo, graphics and images are used. The datasets are collected and captured by Ethiotelecom E-CAF system. The dataset doesn't contain handwritten and typewritten ID card images; rather it only contains printed documents.

The system was tested on 42 distinct RGB Amharic and English different printed ID card images with different noise skewness and perspective level and all the test images were taken by Portable Document Camera Scanner (Model –DY1030) and have saved in JPG format. The dataset contains both simple and complex ID cards including complex backgrounds, logos and watermarks. Some cards contain multiple logos and some logos are combination of text and graphics. Most of the images are skewed, perceptively distorted, and degraded. Table 4-1 shows the different document contents used for this study.

ID Content	*No of ID Cards*
Text & Photo	12
Text, Photo & Logo	22
Text, Photo & Water Mark	6
Text, Photo & Both (Logo and Water Mark)	2

Table 4-1: summary of the dataset used for this study

4.2 Pre-processing

Preprocessing is vital use for any digital image recognition processes. It improves the recognition rate of the OCR system. Moreover, E-CAF scanned documents have more challenges than the desktop scanners because of their poor scanning capacity, lighting condition and their perspective is not fixed like computer scanner. This makes the digital image to have more noise, skewed and more perspective, so this has to be handled for better recognition.

We implemented and used built-in methods in MATLAB Image Processing Toolbox to reduce noise and convert images types. The original images were scanned as color (RGB - Red Green Blue) so we converted the image file from RGB to gray-level using *rgb2gray (image)* built-in method in MATLAB to make processing efficient and easy.

Once the document image becomes ready for processing, the first image preprocessing method implemented in this study is skew and perspective correction and then binarization followed.

4.2.1 Skew and Perspective correction

In E-CAF Scanned ID card images we face not only the skewness of the image, but there is also a perspective distortion due to the scanning process. A skew detection and perspective detection process is needed before proceeding to the segmentation and recognition process. If any skew or perspective distortion detected, an auto-rotation procedure is performed to correct the skew and the perspective before processing text further.

A perspective and skewed image of a ID card can be easily rectified with the help of its bounding rectangle. When the rectangular boundary is clearly distinguishable by locating the four corners of the card, it is possible to correct the image.

Figure 4-1 (a) Perspective distorted (b) Skewed document

We follow the following steps to correct the skewed and perspective distortion in this research:

First, identify the corners of the bounding quadrilateral (vertex of the card) in the given ID card image.

The first step toward image perspective correction or skew correction is pointing out the quadrilateral containing the ID card edges. This operation can be carried out manually or with aid of the technique. But in this research the researcher uses manual way to put the quadrilateral points.

The user after capturing the image it identifies the vertexes of the card by pointing to them in the capturing E-CAF device, the coordinates will be sent to the system. For the experiment purpose the researcher use a MATAB built in function *ginput*() to get the coordinates.

*[*X Y*] = ginput(4)* function assign the four coordinates to the array variables X and Y.

Second, map each vertex of the quadrilateral to the corresponding vertex in the known rectangle.The mapping of the retrieved coordinates from the MATLAB function ginput() is used for this study, because we just get the four coordinates in an array format that does not tell which one is for which coordinate X (x1, x2, x3, x4) and Y (y1, y2, y3, y4). After sorting the Y values, the 1st two high values for the y-axis are the top two edges so the upper corners of y axis are identified. Then we sort the coordinates according to X values. The sorting is done along different dimensions of the array and arranges those elements in ascending order. We use sort () MATLAB built in function to sort them. Get the vertex of the quadrilateral coordinate's location upper/lower and left/right, the upper left and right corners and the lower left and right corners will be identified.

v1 = (x_i, y_i)*= left_top*

v2 = (x_i, y_i) *= left_bottom v3 =*

(x_i, y_i) *= right_top*

v4 = (x_i, y_i) *= right_bottom*

Third, Transform using equations (eq. 3.11) and an additional constraint (eg. $\| h \|$ is of unit norm), and the corresponding coefficients h_{ij} by computing the Homography H. The concept of Homography is to explain and study the difference in appearance of two plane objects viewed from different points of view.

The homography H is a matrix of size 3*3. This is defined only up to a scaling and hence has only 8 unknowns. Given four corresponding points (8 equations) in a general position, H can be uniquely computed. The four corresponding points in this research are (1,1), (2400,1), (2400,1500) and (1, 1500) which are the target corrected corner (vertex) point of the document image, the researcher set this values based on the ID card size and the size of the image.

Mostly, we have 2 collections of 4 points, A and B rectangles, in which we want to map a point in rectangle A to rectangle B (target point).

A = { p1x,p1y ; p2x,p2y ; p3x,p3y ; p4x,p4y }

B = { p1x,p1y ; p2x,p2y ; p3x,p3y ; p4x,p4y } = {1,1; 2400,1; 2400,1500; 1,1500}

Then we build a 8 x **n** matrix (M1), where **n** is number of points * 2(x,y). (Result will be a 8x8 matrix), the MATLAB function *zeros*(8,8) will create and set zeros value to the matrix.

For each point:

{xA, yA, 1, 0, 0, 0, -xA*xB, -yA*xB}

{0, 0, 0, xA, yA, 1, -xA*yB, -yA*yB}

We also make a 1 x 8 matrix (M2) with the "target"(B) points. Result must be 1x8 for each point:

{xB}

{yB}

Once calculated, these 8 parameters can easily be used to transform any point from the first reference system to the second. Then we can calculate the homograph, result will be a 3 x 3 matrix (H). H = M1 / M2, and we reshape the matrix size to 3*3

Fourth, using H rectify the image to the frontal view.

To rectify the image from the original position to the frontal view we use this MATLAB built in functions *maketform()* and *imtransform()*. The *maketform()* function creates a multidimensional spatial transformation structure that can be used with the *imtransform* function. The *imtransform()* transforms the image according to the 2-D spatial transformation defined by tform (Transformation Structure), in RGB image it applies the same 2-D transformation to all 2-D planes along the higher dimensions.

Projective transformation code:

```
T=maketform('projective',H'); TImage=imtransform(image,T,'XData',[1 2400],'YData',[1 1500]);
```

By using this method all document that are skewed, or perspective distorted scanned (captured) image is corrected, if we correctly input the corner coordinates of the card. If we select the boundary corners of the ID card incorrectly, we will get a wrongly resolved image.

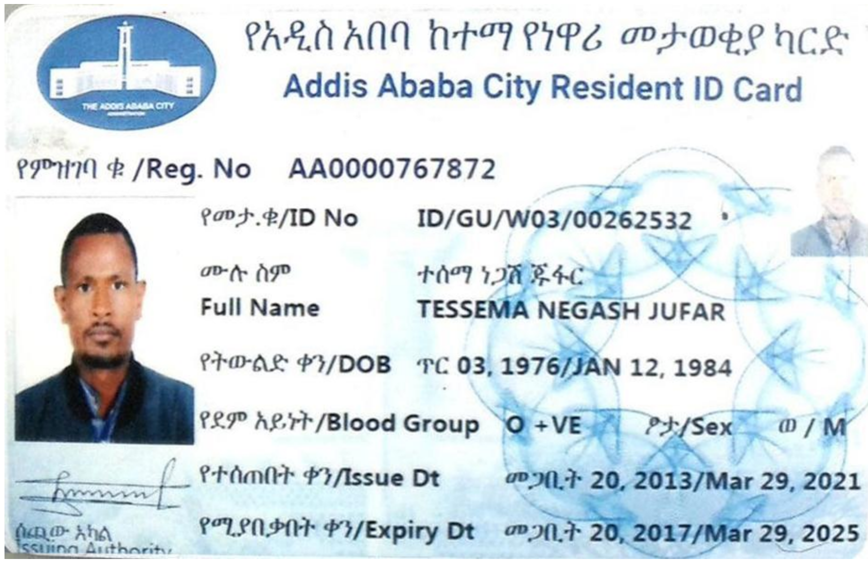


Figure 4-2 Skew and Perspective corrected document image

An experiment shows that the text image surrounding or the image background surrounding of the ID card makes it hard to use another method for the skew and perspective correction. The experiment and Literatures also show traditional methods based on morphological operations and projection methods are extremely slow and tend to fail for low quality images.

4.2.2 Binarization

Binarization is the next stage after skew and perspective correction, which is converting the gray level images to binary images. Binarization has long been recognized as a standard method to solve the lightning issue. The goal of binarization process is to classify image pixels from the given input grayscale or color document into either foreground (text) or background and as a result, reduces the candidate text region to be processed by later processing steps[27]. When a grayscale or colored image is binarized, a certain threshold value must be set so that pixel intensity levels that are less than this value is assigned to black pixels and values greater are assigned to white pixels. There are lot binarization methods but in this study the experimentation is done on three thresholding that are Otsu global thresholding, Sauvola local adaptive thresholding and Niblack local adaptive thresholding.

4.2.2.1 Otsu Global Thresholding

The Otsu thresholding algorithm is easily implemented because there is a built-in function in MATLAB Image Processing Toolbox. The MATLAB function *im2bw (i, level)* is used to find the binarized image from the grayscale image. Where *i* is the input image and *level* the threshold value of the grayscale image[27]. The threshold value is calculated by the built-in function *graythresh(i)*. The function *graythresh*() can be used to compute the level which is a value between 0 and 1.The output image replaces all pixels in the input image with luminance greater than level with the value 1 (white) and replaces all other pixels with the value 0 (black). By using this function, we can easily find the global threshold of the image.

The result of the algorithm test image is shown in Figure 4-3(b).

Implementation of Otsu Binarization:

```
function [outputImage] = otsusThreshold(Image) image = rgb2gray(Image);
        level = graythresh(Image);
        outputImage = im2bw( Image, level);
end
```

4.2.2.2 Niblack Local Adaptive Thresholding

The reason we choose this algorithm and Sauvola algorithm in our experiment is that it performs well as we have seen from earlier literature experimental results. The local adaptive binarization technique is required in a complicated document images which has different background in the same document, in this case a local threshold needed which is calculated for each pixel[27].

The Niblack filter is a local thresholding algorithm that separates white and black pixels given the local mean and standard deviation for the current window. Setting the filter radius is required and it is critical for the algorithm because the quality of the image is determined by this.

Filter Radius	MSE	PSNR
25	0.30	59.17
60	0.26	57.78
70	**0.22**	**55.13**
80	0.24	56.09

Table 4-2: Experimental result to select better filter radius

We use a filter radius of 70 and weight parameter(*w*) which is a constant (between 0 and 1) is 0.01 in our experiment as shown in the above Table 4-2 the values are found based on different experimental results. The filter size of 70 cleans the noise better than any other values. The step of this algorithm is first we build a filter of size 70 and then we calculate the mean and the standard deviation based on the filter we create. Then we calculate the threshold based on the formula, which is multiplying all the three variables the mean, the standard deviation and the constant 0.01. Finally, the pixel values that are greater than the threshold will be black and the others white this is the binary image. We found out that Niblack's binarization method produces more background noise than other (see Figure 4-3(c).

The implementation code for Niblack thresholding algorithm in MATLAB is presented below.

```
function [outputImage] = Niblack(img,varargin)
filt_radius = 70; % filter rad [pix] k_threshold = 0.01; % std
threshold parameter
= rgb2gray(img); X =
double(i);
X = X / max(X(:)); % normalize to [0, 1] range
%% build filter
fgrid = -filt_radius : filt_radius; [x, y] =
meshgrid(fgrid);
filt = sqrt(x .^ 2 + y .^ 2) <= filt_radius; filt = filt / sum(filt(:));
%% calculate mean, and std
local_mean = imfilter(X, filt, 'symmetric');
local_std = sqrt(imfilter(X .^ 2, filt, 'symmetric'));
%% calculate binary image
outputImage = X >= (local_mean + k_threshold * local_std);
end
```

The result of the algorithm test image is shown in Figure 4-3(c).

4.2.2.3 Sauvola Local Adaptive Thresholding

Sauvola recently presented promising results using a variation of Niblack's binarization [7]. The local mean $m(x,y)$ and standard deviation $s(x,y)$ adapt the value of the threshold according to the contrast in the local neighborhood of the pixel. When there is high contrast in some region of the image, the standard deviation $s(x,y) \sim R$ which, results in the threshold $T(x,y) \sim m(x,y)$ with the mean.

This is the same result as in Niblack's method. However, the difference comes in when the contrast in the local neighborhood is quite low. In that case the threshold $T(x,y)$ goes below the mean value thereby successfully removing the relatively dark regions of the background. The parameter k controls the value of the threshold in the local window such that the higher the value of k, the lower the thresholds from the local mean [27].

Window Size	MSE	PSNR
5	0.384	52.286
10	0.378	52.356
25	0.361	52.546
30	**0.357**	**52.594**
40	0.363	52.525

Table 4-3: Experimental result to select the better window size

The value of the parameter k is a bias, which takes positive values in the range [0.2, 0.5]. A value of $k = 0.5$ was used by J. Sauvola. Experiments with different values of k show that $k = 0.34$ gives the best results. The value of R which is the maximum deviation is calculated based on the mean square and the mean of the image. A block size of 30 × 30 window size was used for the research experiment because as it is shown in the Table 4-3 this window size performs good quality. Finally we use the Sauvola formula to compute the threshold (see formula 3.3). If the pixel value is greater than the optimal threshold it is black otherwise it is white, this is a binary image.

The implementation code for Sauvola thresholding algorithm in MATLAB is presented below.

```
function output=sauvola(img,varargin)
    % Initialization
    img = rgb2gray(img); numvarargs =
    length(varargin); if numvarargs > 3
        error('myfuns:somefun2Alt:TooManyInputs', ...
          'Possible parameters are: (image, [m n], threshold, padding)');
    end
    optargs = {[25 25] 0.4 'replicate'}; % set defaults optargs(1:numvarargs) = varargin;        % use memorable
    variable names [window, k, padding] = optargs{:};
    if ndims(image) ~= 2
        error('The input image must be a two-dimensional array.'); 121
    end
```

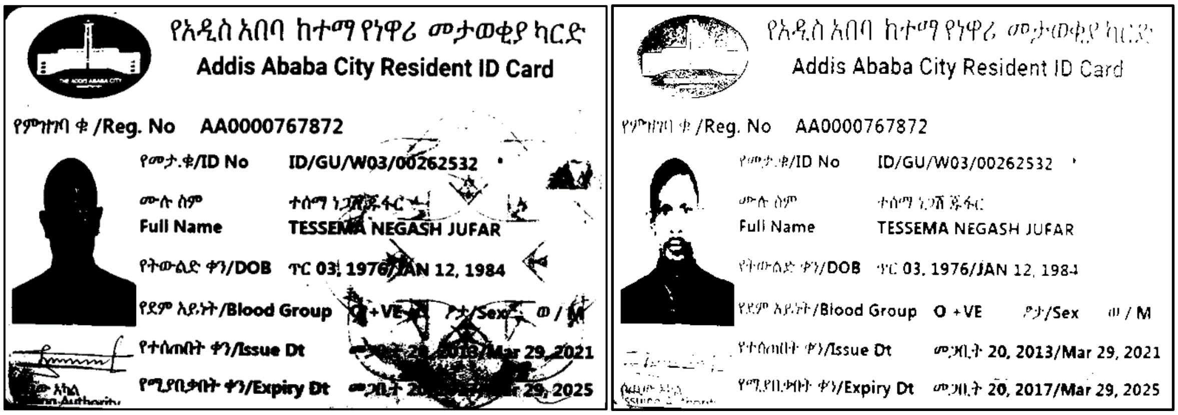
```
% Convert to double image =
double(image);
% Mean value
mean = mean_filter(image, window, padding);
% Standard deviation
meanSquare = mean_filter(image.^2, window, padding);
mean1 = mean.^2;
if meanSquare > mean1
    deviation = (meanSquare - mean.^2).^0.5; else
    deviation = (mean.^2 - meanSquare).^0.5;
end
% Sauvola
R = max(deviation(:));
threshold = mean.*(1.3 + k * (deviation / R-1)); outputImage = (image >
threshold);
end
```

The result of the algorithm test image is shown in Figure 4-1.

4.2.2.4 Experimentation on Binarization

This study has mentioned in the previous section that Wiener filter performs better in our experiment and also from previous literatures. We use this filter and combine with the three thresholding techniques to see which binarization method is better. The Figures 4-3 shows the result the three binarization methods of a ID card that is perspective & skew corrected and noise removed in Wiener filter.

(a) (b)



(c) (d)

Figure 4-3 Experimental result of Image Binarization (a) Original image after skew and perspective corrected; (b) Otsu threshold; (c) Niblack threshold; (d) Sauvola threshold

Noise Level	With Otsu		With Niblack		With Sauvola	
	MSE	PSNR	MSE	PSNR	MSE	PSNR
Low	0.20	55.09	0.16	56.07	**0.07**	**59.75**
Medium	0.36	52.52	0.19	55.21	**0.08**	**59.06**
High	0.43	49.16	0.29	52.39	**0.25**	**54.05**

Table 4-4: Experiment Result of performance evaluation

Global binarization methods are very fast and that gives a good result for typical scanned ID cards. However, if the illumination over the document is not uniform, for instance, in the case of camera-captured documents, global binarization methods tend to produce marginal noise along the ID card borders. Local binarization methods, such as Niblack's algorithm, and Sauvola's algorithm, compute thresholds individually for each pixel using information from the local neighborhood of that pixel. That is why we get good result from the MSE and PSNR for the algorithms of Niblack and Sauvola. Based on the experiments as we have seen above in Table 4-4. Sauvolas's algorithm worked better and faster. An effective pre-processing is followed up by the process of segmentation. Text area segmentation is presented in the succeeding subsection.

4.3 Page Layout Segmentation

Page layout segmentation is the next step to follow after we finished preprocessing of an image. It is performed to separate text from non-text region and to store layout information of the segmented objects then, the subsequent image processing is applied over the text area to recognize text by Appling OCR[59]. In the proposed methods, once the ID card image is converted into binary document image, the next stage is segmenting the non-text area into a set of images.

The proposed system segmentation techniques used morphological dilation, connected component (CC) analysis techniques for separating graphics from text area, detecting and segmenting text line areas, collecting information about segmented column and identifying text areas from the ID card images for further processing.

4.3.1. Text and Photo segmentation

Text/graphic segmentation in document images, which separate graphics from text area, is a procedure that must be applied over the image before the performing the next stage of system. It is crucial for the next sub-sequential stage, text segmentation which has a great impact to improve the performance of character recognition for textual regions. In this study, morphological Dilation and Connected Component (CC) analysis are applied for the purpose of segmenting texts and non-textual areas from ID card images that contain both text and graphics.

The first step in the course of text/graphics separation is dilation. It is used to connect the space between characters, word, lines and so on by increasing the pixel values of a ID card images. Depending on the structuring element there are different types of dilation. In this study vertical and horizontal direction dilations are applied.

After the ID card image is connected the next step is CC labeling analysis. Connected components (CC) labeling algorithm is applied to identify and label each connected component in a given binary image. MATLAB built-in method bwconncomp() and bwlabel() are used to identify connected component and to label them in a given binary image respectively. In this study 8 connectivity of pixel are used to identify connected components. Algorithm 4.1 shows CC algorithm that identifies connected components in a given image.

```
function [cc, num] = ConnectedComp(binary_image)
cc = bwconncomp (binary_image, 8); % 4 connectivity
Num=cc.NumObjects;
```

Algorithm 4-1: Identifying Connected Component

Once connected components it is used to identify big connected elements like Photos and logos and small connected elements like punctuation marks and small dots. Graphics and Images usually have larger height, width and area than normal text while punctuation marks, dots and others have smaller area as well as height and width. Thus, a threshold value in order to separate text from graphics is set by taking the fact that graphics have larger area than text. The width and height of the bounding box are used to compute the area for each component and saved on array size info to compare the results. This study adopts a threshold values from privies conducts study experiment 8000 is found to be a better threshold value.

```
size_info = [];
cc = 1;
forcnt = 1:num
        x = Ibox(:,cnt);
        size_info (cc,1) = x(3,:,1);
        size_info (cc,2) = x(4,:,1);
        size_info (cc,3) = x(3,:,1) * x(4,:,1);
        cc = cc + 1;
        if (size_info(cnt,3) >8000)
        rectangle ('position',Ibox(:,cnt),'edgecolor','r');
end
end
```

Algorithm 4-2: Analysis for Text\Graphics separation

4.3.2 Text Line Segmentation

After separating text area from graphics, the next step in this sequential layout segmentation is text line detection and segmentation. ID card images might contain different line of information such as name, sex, address and other information's. So it is important to detect and segment those regions for the next OCR process. Likewise, text/graphics separation, these study proposed similar technique for text line detection. It includes morphological dilation and CC labelling[31].

Dilation technique used for text/graphics separation is also applied here for text line segmentation. But, in order to keep the white space between the text lines horizontal dilation is applied Figure 4-4 shows the result of connected pixels after dilation algorithm that connects characters, words and text lines is performed.

```
textRegions = bwconvhull(textRegions, 'objects');
textRegions = bwareafilt(textRegions, [, inf]);
se = strel('line', ,);
```

```
Verticald = imdilate(textRegions,se);
CC = bwconncomp(Horizontald);
imshow(Horizontald);
stats = [regionprops(CC)];
for i = 1:numel(stats)
   rectangle('Position', stats(i).BoundingBox, ...
   'Linewidth', 1, 'EdgeColor', 'r', 'LineStyle', '-');
End
```

Algorithm 4-3:Analysis for Text Line segmentation

Figure 4-4 shows results of experiment after the proposed text/graphics separation techniques which include dilation, connected component analysis

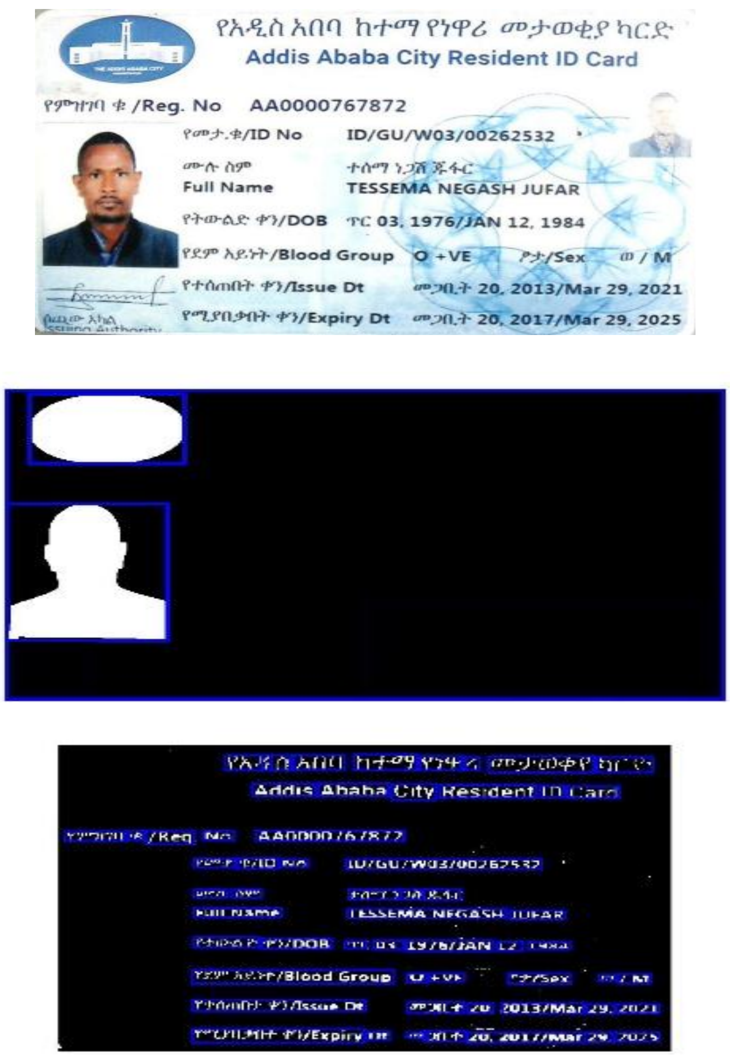


Figure 4-4 Experimental result after the proposed text/graphics separation techniques applied

4.4 Objects Extraction

After all ID card images segmentation is performed ID card Image objects extraction process is started based on the above segmentation process output. In order to extract each text and non-text region the proposed approach applies the bounding box (x,y) coordinates for each objects[24].

L = {S1, Sn}, where L and S represent layout and segment respectively

A segment is a pixel collection encapsulated within a bounding box defined by its lower left and upper right corner pixels:

S = (P1, P2), where S and P represent segments and pixels accordingly

Each pixel is defined by a coordinate pair:

P = (x, y)

Based on this structure, in the proposed system object extraction is done through the following steps:

- computing the bounding box for the connected components
- finding all edge pixels of the bounding box
- calculate maximum y, minimum y, maximum x and minimum y coordinate values
- (max Y, min X) and (min Y, Max X) will be the left upper and the right lower corner values of the segmented image
- Based on the above coordinate the object is cropped from the original RGB ID card Image.

4.4.1 Non-Text Components Extraction Such as Photos and Logos

This Stage performs document extracting non text areas such as Images pictures graphical areas based on the above segmentation result.

```
max(max(Textlabel)); % This Command gives us the maximum objects detected.
mkdir([‘Output path', savefilename,'\\photo]);
pagefolderColumens=['C Output path ', savefilename,'\photo];
measurements2 = regionprops(BW3, 'Area', 'BoundingBox');
allAreas2 = [measurements2.Area];
% Crop out each word
for blob = 1 : length(measurements2)
 % Get the bounding box.
 thisBoundingBox = measurements2(blob).BoundingBox;
 % Crop it out of the original gray scale image.
 thisWord = imcrop(rgbImage, thisBoundingBox);
```

```
 x=fix(thisBoundingBox(1));
 y=fix(thisBoundingBox(2));
 FileName = sprintf('%d, %d.jpg', x, y);
 fullFileName = fullfile(pagefolderColumens, FileName);
 imwrite(thisWord, fullFileName);
end
```

After figure extraction from the page has been completed, the figure block is appended to its own folder.

4.4.2 Textual Components Extraction

After non textual objects extracted the next steps is extracting textual objects extracting from the above page columns segmentation. The output of this stage is the input for the next OCR application.

```
max(max(Textlabel)); % This Command gives us the maximum objects detected.
mkdir([‘Output path', savefilename,'\\Text']);
pagefolderColumens=['C Output path ', savefilename,'\Text'];
measurements2 = regionprops(BW3, 'Area', 'BoundingBox');
allAreas2 = [measurements2.Area];
% Crop out each word
for blob = 1 : length(measurements2)
 % Get the bounding box.
 thisBoundingBox = measurements2(blob).BoundingBox;
 % Crop it out of the original gray scale image.
 thisWord = imcrop(rgbImage, thisBoundingBox);
 x=fix(thisBoundingBox(1));
 y=fix(thisBoundingBox(2));
 FileName = sprintf('%d, %d.jpg', x, y);
 fullFileName = fullfile(pagefolderColumens, FileName);
 imwrite(thisWord, fullFileName);
end
```

4.5 OCR Application

Once the above page layout detection, segmentation and extraction is accomplished, the next stage is extracting the text from segmented textual objects. In this study, Google developed tesseract OCR engine is used and all processes regards to OCR application is performed by tesseract OCR. Python built-in method pytesseract () is used to perform OCR process.

```
for file in file_list:
    name=os.path.splitext(file)[0]
    # selecting image file type
    if file.endswith(".jpg"):
        txt=ocr(file) # calling the ocr function
        os.remove(name+".jpg")
        file = open(directory+"\\"+name+".txt",'a+', encoding="utf-8")
        #os.remove(file)

        file.write(str(txt))

print("completed")
```

4.6 Performance Evaluation of the Proposed Technique

The proposed technique is found based on different experimental result of different ID cards. The study proposed for E-CAF system captured ID cards to use a document boundary user assisted homography skew and perspective correction, Wiener filtering, Sauvola binarization, and Connected Component labeling and Dilation to extract the text region from the ID card[27].

The input RGB digital image is passed to the skew and perspective correction technique which uses a bounding rectangle user assisted homography that the user has to select the vertex of the ID card. The proposed technique will rectify the image from the original position to the frontal view by using the homography and the vertex of the ID card. This technique will help the OCR system not only for correcting the perspective or the skew it is also used to eliminate the border noise and other external images that appear around the ID card.

After the image skew and perspective corrected then the Sauvola binarization algorithm is used for thresholding the text from the background. Finally, for the text region extraction, we propose Connected Component analysis and Dilation algorithm. In CC analysis we apply the height, width, and area analysis to differentiate the text from the logo, picture, and other noise that are found in the image. The proposed system outputs the candidate text region for next stage segmentation and character recognition[27].

The performance of the proposed technique is evaluated based on the precision rate and the recall rate of the text region extracted.

To quantify the text region separation precision and recall we have designed the following method. ID card components have a text and non-text (photo) region. The non-text region includes background texture, logos, pictures, and watermarks. All the remaining connected regions are identified as texts. While estimating the classification precision and recall four different situations may arise in the text region (II, BB, BI, and IB).

We will put all the non-text regions as an Image (I) and the text regions as Text (T). Table 4-5 shows all such possible cases.

CC/Region	**Identified as**	**Justification**
Image	Image	BB(True)
Image	Text	BT (False)
Text	Image	TB (False)
Text	Text	TT (True)

Table 4-5: Justification of classification rules

The evaluation of the system performance with the selected noise filter to other thresholding algorithms for low, medium, and high-level noise is shown in Table 4-6. The average results of text region extraction at different noise level ID card documents as we have seen Wiener with Sauvola have highest precision and recall.

Type of document image	Evaluation methods	Wiener with Otsu	Wiener with Niblack	Wiener with Sauvola
low-level noisy	Precision	82.36	87.66	90.5
	Recall	84.6	89.46	96.9
medium-level noisy	Precision	81.82	87.13	89
	Recall	84.03	87.63	96.6
high-level noisy	Precision	76.79	84.22	88.66
	Recall	82.37	87.57	96.38
Average	Precision	80.32	86.336	89.38
	Recall	83.67	88.22	96.6267

Table 4-6: System performance of thresholding

The precision and recall of a text region extraction of the proposed skew and perspective corrected image increases. The result is shown in the table 4-7. The average system performance is shown an increase in precision and recall in the skewed and perspective corrected image.

	Skewed and Perspective Distorted Image				Skew and Perspective Corrected Image			
	Low level	Medium Level	High level	Average	Low Level	Medium level	High level	Average
Precision	90.87	90.5	54.43	78.6	93.6	92.1	91.76	92.486
Recall	98.40	98.21	95.86	97.49	100	99.9	99.78	99.899

Table 4-7: Average system performance

For document images that contain small amount of noise, the performance of the system without skew correction shows 90.87% Precision and 98.40% Recall. After the proposed skew and perspective rectification applied a 93.6% Precision and 99.89% Recall is registered.

Experimental result for OCR Recognition

For this study we used tesseract OCR which has implemented a neural network model. Training data that is used is default data that has been obtained from tesseract that contains text data in a different language with different fonts. The calculation of the accuracy rate of OCR on each extracted text image is determined by the number of character extraction results.

Accuracy rate:

$character\ count - character\ error\ character\ count$ / $character\ count\ x$ 100%

Table 4-8 presents the performance of the different OCR engines to recognize texts from segmented text blocks.

file	**Text Blocks**	**Tesseract OCR (%)**
ID 1	22	97.37
ID 2	22	95.91
ID 3	15	92.11
ID 4	14	93.33
ID 5	18	95.72
ID 6	18	96.67
ID 7	8	97.04
ID 8	10	95.54
Average		95.47

Table 4-8: Performance of the three OCR Engines

The average accuracy of character recognition using tesseract OCR shows 95.47%. While in some extracted text images the acknowledgment of character recognition of names does not show accurate accuracy because there are errors in some letter characters. This is because the train data tesseract is not enough so the tesseract could not recognize all characters of letters.

4.7 Findings and Challenges

This study attempted to show what preprocessing and page layout segmentation algorithms are good for ID card detection that are captured by a E-CAF system. We found a best and easy way to rectify a distorted or skewed document image by a support from the user to select the four vertex of the ID card, using an advantage of the shape and size of the ID card image. And Sauvola thresholding has shown an attractive result on the captured document which face lighting problem, we propose to use this algorithm when we are inputting a document which has different type of texture. For example, if the ID card has a complicated graphics, logo, picture and watermarks.

The challenge is to find and determine the skewness of the text, when we scan a document by a desktop scanner there is no additional surrounding noise to the image. But when we scan it by the E-CAF there will be a lot of noise and external image found in the documents which are the surroundings of the ID card. So finding and determining the skewness of the text is challenging. That is why we went to use a document boundary rectification. However, correcting the ID card based on its boundary is also challenging because of finding the bounding rectangle of the ID card. Based on this challenge we propose to use a user support for the rectification of the text skewness.

The other challenge is finding the candidate text region from the image. The proposed technique may find the correct text region as a non-text region (True Negative) and find a correct non-text region as a text region (False Positive). The selected non- text region may be eliminated in the segmentation or post-processing stage but the text regions that are selected as non-text region will be ignored and missed.

It is also observed that, there are various designs and different ID Templates with several fonts, designs, and templates for different types of ID cards. Characters of various fonts have large within-class variations and form many pattern sub-spaces, making it difficult to perform accurate recognition when the character class number is large. Sometimes few ID cards are printed in different languages. In multilingual situations, capturing information from the scanned documents stays as a primary research issue since information in complex symbolism is more troublesome.

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

In this thesis an attempt is made to automatically detect and recognize text from identity card of customers which is a gap observed in obtaining information of a customer to perform registration process. For this purpose effective page segmentation technique developed, that can identify text and non-text blocks and applying text recognition using OCR. In digitizing information, customers' identity card text recognition plays a great role specially in business areas and government offices in daily customer acquisition activities. It reduces costs, save time, and improves the user experience, thereby providing accelerated customer onboarding processes with an improved data accuracy.

5.1 Conclusions

The main objective of this study is detecting and identifying Amharic text from identity card of customers by applying effective page segmentation that can identify text and non-text blocks from ID cards. Towards achieving this goal, effective page layout segmentation is performed to detect and identify object information that is captured from the ID cards.

The proposed approach is as follows, first image pre-processing techniques skew, and perspective correction method implemented to made collected document images ready for processing. Then, binarization methods namely Otsu global thresholding, Sauvola local adaptive thresholding and Niblack local adaptive thresholding used to solve lightning issues. Based on the experiment Sauvola's method worked better and faster (PSNR-54.05). The second process is segmentation. This is done by applying page layout segmentation techniques, morphological dilation and connected component (CC) for separating graphics from text area and segmenting text line areas. Finally, the proposed approach performance will be evaluated based on the precision and the recall rate. For document images that contain small amount of noise the performance of the system without skew correction shows 90.87% precision and 98.40% recall. After the proposed skew and perspective rectification applied a 93.6 % precision and 100% recall were registered.

The major challenges that the proposed approach faces were finding and determining the skewness of the text and correcting the ID card based on its boundary is challenging and existence of many font types which should be considered in the training phase, so the training phase will be very complex.

5.2 Recommendations

This study tried to detect and identify Amharic text from ID cards of customers. Based on the finding of the study, the following recommendations are formulated to future research direction.

- Customer ID cards have different physical and logical layouts such as complicated graphics, logos, pictures, etc. Hence, future studies can explore on different layouts of real-life ID cards.
- The proposed study used google tesseract OCR for Amharic ID card documents recognition; However, the recognition accuracy depends on the quality of ID cards; future research also needs to explore a better recognition algorithm in the course of developing applicable Amharic OCR
- The study focuses on determining and identifying sample attributes. Therefore, to determine the overall layout of every scanned ID card, extracting a sense of the format as well as the content of every scanned ID card and to keep the output in the database needs further research to be conducted.

Reference

[1] A. K. Boyat and B. K. Joshi, "A Review Paper : Noise Models in Digital Image Processing," *Signal Image Process. An Int. J.*, vol. 6, no. 2, pp. 63–75, 2015, doi: 10.5121/sipij.2015.6206.

[2] Rafael C. Gonzalez& Richard E. Woods, "Image Processing and Related Fields," 2008.

[3] S. Padmappriya and K. Sumalatha, "Digital Image Processing Real Time Applications," *Int. J. Eng. Sci. Invent.*, vol. 4, pp. 46–51, 2018.

[4] . U. K. and M. Vanitha, "A Study on Text Recognition using Image Processing with Datamining Techniques," *Int. J. Comput. Sci. Eng.*, vol. 7, no. 2, pp. 783–787, 2019, doi: 10.26438/ijcse/v7i2.783787.

[5] B. T. Negussie, "Citizen Identification System: the Case for Ethiopia," 2007.

[6] A. R. Khan, "National Identity Card: Opportunities and Threats," *J. Asian Res.*, vol. 2, no. 2, p. 77, 2018, doi: 10.22158/jar.v2n2p77.

[7] R. Clarke, "Human Identification in Information Systems: Management Challenges and Public Policy Issues," *Inf. Technol. People*, vol. 7, no. 4, pp. 6–37, 1994, doi: 10.1108/09593849410076799.

[8] X. Fang, X. Fu, and X. Xu, "ID card identification system based on image recognition," *Proc. 2017 12th IEEE Conf. Ind. Electron. Appl. ICIEA 2017*, vol. 2018-Febru, no. 3, pp. 1488–1492, 2018, doi: 10.1109/ICIEA.2017.8283074.

[9] Ethiotelecom, "https://www.ethiotelecom.et," 2021.

[10] R. Akhter, H. Bhuiyan, and M. S. Uddin, "Extraction of Words from the National ID Cards for Automated Recognition," no. October, 2011, doi: 10.1117/12.913478.

[11] ethiotelecom, "Customer Filling Agent.pdf."

[12] X. Wang, X. Ding, and C. Liu, "Gabor filters-based feature extraction for character recognition," vol. 38, pp. 369–379, 2005, doi: 10.1016/j.patcog.2004.08.004.

[13] V. J. Dongre, G. Polytechnic, V. H. Mankar, and G. P. Nagpur, "Devnagari document segmentation using histogram approach," no. August 2011, 2014, doi: 10.5121/ijcseit.2011.1305.

[14] M. Ryan and N. Hanafiah, "An Examination of Character Recognition on ID card using Template Matching Approach," vol. 59, no. Iccsci, pp. 520–529, 2015, doi: 10.1016/j.procs.2015.07.534.

[15] W. Satyawan, M. O. Pratama, R. Jannati, G. Muhammad, and B. Fajar, "Citizen Id Card Detection using Image Processing and Optical Character Recognition," *Wira Satyawan al 2019 J. Phys. Conf. Ser. 1235 01204*, 2019, doi: 10.1088/1742-6596/1235/1/012049.

[16] K. J. Kuipers and S. J. Hysom, *Common Problems and Solutions in Experiments*, Second Edi. Elsevier Inc., 2014. doi: 10.1016/B978-0-12-404681-8.00007-8.

[17] U. I. Basics, "Understanding Image Basics." https://nanonets.com/image-recognition

[18] T. K. Nguyen, "A Method for Segmentation of Vietnamese Identification Card Text Fields," no. October, 2019.

[19] O. C. Recognition, O. C. Reader, and I. C. Recognition, "Optical Character Recognition (OCR) Technology Overview."

[20] S. A. Butscher, "What is a customer club?," *Cust. Loyal. Program. Clubs*, pp. 14–19, 2019, doi: 10.4324/9781315575629-4.

[21] A. Ababa, "Ethiopia : Appearance of identity cards and birth certificates issued in Addis Ababa , including description of features , variation in appearance , and issuing authorities (2012-2014)," 2021. [Online]. Available: https://www.refworld.org/docid/54bf5d616.html

[22] Deeksha Aggarwal, "Image Segmentation Techniques using Digital Image Processing, Machine

Learning and Deep Learning Methods.," *Https://Medium.Com/Analytics-Vidhya/Image-Segmentation-Techniques-Using-Digital-Image-Processing-Machine-Learning-and-Deep-Learning-Ccf9E4589E94*, vol. 1, no. Part 1, pp. 1–11, 2020.

[23] R. Articles, "Digital Image Processing Basics," pp. 1–7, 2021.

[24] S. Joseph, "Image Processing Techniques and Its Applications : an Overview," *Ijariie*, vol. 4, no. 3, pp. 2168–2174, 2018, [Online]. Available: http://ijariie.com/AdminUploadPdf/IMAGE_PROCESSING_TECHNIQUES_AND_ITS_APPLICATIONS__AN_OVERVIEW_ijariie8745.pdf

[25] V. K. Mishra, S. Kumar, and N. Shukla, "Image Acquisition and Techniques to Perform Image Acquisition," *SAMRIDDHI A J. Phys. Sci. Eng. Technol.*, vol. 9, no. 01, pp. 21–24, 2017, doi: 10.18090/samriddhi.v9i01.8333.

[26] H. K. Ranota and P. Kaur, "Review and Analysis of Image Enhancement Techniques," *Int. J. Inf. Comput. Technol.*, vol. 4, no. 6, pp. 583–590, 2014, [Online]. Available: http://www.irphouse.com

[27] E. H. Barney Smith, L. Likforman-Sulem, and J. Darbon, "Effect of pre-processing on binarization," *Doc. Recognit. Retr. XVII*, vol. 7534, no. January, p. 75340H, 2010, doi: 10.1117/12.840606.

[28] S. G. K. Patro and K. K. sahu, "Normalization: A Preprocessing Stage," *Iarjset*, no. March, pp. 20–22, 2015, doi: 10.17148/iarjset.2015.2305.

[29] M. S. Yasein and P. Agathoklis, "An image normalization technique based on geometric properties of image feature points," *ISSPIT 2007 - 2007 IEEE Int. Symp. Signal Process. Inf. Technol.*, pp. 116–121, 2007, doi: 10.1109/ISSPIT.2007.4458129.

[30] Changming Sun and Deyi Si, "Skew and slant correction for document images using gradient direction," vol. 2, pp. 142–146, 2002, doi: 10.1109/icdar.1997.619830.

[31] S. Sajeev and O. T. Lee, "Text_Line_Detection_Using_Connected_Components," vol. 7, no. 3, pp. 3–8, 2019.

[32] A. Choudhary, "A Review of Various Character Segmentation Techniques," vol. 4, no. 6, pp. 559–564, 2014.

[33] V. Bolón-Canedo and B. Remeseiro, "Feature selection in image analysis: a survey," *Artif. Intell. Rev.*, vol. 53, no. 4, pp. 2905–2931, 2020, doi: 10.1007/s10462-019-09750-3.

[34] T. W. Ryan, "Image Segmentation Algorithms," *Archit. Algorithms Digit. Image Process. II*, vol. 0534, p. 172, 1985, doi: 10.1117/12.946577.

[35] C. R. Kulkarni and A. B. Barbadekar, "Text Detection and Recognition : A Review," pp. 179–185, 2017.

[36] K. Wang and S. Belongie, "Word spotting in the wild," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6311 LNCS, no. PART 1, pp. 591–604, 2010, doi: 10.1007/978-3-642-15549-9_43.

[37] M. Mulchandani, P. Dudhe, and R. Patil, "Implementing Object Detection System in Image," vol. 13, no. 2, pp. 1235–1239, 2020.

[38] P. Vashist and K. Hema, "Character Recognition with Minimum Edit Distance Method," vol. 2, no. 4, pp. 457–462, 2013.

[39] G. A. Nagargoje and A. L. Wanare, "An innovative approach towards Virtual Pen with real-time OCR feature," *Int. J. Res. Adv. Eng.*, vol. 1, no. 4, p. 1, 2015, doi: 10.26472/ijrae.v1i4.34.

[40] A. LETTA, "FEATURE EXTRACTION AND MATCHING IN AMHARIC DOCUMENT IMAGE COLLECTIONS ADANE LETTA," 2011.

[41] N. Features, "Ge ' ez script," 2021. http://www.omniglot.com/writing.html
[42] B. W. Agajyelew, "Information Extraction from Amharic language Text : Knowledge," 2015.
[43] W. Kelemework, "Automatic Amharic text news classification: Aneural networks approach," *Ethiop. J. Sci. Technol.*, vol. 6, no. 2, pp. 127–137, 2013.
[44] M. Meshesha and C. V Jawahar, "Optical Character Recognition of Amharic Documents," *African J. Inf. Commun. Technol.*, vol. 3, no. 2, 2007, doi: 10.5130/ajict.v3i2.543.
[45] R. Valiente, M. T. Sadaike, J. C. Gutiérrez, D. F. Soriano, G. Bressan, and V. Wilson, "A process for text recognition of generic identification documents over cloud computing," *142 Int'l Conf. IP, Comp. Vision, Pattern Recognit. | IPCV'16 |*, vol. 16, no. 1, pp. 142–146.
[46] K. Abebe, R. S. R, P. Y. K. Chung, and D. Zerihum, "Vehicle Plate Recognition for Ethiopian License Plate : based Sift Feature," vol. 8, no. 06, pp. 568–570, 2019.
[47] Villaverde and Martin, "Automatic Reading of Spanish Identity Cards".
[48] A. Kumar and A. Tiwari, "A Comparative Study of Otsu Thresholding and K-means Algorithm of Image Segmentation," *Int. J. Eng. Tech. Res.*, vol. 9, no. 5, pp. 12–14, 2019, doi: 10.31873/ijetr.9.5.2019.62.
[49] T. R. Singh, S. Roy, O. I. Singh, T. Sinam, and K. M. Singh, "A New Local Adaptive Thresholding Technique in Binarization," vol. 8, no. 6, pp. 271–277, 2012, [Online]. Available: http://arxiv.org/abs/1201.5227
[50] W. J. Teahan, P. Wu, and W. Liu, "Adaptive compression-based models of Chinese text," *ICALIP 2014 - 2014 Int. Conf. Audio, Lang. Image Process. Proc.*, no. July, pp. 874–881, 2015, doi: 10.1109/ICALIP.2014.7009920.
[51] M. Javed, P. Nagabhushan, and B. B. Chaudhuri, "Extraction of line-word-character segments directly from run-length compressed printed text-documents," *2013 4th Natl. Conf. Comput. Vision, Pattern Recognition, Image Process. Graph. NCVPRIPG 2013*, no. March, 2013, doi: 10.1109/NCVPRIPG.2013.6776195.
[52] C. Science and M. Meshesha, "Recognition and Retrieval from Document Image Collections Thesis submitted in partial fulfillment of the requirements for the degree of International Institute of Information Technology," *Technology*, no. August, 2008.
[53] T. Randen and H. John, "Segmentation of text / image documents using texture approaches 1 Introduction," *Int. Conf. Acoust. Speech Signal Process.*, pp. 60–67, 1999.
[54] H. J. He, C. Zheng, and D. W. Sun, "Image Segmentation Techniques," *Comput. Vis. Technol. Food Qual. Eval. Second Ed.*, no. December, pp. 45–63, 2016, doi: 10.1016/B978-0-12-802232-0.00002-5.
[55] S. Ruparelia, "Implementation of Watershed Based Image Segmentation Algorithm in FPGA," Stuttgart University, 2012.
[56] F. Shafait, D. Keysers, and T. M. Breuel, "Performance comparison of six algorithms for page segmentation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3872 LNCS, no. July 2014, pp. 368–379, 2006, doi: 10.1007/11669487_33.
[57] S. V Khangar and L. G. Malik, "Compression Method for Handwritten Document Images in Devnagri Script," vol. 3, no. 3, pp. 4305–4309, 2012.
[58] S. Mandal, S. P. Chowdhury, A. K. Das, and B. Chanda, "Automated detection and segmentation of table of contents page from document images," *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR*, vol. 2003-Janua, no. August 2014, pp. 398–402, 2003, doi: 10.1109/ICDAR.2003.1227697.

[59] S. S. Kumar, P. Rajendran, P. Prabaharan, and K. P. Soman, “Text/Image Region Separation for Document Layout Detection of Old Document Images Using Non-linear Diffusion and Level Set,” *Procedia Comput. Sci.*, vol. 93, no. September, pp. 469–477, 2016, doi: 10.1016/j.procs.2016.07.235.