UNIVERSITY OF NAIROBI

SCHOOL OF ENGINEERING

ENHANCEMENT OF LOW RESOLUTION MOBILE PHONE IMAGES FOR IMPROVED INTERPRETATION OF VISUAL INFORMATION

By

Kiragu Henry Macharia

B Tech. in Electrical and Communications Engineering

Moi University

A thesis submitted in partial fulfillment of the degree of Master of Science in Electrical and Electronic Engineering in the department of Electrical and Information Engineering in the University of Nairobi

23rd July 2013

1

DECLARATION AND APPROVAL

Declaration

This thesis is my original work and has not been presented for a degree in any other university.

Name: Kiragu Henry Macharia

Registration number: F56/71960/08

Sign:…………………………………..Date:………………………………

Approval

This thesis has been presented for examination with my approval as the university supervisor.

*S*upervisor: Professor Elijah Mwangi

Department of Electrical and Information Engineering

University of Nairobi.

Sign:………………………………….Date:………………………………

ACKNOWLEDGEMENT

ABSTRACT

The enhancement of digital images is necessary in many fields of application of images. It makes it easier for human viewers and machines to extract, interpret, and where necessary perform further processing on the information contained in the images.

In this thesis, a spatial-spectral enhancement solution is proposed for degraded images captured using a cell phone camera. The degradations addressed here are those caused by factors such as poor resolution, fading, noise, shadows, poor illumination, geometrical distortions and ink smears which may lead to erroneous interpretation of the image information.

A Nokia model 1680 cell phone camera was used to capture the images which were then loaded to a computer using the Nokia PC suite software version 7.1 and converted to 8-bit grey-scale images using MATLAB version 7.14. The greyscale images were then interpolated using a bicubic method to a size of 480x640 pixels. The interpolated images were enhanced by contrast stretching followed by single-scale retinex enhancement. Optimal global thresholding was then used to binarise the images and finally, morphological dilation was performed to yield enhanced images. For geometrically distorted document images, an additional procedure based on 2-D processing was carried out on the morphologically dilated images to rectify the distortions.

Computer simulation based experiments have been used to demonstrate the effectiveness of the proposed algorithm. The results show a significant improvement over those obtained using other proposed methods in terms quality measures and performance of optical character recognition applications.

TABLE OF CONTENTS page

6

LIST OF FIGURES

page

LIST OF ACRONYMS

CCD   Charge Coupled Device

CFA   Colour Filter Array

CMOS Complementary Metal Oxide Semiconductor

CORR Correlation

dB   Decibel

DCT   Discrete Cosine Transform

DFT   Discrete Fourier Transform

DWT   Discrete Wavelet Transform

JPEG   Joint Photographic Experts Group

LCD   Liquid Crystal Display

LSE   Least Square Error

MRF   Markov Random Field

MI   Mutual Information

MLV	Mean of Least Variance
MWIS	Multiple Windowed Inverse Sigmoid
NLSE	Normalised Least Square Error
OCR	Optical Character Recognition
PC	Personal Computer
pdf	Probability Density Function
PSNR	Peak Signal to Noise Ratio
RAM	Random Access Memory
RLSA	Run Length Smoothing Algorithm
RMSE	Root Mean Square Error
SE	Structuring Element
SSR	Single Scale Retinex
USB	Universal Serial Bus
VSB	Vertical Stroke Boundaries

LIST OF TABLES

Page

CHAPTER 1

INTRODUCTION

1.1. Background

An image is a two dimensional function $f(x, y)$ of spatial coordinates. The value of $f(x, y)$ is the grey level of the image at a point represented by the spatial coordinates (x, y). When x, y and the grey levels are finite discrete quantities, $f(x, y)$ is a digital image. A digital image consists of a finite number of elements called pixels which have specific values and occur at specific positions [1, 2, 3]. The values of $f(x, y)$ for a digital image are in the range $0 \leq f(x, y) \leq K$ where K is an upper bound value.

1.2 Image Generation Methods

For an image to be generated, energy from a suitable source is made to interact with the scene to be imaged. The portion of the energy transmitted or reflected from the scene is sensed by a suitable transducer which produces the image signal. The sources of energy used in image generation are: Electromagnetic waves, sonar, ultrasound and electron beams. Synthetic images or fractals can also be generated by use of computers [1, 2].

The electromagnetic spectrum is partitioned into the following bands of frequencies: Gamma rays, X-rays, Ultraviolet, Visible light, Infrared, Microwaves and radio waves. Any of these bands can be used to generate images for different applications [1, 2].

Gamma rays imaging is applied in nuclear medicine and astronomy whereas X-rays imaging is mainly used in medical diagnostics and industrial inspection of products. Ultraviolet imaging is used mainly in lithography, industrial inspection and microscopy. Imaging in the visible light band has applications in many fields such as entertainment and

multimedia communication. Imaging in the infrared band is used in security and wildlife monitoring. Imaging in the microwave band is mainly used in radar systems and meteorology. Imaging in the radio band is mainly used in astronomy and medical fields [1].

Sonar imaging utilises the intensity and time delay of reflected sound waves (echo) to detect objects (targets). The sonar image provides information about the size, shape and reflection characteristics of the target. It is mainly applied in routine inspection of underwater areas and structures such as pipelines and submarine cables. The inspection is done for security and maintenance reasons [2].

Other imaging techniques include ultrasound imaging which is mainly used in geological and medical fields as well as imaging using a beam of electrons which is used in electron microscopy [1].

1.3 Photographic Image Formation

During digital camera photographing, light is reflected by the objects in the imaging scene, collected by a series of lenses and focused onto the image plane. The image plane consists of a matrix of closely spaced light sensors each corresponding to a single image pixel. Most digital cameras use charge coupled device (CCD) sensors although some use CMOS based sensors. A CCD sensor consists of a rectangular grid of electron collection sites laid over a silicon wafer to record the amount of light energy reaching each site. When light from a scene strikes these sensor sites, electron-hole pairs are generated. The electrons at each site are then collected after a certain period of time and converted to pixel values [4].

Each type of sensor has a specific spectral response which is a function of wavelength of the illumination used, the optical blur and the spatial integration at each site of the sensor.

When a digital image is recorded, the camera performs several processes on the image. These processes are such as correction for sensor nonlinearities and nonuniformities, white balance adjustment, image compression and colour filter array (CFA) interpolation.

A color image requires at three colour samples at each pixel location. A colour camera would therefore need three separate sensors for the primary colours red (R), green (G), and blue (B). Each of these sensors would also require its own driving electronics and all of them have to be precisely registered. Such an imaging system would be quite expensive.

In order to reduce the expense of a digital camera, a single colour sensor covered with a CFA which allows only one of the three colour samples to be measured at each pixel location is used. The missing two color values at each pixel are then estimated by the camera using an interpolation process termed demosaicking. The most commonly used demosaicking filter array is the Bayer CFA. It measures the G component of the colour image and then estimates the R and B components by interpolation [4].

1.4 Image Pre-processing

The pre-processing of the generated or acquired image is performed so as to produce an image signal that is suitable as an input for the selected digital image enhancement technique. Pre-processing methods include: geometric transformation, scanning, denoising and Restorative pre-processing. Spatial geometric transformations include: zooming, shearing, shrinking, rotation and flipping. Restorative pre-processing involves noise filtering, text warping rectification, shadow and other degradations removal [2]. When an image is acquired as a hardcopy, it is scanned so as to transfer it into a digital computer in form of a two-dimensional digital signal.

1.5. Problem Statement

Interpretation of the information contained in images is a complex task since an image is a complex interaction of many factors. These factors are such as illumination, interaction of the radiation with matter, geometry of projection of the imaging energy onto a plane, characteristics of the image sensors, camera lens imperfections and dark current. Most images have important details of information obscured due factors such as poor illumination, low resolution, noise, fading and geometrical distortions [1]. None of the existing image enhancement methods can remove the effects of all these factors. This problem requires development of algorithms that can improve the quality of images degraded by a combination of the aforementioned factors. In this research, a general purpose image enhancement procedure has been proposed for enhancing such images.

1.6. Objectives

The main objective of this research was to formulate an image enhancement procedure for improving the visual quality of degraded images captured using a cell phone camera to give high quality output images that are comparable to those taken using a higher resolution digital camera such as the Fujifilm model 4800Z.

The specific objectives were:
   i.   To reveal important details of information hidden in the digital images.
   ii. To reduce the effects of noise and other degradations on the quality of images.
   iii. To reduce the effects of geometric distortions in text document images.

1.7. Scope of the Study

This thesis focused on degraded greyscale images captured using a Nokia 1680 cellphone camera for which suitable enhancement solutions have been proposed. The images used were captured in the visible light band of the electromagnetic spectrum. Enhancement of geometrically distorted document images based on two dimension image processing was also covered in the study. The geometric distortions covered in this thesis are those that arise from the orientations of the imaged objects but not due to camera lens imperfections.

1.8 Organisation of the Thesis

This first chapter gives a brief introduction to digital images, image generation methods and formation of photographic images by digital cameras. A statement of the problem that has been addressed in this research, the objectives as well as the scope of the research are also presented in this chapter.

Chapter two covers some of the literature on digital image enhancement techniques published by researchers in recent years. Recent works that are related to this study have also been highlighted in the chapter.

In chapter three, some of the published theory in spatial domain image enhancement methods is presented. Morphological image processing, the single scale retinex algorithm, image thresholding as well as theory on enhancement by noise and degradation reduction methods are included in the chapter.

Chapter four covers the theory of transform domain enhancement techniques. Two-dimensional Discrete Fourier Transform (DFT), Discrete Wavelet Transform (DWT) and the Hough transform are presented.

Chapter five covers the theory of enhancement techniques for geometrically distorted document images. Only the enhancement based on two dimensional image processing is covered here.

The sixth chapter presents the materials and methods that were used in carrying out the research. The proposed enhancement method is presented in the chapter.

Computer simulation experimental results obtained during the research are presented in chapter seven. The simulations were done in MATLAB.

Conclusions are drawn in chapter eight which also contains recommendations for further work. A list of published reference documents and appendices are included at the end of this report.

CHAPTER 2

LITERATURE REVIEW

Digital image enhancement is a process of improving the perception of the information in images by human viewers. The objective of digital image enhancement is to remove unwanted signal components corrupting the image in order to improve the visual appearance and intelligibility of the image for a given purpose [1, 2, 3]. This has applications in various fields such as in the entertainment industry, medical imaging, surveillance, astronomy, robotics, business industry, traffic, military, geology, meteorology, industrial inspection and the aviation industry. Various techniques are used in image enhancement such as contrast and dynamic range enhancement, image thresholding, edge sharpening, image interpolation, false colouring, image denoising and degradation reduction. In addition, image enhancement techniques are employed as the first stage in processes such as image restoration, recognition and compression.

Image restoration refers to the techniques that improve the image appearance based on mathematical and probabilistic models of the image degradation and noise. Unlike enhancement whose outputs are tested subjectively, testing of image restoration results are objective [5].

Restoration attempts to recover the image that has been degraded by using prior knowledge of the degradation phenomenon. Degradations can be introduced by the image sensor, transmission medium or the display system. Like enhancement, restoration techniques can be formulated in either spatial domain or frequency domain [5, 6].

2.1 Contrast and Dynamic Range Enhancement

Contrast stretching is a process that expands the range of intensity levels in an image so that it spans the full intensity range of the recording medium or display device [7]. For an 8-bit greyscale image $f(x,y)$, a linear contrast stretching operation can be represented by:

$$g(x,y) = \frac{255}{(f_{max} - f_{min})}(f(x,y) - f_{min})$$

(2.1)

for $0 \leq f_{min} < f_{max} \leq 255$.

Where f_{max} and f_{min} are the maximum and minimum intensity levels in the image $f(x,y)$ respectively and $g(x,y)$ is the contrast stretched image [7].

The enhancement methods employ techniques such as Wiener filtering, contrast stretching, Multiple Windowed Inverse Sigmoid (MWIS), Single Scale Retinex (SSR) and homomorphic filtering.

Gatos et al. [8] have proposed an image pre-processing method that uses an adaptive low-pass Wiener filter in the binarisation of degraded document images. The filter serves the functions of eliminating noisy areas, smoothing of background texture as well as contrast enhancement between background and text areas. Use of the filter proved efficient for all the above goals. The adaptive Wiener filter used is based on statistics estimated from a neighbourhood around each pixel. The greyscale source image $I_s(x,y)$ is transformed to the filtered greyscale image $I(x,y)$ according to equation (2.2).

$$I(x,y) = \frac{\mu + (\sigma^2 - v^2)(I_s(x,y) - \mu)}{\sigma^2}$$

(2.2)

where μ is the local mean, σ^2 is the variance at a $N \times N$ neighbourhood around each pixel and v^2 is the average of all estimated variances for each pixel in the neighbourhood. Substantial enhancements are obtained for values of N=3 and N=5.

Mokhtar et al. [9] have proposed several contrast enhancement techniques for noisy and blurred blood sample images to determine the levels of white blood cells in the blood. Abnormal levels of white blood cells suggest the possibility of leukaemia and requires further diagnostic tests. Blurring and effects of unwanted noise on blood cell images usually result in false diagnosis. The contrast enhancement techniques proposed are local contrast stretching, global contrast stretching, partial contrast stretching, bright and dark contrast stretching. Comparison of all the proposed techniques was carried out to find the best technique in enhancing acute leukaemia images. Experimental results show that the partial contrast stretching is the best technique for such images.

Both linear and nonlinear contrast stretching algorithms are employed in enhancement of medical images to improve their quality for further analysis by pathologists [10].

Bin et al. [11] have proposed an enhancement method that combines the fuzzy and retinex theories for enhancing vein images degraded by low contrast, narrow dynamic range and asymmetrical illumination intensity distribution. A two-parameter optimization of the fuzzy logic membership function is proposed in which an adaptive approach is used to select the parameters. Experimental results show that this method can greatly enhance the contrast between the vein patterns and its surrounding areas.

A nonlinear image enhancement algorithm named Multiple Windowed Inverse Sigmoid (MWIS) has been proposed by Ucan and Oguslu [12] for enhancing images captured in extremely non-uniform lighting environments. The proposed algorithm is capable of compressing bright regions and at the same time enhancing dark regions by preserving the main structure of the illuminance-reflectance modality. Experimental

results show that the proposed algorithm yields visually optimal results for images captured under extreme lighting conditions.

Mei et al. [13] propose an edge detection algorithm based on homomorphic filtering. The procedure starts with homomorphic filtering followed by a logarithmic transformation of the resulting image. The discrete Fourier transformation is then performed from which the low frequency components are fused with the illumination component of the image while the high frequency components are fused with the reflectance component. Results show both contrast and edge enhancements.

2. 2 Image Thresholding

Image thresholding is a statistical-decision theory problem whose objective is to assign pixels to two or more classes [7, 14]. Thresholding is used as a preprocessing stage in many image enhancement procedures [7]. Some of the published literature employing both global and local thresholding techniques are highlighted in this section.

Otsu [14] proposes a non-parametric optimal global thresholding method for image segmentation. An optimal threshold is selected using the discriminant criterion that maximises the separation between the two classes of grey levels. The procedure utilises only the zeroth and the first-order cumulative moments of the histogram of the image. Experimental results presented show that the technique works well for images having uniform background regions.

Gatos et al. [15] propose an adaptive image binarisation method that uses a combination of multiple binarisation techniques and the edge information of the greyscale source image. An enhancement step based on mathematical morphology operations is incorporated in order to produce a high quality result while preserving stroke information. Subjective performance evaluation on some degraded handwritten and printed documents demonstrate superior performance of the technique

against other techniques such as the Otsu's global method and the Niblack's adaptive method.

Bukhari et al. [16] propose an adaptive binarisation technique for degraded hand-held camera-captured document images. The method uses two different sets of parameters values to select the threshold levels for the foreground and background regions of an image. This overcomes the effects of distortions like non-uniform illumination, bad shading, blurring, smearing and low resolution on the output image. The detection of ridges is used in rough estimation of foreground regions in a document image. The ridges information is then used to calculate the different set of parameters values for the foreground and background regions. The evaluation of the method using an OCR-based measure shows that the method achieves better performance as compared to state-of-the-art global and local binarisation methods such as the Otsu's method.

2.3 Enhancement by Noise and Degradation Reduction

This section presents some of the published methods that have been proposed for noise and degradation reduction. These enhancement methods include denoising, deblurring and geometrical distortion reduction techniques.

Gatos et al. [15] have used a shrink filter to remove noise from the background of an image. The entire binary image is scanned and each foreground pixel is examined. If P_{sh} is the number of background pixels in an $N \times N$ sliding window which has a foreground pixel as the central pixel, then this foreground pixel is changed to background if $P_{sh} > k_{sh}$ where K_{sh} is a constant defined experimentally.

Wu and Tang [17] have proposed a random-valued impulse noise removal method based on controlling functions. The image pixels are first classified as edge pixels, noisy pixels, and interior pixels. A controlling

speed function and a controlling fidelity function are then defined. According to the two controlling functions, the diffusion and fidelity process at the edge pixels, noisy pixels, and interior pixels are selectively carried out. Next, a class of second-order improved and edge-preserving partial differential equation denoising models based on the two controlling functions are determined. Performance of the method is demonstrated using test images corrupted by random-valued impulse noise with various noise levels. The results show better performance of the method than other related denoising models such as the Perona-Malik method.

Gatos et al. [18] propose a recovery technique for arbitrarily warped document images captured by a digital camera or a scanner. The technique recovers the documents based on text lines and words detection .The recovery starts with a draft binary image dewarping step that is based on word rotation and translation according to the upper and lower word baselines. The recovery of the original image is then guided by the draft binary image de-warping result. Experimental results on arbitrarily warped documents image show effectiveness of the technique.

Masalovitch and Mestetskiy [19] have proposed a document images dewarping method based on usage of continuous skeletal image representation. The method utilizes an approximation of the deformation of interlinear spaces of image based on elements of image's skeleton that lies between the text lines. A method for approximation of the whole image deformation as a combination of single interlinear space deformations in the form of two-dimensional cubic Bezier splines [20] is also proposed. Experimental results for several deformed document images prove that efficiency of optical character recognition increases after de-warping process.

Stamatopolous et al. [21] propose a two-step dewarping technique for camera captured document images that improve both the document

legibility and the accuracy of OCR results. The first step involves coarse dewarping which is accomplished with the help of a transformation model that maps the projection of a curved surface onto a 2D rectangular area. The projection of the curved surface is defined by the two curved lines which fit the top and bottom text lines together with the two straight lines which fit to the left and right text boundaries. The second fine dewarping step is achieved based on words detection. Positions and orientations of all the words are then normalised based on detected lower and upper word baselines. Experimental results obtained with several document images have demonstrated the effectiveness of the method.

Lavialle et al. [22] have proposed a method based on straightening of the distorted text lines by fitting a model to each text line. The approach is not efficient if the line spacing is not uniform.

A method that uses an analytical model with cubic B-splines has been proposed by Wu and Agam [23]. This technique requires that the user should interactively specify four corner points of the geometrically distorted image. It cannot be used to enhance document images that have non-uniform columns.

L. Zhang and Tan [24] have proposed a method that uses a non-linear curve for each text line to approximate the text base lines. The approach assumes that the book spine is found along iso-parametric lines which is not always the case.

A method that uses modeling of the text lines using natural cubic splines has been suggested by Ezaki et al. [25]. The method uses complex calculations that have been noted to be computationally expensive.

Mischke and Luther [26] have proposed an enhancement method that is based on approximation of each distorted text line using a polynomial. The method requires a pre-processing step to correct the

skew of the warped document. It is also hard to generalise since it is formulated for correction of a fixed type of warping.

Ulges et al. [27] have proposed a technique that uses estimation of a quadrilateral cell for each letter based on local baselines followed by mapping onto a rectangular plane of correct size and position in the dewarped image. This method is not generic since it assumes that the document image contains only straight lines that are approximately equally spaced and sized.

The technique proposed by Lu et al. [28] is based on restoring images of documents that contain Latin characters by dividing them into multiple quadrilateral patches based on the exploitation of the vertical stroke boundaries (VSBs) and the text baselines. Use of this method is limited to documents that contain Latin characters only.

Schneider et al. [29] have proposed a method that uses local orientation features extracted by text baselines to interpolate a vector field from which a warping mesh is derived. The problem with this approach is that it is not easy to define the characteristic points of transitions so that consistent approximation of baselines is achieved.

Bukhari et al. [30] have proposed a method for mapping characters over each curled baseline pair (upper and lower) to its corresponding straight baseline pair. The mapping is not accurate if the image is heavily distorted.

A method that involves modeling of the image surface as a cylinder followed by a transformation to flatten the document image has been proposed by Fu et al. [31]. The drawback of this method is that it requires complex computation. The assumption that a single cylinder can fit to a deformed page is also not generic.

Y. Zhang et al. [32] have suggested a method based on performing a rough text line and character segmentation to estimate the warping

direction. Interpolation using Splines is then used to restore the image. Text line and character segmentation is performed using projections at the original warped document this causes many segmentation errors.

2.4 Edge Sharpening

Image sharpening results from enhancement of the high frequency components of the image. Sharpening is achieved by differentiation of the image which is carried out by means of filters in either transform or spatial domains. The effect of image sharpening is to highlight fine details such as points, lines and edges. Edge sharpening is particularly important in enhancing boundaries of objects in blurred images.

Schulze [33] uses a non-linear Mean of Least Variance (MLV) filter to sharpen biomedical images. The results make it easier to distinguish between the various parts of the images leading to easier diagnosis.

Guoxin et al. [34] have used the single scale retinex algorithm based on the wavelet theory to achieve sharpening of images captured using cameras under conditions of insufficient light. The resulting images are sharper than the inputs and also are of high fidelity.

2.5 Image Interpolation

Image interpolation is the process of using known intensity levels of an image to estimate unknown intensities during image resizing [1]. The three types of image Interpolation methods that are commonly used in image enhancement are the nearest neighbour, bilinear and bicubic interpolations. In nearest neighbour interpolation, each new location is assigned the intensity level of its nearest neighbour in the original image. For bilinear interpolation, the intensities of the four nearest neighbours in the original image are used to estimate the intensity at a given location in the resampled image. Bicubic interpolation is the standard used in commercial image editing programs due to its modest complexity and

distortion. The method uses the nearest sixteen neighbours to a point to determine the intensity value v to be assigned to the point (x, y) in the resampled image as shown in equation (2.3)

$$v = \sum_{i=0}^{3} \sum_{j=0}^{3} a_{ij} x^{i} y^{j}.$$ (2.3)

Where a_{ij} are coefficients obtained by solving the sixteen equations written for the sixteen nearest neighbours of point (x, y) [1].

Gatos et al [15] have used bicubic interpolation to enlarge images since it produces better image upsampling results than the other methods.

2.6 False Colour Enhancement

False colour image enhancement is a process that assigns colours to grey scale images following a specified criterion. The process leads to improved human visualisation and interpretation. It is motivated by the fact that humans can discern more colour shades as opposed to grey shades. The two false colouring approaches employed in image processing are intensity slicing and intensity to colour transformation. In intensity slicing, grey levels are partitioned by planes parallel to the black grey level and a different colour assigned to either side of each plane. In intensity to colour transformation, three independent transformations are performed on the grey values of the image. The three results are then combined as colour components of a colour image. This approach achieves a wider range of pseudocolours but it involves more computation steps [5].

Vendhan et al. [10] have proposed a thin-prep monolayer cytology method that uses fuzzy logic and false colour enhancement to overcome the limitations of conventional Pap-smear test in screening of cervical cancer. Experimental results show that the technique improves the image quality and also makes it easier for pathologists to perform further analysis.

2.7 Related Works

Some of the recently published works that are related to the enhancement algorithm proposed in this thesis are summarised in this section. A comparison with the proposed procedure is also made to highlight the drawbacks in these methods that are addressed by the proposed algorithm.

Niblack [35] proposes a method for determining local threshold values using a rectangular neighbourhood mask. The threshold value for the pixel at the coordinates (x, y) is given by equation (2.4),

$$T_{xy} = k_1 \sigma_{xy} + \mu_{xy} \qquad (2.4)$$

where μ_{xy} and σ_{xy} denote the mean and standard deviation of the set of pixels contained in a neighbourhood that is centered at the coordinates (x, y) whereas k_1 is a constant that determines the proportion of the total boundary considered to be a part of the object. The value of k_1 is usually set at 0.2 through experimentation.

The result of binarisation using this method retains background noise when the objects in the image are sparse [8]. The low frequency background noise is suppressed by the method proposed here by the high pass filtering effect of the single scale retinex algorithm.

Sauvola and Pietikainen [36] have proposed an adaptive document image binarisation method where the page is considered as a collection of sub-components such as text, background and pictures. The method addresses the problems caused by noise, illumination and many source related degradations. The method adds a hypothesis on the intensity levels of the object and background pixels resulting in the expression for the threshold level given in equation (2.5).

$$T_{xy} = \mu_{xy}\left[1 + k_2\left(\frac{\sigma_{xy}}{R} - 1\right)\right] \qquad (2.5)$$

where T_{xy} is the threshold level for the pixel at the location (x, y), R is the dynamic range of the standard deviation which is fixed at 128 and k_2 is a positive constant which is usually set at 0.5.

Performance evaluation of the algorithm is performed using test images, ground-truth images, evaluation metrics for binarisation of textual and synthetic images, and a weight-based ranking procedure. The algorithm is also tested with different types of document images and degradations. The method produces better results compared with a number of known techniques in published literature such as the Niblack's, Barnsen's and the Eikvil's local methods. In addition, evaluation of local binarisation methods have reported that the Sauvola binarisation method is better than other types of local binarisation methods [15]. However the method is computationally complex, requires continuous adjustment of parameters and does not perform well on heavily faded images. The method proposed in this thesis overcomes the complexity problem by using global thresholding which involves less computation time and less complex algorithms. The global thresholding method used in the proposed procedure does not require continuous adjustment of parameters. Use of the Single Scale retinex algorithm in the method proposed in this thesis reduces the effects of fading and other low frequency degradations.

Gatos et al. [37] propose a technique for improving the Sauvola method. The procedure starts with pre-processing using a Wiener filter followed by foreground and background regions estimation. Next, adaptive thresholding is performed followed by up-sampling. Finally post-processing using shrinks and swell filtering is performed. The procedure yields good results but it is complex, slow and does not compensate for shadows and warping degradations. All these demerits are addressed in

the method proposed here. By using a global thresholding method, the proposed method is computationally simple and fast. Use of the single scale retinex enhancement results in shadow removal while the inclusion of a geometrical distortion reduction stage in the proposed method results in document image dewarping.

Guoxin, Pei and Qiang [34] have proposed a wavelet based image enhancement algorithm. Although their approach improves the dynamic range compression of the image and also reduces the effect of poor illumination, it does not address the problem of non-uniform illumination and geometrical distortions. The proposed algorithm employs the single scale retinex enhancement to reduce the effects of non-uniform illumination.

Wang and Wang [38] have applied a Retinex based algorithm for detection and compensation of shadows in high resolution images. The method improves the visual quality of the images but does not address the effects of other causes of degradation such as warping and gaps in the characters of document images. The algorithm proposed in this thesis incorporates a geometrical distortion reduction stage to reduce the effects of warping. Morphological dilation is employed in the proposed algorithm in order to preserve stroke connectivity of document image characters by bridging any gaps resulting from the thresholding process.

The algorithm proposed by Bin, Janwen and Xinyan [11] based on fuzzy sets theory assumes that the illumination component of the image is constant. The reflectance component can therefore be obtained by dividing the image function by the constant illumination component. This assumption does not hold for non-uniform illumination environments.

The method proposed in this report models the illumination component of an image as a low frequency two-dimensional signal. This modeling is done by convolving the image with a low pass Gaussian filter

kernel to obtain an estimate for the illumination component. This component is then suppressed by subtracting its logarithm from the logarithm of the input image according to the single scale retinex algorithm. This pre-processing results in an output enhanced image of higher visual quality.

CHAPTER 3

SPATIAL DOMAIN ENHANCEMENT TECHNIQUES

The objective of image enhancement is to make the resulting image better than the original image for a particular application [3]. The enhancement methods can be classified as either spatial domain or frequency domain methods. In spatial domain methods, the image is manipulated directly on the plane containing the pixels. The spatial domain methods can further be classified as either point operations or mask operations.

In point operations, individual pixels of an image are transformed independently according to either a linear or a nonlinear transformation to yield the corresponding individual pixels in the enhanced image. Point operations are such as grey level transformations, arithmetic operations and logic operations. In mask operations, a linear or nonlinear transformation is applied to a block of pixels at a time to produce a single pixel in the enhanced image. Mask spatial domain operations include smoothing and sharpening filtering [1].

Enhancement techniques are problem-oriented in that the best enhancement technique used to enhance one type of image for a given application may not be the best technique to enhance another image for a different application. Enhancement techniques cannot be generalised because the viewer is the ultimate judge as to how well a particular technique works [2].

In this chapter, image enhancement techniques based on spatial domain operations are presented and applied in removal of noise and blur.

3.1 Point Processing Operations

A point spatial domain process can be denoted by the expression given in equation (3.1).

$$g(x,y) = T_1\{f(x,y)\} \tag{3.1}$$

where $f(x,y)$ is the input image, $g(x,y)$ is the output enhanced image and $T_1\{\}$ is an operator applied on a single pixel r of the input image at a time to yield a corresponding single pixel s in the output image. The transformation function for point processing is given in equation (3.2).

$$s = T_1(r) \tag{3.2}$$

Different types of point processing techniques are used in image enhancement. These techniques are such as contrast stretching, grey level transformations, arithmetic and logical operations [2, 3].

Two of the commonly used grey level transformations enhancement methods are linear transformation and logarithmic transformation. Linear transformation includes identity and negative transformations. Logarithmic transformation also includes inverse logarithmic transformation. [1, 6].

3.1.1 Linear Transformation

The linear identity transformation can be represented by equation (3.3).

$$s = kr \tag{3.3}$$

where r is the intensity level of the pixel located at the coordinates (x, y) in the input image, s is the intensity level of the pixel at the corresponding location in the output image and k is a positive constant.

In linear negative transformation, the intensity levels of an image are reversed. This transformation is useful for enhancing white or grey details embedded in dark regions of an image especially when the dark

regions are dominant in size. The negative of an image with intensity levels in the range [0, L-1] is expressed as,

$$s = L - 1 - r. \qquad (3.4)$$

where *r* is the intensity level of the pixel located at the coordinates (x, y) in the input image while *s* is the intensity level of the pixel at the corresponding location in the output image [2,5].

3.1.2 Logarithmic Transformation

This transformation maps a narrow range of low grey level values in the input image onto a wider range of output grey levels. The values of dark pixels in an image are expanded and higher level pixels values are compressed. This transformation compresses the dynamic ranges of images that have large variations in pixel values. The transformation can be expressed as in equation (3.5).

$$s = c \log(1 + r) \qquad (3.5)$$

where *r* is the intensity level of the pixel located at the coordinates (x, y) in the input image while *s* is the intensity level of the pixel at the corresponding location in the output image and *c* is a positive constant.

The inverse logarithm transformation maps a wide range of low grey level values in the input image onto a narrow range of output levels. This expands the dynamic range of images with small variations in pixel values [2, 5].

3.1.3 Arithmetic and Logic Operations

Arithmetic and logical operations are carried out between corresponding pixels in two or more images. The logical operations that can be performed on an image are: AND, OR and NOT operations. The NOT operation is identical to linear negative transformation. The AND and OR operations are used in image masking.

The logic AND and OR operations between two image f(x, y) and h(x, y) can be expressed as in equations (3.6) and (3.7) respectively.

$$g_1(x, y) = [f(x, y)] \bullet [h(x, y)] \qquad (3.6)$$

$$g_2(x, y) = [f(x, y)] + [h(x, y)] \qquad (3.7)$$

where $g_1(x, y)$ and $g_2(x, y)$ denote the respective output images. The symbols $\bullet$ and $+$ denote the AND and OR operations respectively.

Subtraction and addition are the most widely used arithmetic operations in image enhancement. The difference between two image f(x, y) and h(x, y) can be expressed as in equation (3.8).

$$g(x, y) = f(x, y) - h(x, y) \qquad (3.8)$$

The subtraction is carried out between corresponding pixels in the two input images to produce the output image $g(x, y)$. Image subtraction is widely used in mask mode radiography where blockages in the blood vessels of a patient are detected. An X- ray image of the patient's body is taken and then a contrast medium like iodine dye is injected into the patient's bloodstream before another X-ray image is taken. The two images are subtracted to reveal any blockage in the blood vessels.

For an 8-bit grey-scale image, the grey level values in the difference images can range from a minimum of -255 to a maximum of + 255. Since image intensity values should not lie outside the range 0 to 255, scaling is therefore necessary for the difference image. Scaling is carried out using either of two algorithms. In the first algorithm, 255 is added to all the pixels in the difference image and then each pixel is divided by 2. This method is simple to implement but there is a loss in accuracy resulting in poor quality image. In the second approach, the minimum intensity value in the difference image is obtained and subtracted from all intensity levels in the difference image. Then, the maximum intensity value in the image is obtained. Each pixel is then divided by this maximum value and multiplied by 255. This algorithm is difficult to implement but the output image has better accuracy and quality. Image addition is used in image averaging during image denoising [2, 5].

Consider a noisy image $g(x, y)$ formed by the addition of noise $\eta(x, y)$ to an original image $f(x, y)$.

$$g(x, y) = f(x, y) + \eta(x, y) \tag{3.9}$$

Zero-mean noise can be reduced by adding a set of K noisy images $\{g_i (x, y)\}$ followed by taking the average as follows;

$$g(x, y) = \frac{1}{K} \sum_{i=1}^{K} g_i(x, y) \tag{3.10}$$

As K increases $g(x, y)$ approaches $f(x, y)$ [1].

3.2 Mask Spatial Domain Operations

Mask spatial domain processes are also referred to as spatial domain filtering which can be represented by the expression in equation (3.11).

$$g(x, y) = T_2\{f(x, y)\} \tag{3.11}$$

where $f(x, y)$ is the input image, $g(x, y)$ is the output enhanced image and $T\{\}$ is an operator on the input image to produce the output image. The operator is defined over some spatial neighbourhood of (x, y). The neighbourhood is a square or rectangular sub-image containing the point (x, y) and other points around it [2].

Spatial filtering is achieved by using spatial filter masks. Any image can be viewed as consisting of grey level details corresponding to low frequencies and high frequencies. High frequency components correspond to edges and points in an image. High pass filters are used to obtain the boundaries of the objects in an image. Conversely, the low frequency details correspond to the slowly varying components of the image such as the image texture [2, 5, 6]. The process of spatial filtering involves moving the filter mask from point to point on an image. At each point (x, y), the response of the filter is calculated. For linear spatial

41

filtering the response is given by a sum of products of the filter coefficients and the intensities of the corresponding image pixels in the area spanned by the filter mask. Spatial filters are classified into two categories of smoothing and sharpening filters [2, 5,].

3.2.1 Smoothing Spatial Filters

Smoothing filters are used for image blurring and noise reduction. As the size of the filter mask increases the blurring effect also increases. Smoothing filters can be classified as either linear or non-linear filters [2].

In linear averaging filtering, the value of each pixel in an image is replaced by the average of grey levels in the neighbourhood defined by the filter mask. This process reduces sharp grey level transitions in the image. The false contour effect due to insufficient number of grey levels is reduced by using averaging filters. The disadvantage of averaging filters is that they blur the edges. Averaging filters are of two types, namely box filters in which all the coefficients in the filter mask are equal and weighted average filters. In weighted average filters, the pixel in the center of the mask is multiplied by a higher value, thus giving more weight in the calculation of the average. The other pixels are inversely weighted as a function of their distance from the center of the mask [2, 6].

Non-linear smoothing filters are also termed statistics filters and are characterised by an ordering of the pixels contained in the image region covered by the filter. The three commonly used categories of statistical filters are median filters, maximum filter and minimum filters. In median filters, the value of a pixel in an image is replaced by the median of the grey levels in the neighbourhood of that pixel. Median filters are effective in removal of salt and pepper noise. A maximum filter gives the brightest point in an image whereas a minimum filter gives the darkest point in an image [1, 2, 6].

3.2.2 Sharpening Filters

Sharpening filters highlight fine details in an image and are used to enhance image details that are blurred. Sharpening can be accomplished by spatial differentiation. Image differentiation enhances edges and also de-emphasizes areas with slowly varying grey levels.

The second derivative is zero in flat segments, non-zero at the onset of grey level change and zero along the ramps of constant slopes. A commonly used second derivative sharpening filter is the Laplacian filter. The Laplacian of an image f(x, y) is given by equation (3.12) [2],

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \tag{3.12}$$

where

$$\frac{\partial^2 f}{\partial x^2} = f(x+1,y) + f(x-1,y) - 2f(x,y) \tag{3.13}$$

and

$$\frac{\partial^2 f}{\partial y^2} = f(x,y+1) + f(x,y-1) - 2f(x,y) \tag{3.14}$$

Hence,

$$\nabla^2 f = f(x+1,y) + f(x-1,y) + f(x,y+1) + f(x,y-1) - 4f(x,y). \tag{3.15}$$

The Laplacian operation can be implemented using the mask shown in figure 3.1.

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

Figure 3.1: Laplacian sharpening mask.

First order derivatives in image processing are implemented using the magnitude of the gradient. Given an image f (x, y) the gradient at the point (x, y) is a column vector given as,

$$\nabla f = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \dfrac{\partial f}{\partial x} \\ \dfrac{\partial f}{\partial y} \end{bmatrix} \qquad (3.16)$$

The magnitude of this vector is given as,

$$|\nabla f| = \left[G_x^2 + G_y^2 \right]^{\frac{1}{2}} = \left[\left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2 \right]^{\frac{1}{2}} \qquad (3.17)$$

The magnitude of the gradient filter is 90° rotational invariant and can be approximated as,

$$|\nabla f| \approx |G_x| + |G_y|. \qquad (3.18)$$

First derivative filters include the Canny, Sobel, Prewitt and Robert filters [1, 2].

Other sharpening filters include the unsharp mask filter and the high boost filter. The unsharp Masking filtering is expressed as,

$$(3.19)$$

$$f_s(x, y) = f(x, y) - \bar{f}(x, y).$$

where $f_s(x, y)$ denotes the sharpened image and $\bar{f}(x, y)$ denotes blurred version of the image f(x, y). The high boost filter is a generalisation of the Unsharp mask filter. The high boost filtered image is given by;

$$f_{hb}(x, y) = Af(x, y) - \bar{f}(x, y) \qquad (3.20)$$

where $f_{hb}(x, y)$ denotes the high boost sharpened image, *A* is a positive constant greater than unity and $\overline{f}(x, y)$ denotes blurred version of f(x, y) [2].

3.3 Histogram Processing

A histogram of an image is a plot of the relative frequency of each pixel value that occurs in a given colour of an image. The histogram provides a convenient summary of the intensities in an image, but does not convey any information regarding spatial relationships between pixels. In images that do not contain many very low or very high intensity pixels, it is possible that peaks in the histogram correspond to objects in the image, but it is difficult to be certain about this without visually examining the image. The histogram of a digital image whose size $M \times N$ pixels and having intensity levels in the range [0, L-1] is a discrete function expressed as,

$$h(r_k) = n_k,$$ (3.21)

where r_k is the kth grey level and n_k is the number of pixels in the image having that grey level. The histogram can be expressed as a normalised function as:

$$p(r_k) = \frac{n_k}{MN}$$ (3.22)

for $k = 0, ..., L-1$

The horizontal axis of the histogram corresponds to grey level values. The vertical axis corresponds to $h(r_k)$ or $p(r_k)$. A good quality image has a histogram which is equally distributed over a broad range of grey levels. Histogram equalisation algorithm takes the following procedure:

Let

$$s = T(r)$$ (3.23)

for $0 \leq r \leq L-1$ and $0 =\leq s \leq L-1$.

where *r* is the intensity level of the pixel located at the coordinates (x, y) in the input image and *s* is the intensity level of the pixel at the corresponding location in the output image.

Assuming that the transformation function satisfies the following conditions:

T(r) is single valued and monotonically increasing in the interval $0 \leq r \leq 1$ and $0 \leq T(r) \leq L-1$ for $0 \leq r \leq L-1$, the inverse transformation from s back to r is denoted as,

$$r = T^{-1}(s) \qquad (3.24)$$

The inverse transformation should also satisfy both the above conditions. The grey levels of an image can be viewed as a random variable in the interval [0, L-1] and hence a probability density function can be obtained. Let $p_r(r)$ and $p_s(s)$ denote the *pdfs* of r and s respectively. If $p_r(r)$ and $T(r)$ are known, and $T(r)$ is continuous and differentiable over the entire range of r, then $p_s(s)$ is obtained using equation (3.25).

$$p_s(s) = p_r(r) \frac{dr}{ds}, \qquad (3.25)$$

Let

$$s = T(r) = \int_0^r p_r(w)dw$$

$$\frac{ds}{dr} = \frac{dT(r)}{dr} \quad\quad (3.26)$$

$$= \frac{d}{dr}\left[\int_0^r p_r(w)dw\right]$$

$$= p_r(r)$$

then,

$$p_s(s) = 1 \quad\quad (3.27)$$

Therefore $p_s(s)$ is a uniform probability density function that is independent of $p_r(r)$.

For discrete values,

$$s_k = T(r_k) = \sum_{j=0}^{k} p_r(r_j)$$

$$\quad\quad (3.28)$$

$$= \sum_{j=0}^{k} \frac{n_j}{MN} \quad \text{for} \quad k = 0,\ 1,\ 2,...,L-1$$

The above equation is called the histogram equalisation or histogram linearisation equation.

3.4 Morphological Image Processing

Morphological image processing involves extraction of the image components that are useful in representation and description of the region or object shapes such as boundary information. The concepts of set reflection and translation are used in morphological processing [2, 33, 39, 40, 41]. The reflection of a set of pixels *B* is obtained by shifting every pixel at the location (x, y) to the location (-x, -y). Reflection is represented in set theory as:

$$B^\bullet = \left\{ w/w = -b \quad \text{for} \quad b \in B \right\}$$ (3.29)

where $B^\bullet$ is the reflection of set *B*.

Translation of a set of pixels B by a point $z = (z_1, z_2)$ is the set of points of *B* whose (x, y) coordinates have been replaced by $(x+z_1, y+z_2)$. Translation is represented as:

$$(B)_z = \left\{ c/c = b + z \quad \text{for} \quad b \in B \right\}$$ (3.30)

In image morphology, a small sub image known as a structuring element (SE) is used to probe the image for properties of interest. The structuring elements used in morphological image processing are rectangular arrays which have defined origins. If the origin is not defined, it is assumed to be at the centre of symmetry of the structuring element. The basic morphological operations based on which other morphological operations are defined are: dilation, erosion, opening and closing [2, 5, 6].

Morphological dilation is an operation that grows objects in an image uniformly in spatial extent [6]. Dilation of an image *A* by a structuring element *B* is the set of all displacements *z* such that *A* and *B* overlap by at least one element which is written as:

$$A \oplus B = \left\{ z / [(B^\bullet)_z \cap A] \subseteq A \right\},$$ (3.31)

Dilation thickens the characters and also bridges gaps in the characters as reported in [5].

Erosion of an image *A* by a structuring element *B* is the set of all points *z* such that *B* translated by *z* is contained in *A*. Erosion is represented as:

$$A \ominus B = \left\{ z / (B)_z \subseteq A \right\}$$ (3.32)

Erosion is used in shrinking of the image objects, contracting the boundaries and removing of thin lines [1, 2].

Morphological opening of an image *A* by a structuring element *B* is equivalent to erosion followed by dilation. Opening is represented as:

$$A \circ B = (A \Theta B) \oplus B \qquad (3.33)$$

Opening is used in smoothing image contours, breaking of narrow isthmuses and eliminating of small islands [2, 5].

Closing of an image *A* by a structuring element *B* is equivalent to dilation followed by erosion. Closing is represented as:

$$A \bullet B = (A \oplus B) \Theta B \qquad (3.34)$$

Closing is used in smoothing of image contours, fusing of narrow breaks and elimination of narrow holes [1, 2, 5].

3.5 Single Scale Retinex Algorithm

The Single Scale Retinex (SSR) algorithm is an image enhancement method that improves the brightness perception, contrast and sharpness of a greyscale image. The algorithm achieves these improvements through a combination of spatial and spectral transformations that result in dynamic range compression and a reduction in global variance of the image [42, 43, 44,]. The algorithm belongs to the class of centre-surround functions. Each output value of these functions is determined by the corresponding input value called the centre and its neighbourhood or surround [43]. The mathematical form of single scale retinex is given by:

$$g(x, y) = \alpha(\log[f(x, y)] - \log[s(x, y) \otimes f(x, y)]) - \beta, \qquad (3.35)$$

where $g(x, y)$ is the output image, $f(x, y)$ is the input image, α is a scaling factor, β is an offset parameter, $\otimes$ is the convolution symbol and $s(x, y)$ is a Gaussian filter kernel defined as:

$$s(x, y) = a \exp[-(x^2 + y^2)/\sigma^2], \qquad (3.36)$$

where σ is the standard deviation of the filter which typically ranges from

10 to 250 and a is a constant given by:

$$a = \frac{1}{\sum_x \sum_y s(x, y)}.$$ (3.37)

Due to the large sizes of filter kernels used, the convolution in spatial domain is normally performed as a multiplication in frequency domain to reduce the processing time. The result is then transformed back to spatial domain using Inverse Discrete Fourier Transformation (IDFT) as follows:

$$f(x, y) \otimes s(x, y) = IDFT[F(u, v)S(u, v)],$$ (3.38)

Where $S(u, v)$ and $F(u, v)$ are the discrete Fourier transforms of $s(x, y)$ and $f(x, y)$ respectively.

The single scale retinex algorithm reduces the global variance of an image.

3.6 Image Thresholding

Thresholding techniques can be classified as either global or local methods [8].

In global thresholding techniques, a single optimum threshold level is used to binarise the image. The threshold level is determined as follows: Let $\{0,1,2,...,L-1\}$ denote the L intensity levels of an image of size $M \times N$ pixels and let n_i denote the number of pixels having intensity level i.

The total number of pixels in the image is given by:

$$MN = n_0 + n_1 + n_2 + ... + n_{L-1}.$$ (3.39)

The probability of occurrence of intensity level i is:

$$p_i = n_i / MN$$ (3.40)

where $\sum_{i=0}^{L-1} p_i = 1$ and $p_i \geq 0.$

50

Let a threshold level $T(k) = k,\ 0 \le k \le L-1$ be used to dichotomize the image into two classes C_1 and C_2 where C_1 consists of all the pixels with intensity values in the range $[0, k]$ and C_2 consists of all the pixels with intensity values in the range $[k+1, L-1]$. The probabilities that a pixel is assigned to class C_1 and C_2 respectively are given by [14]:

$$p_1(k) = \sum_{i=0}^{k} p_i = P_1 \tag{3.41}$$

and

$$p_2(k) = \sum_{i=k+1}^{L-1} p_i = 1 - p_1(k) = 1 - p_1. \tag{3.42}$$

The mean intensity values of the pixels assigned to classes C_1 and C2 respectively are:

$$m_1 = \sum_{i=0}^{k} ip(i/C_1) = \frac{1}{p_1(k)} \sum_{i=0}^{k} ip_i, \tag{3.43}$$

and

$$m_2 = \sum_{i=k+1}^{L-1} ip(i/C_2) = \frac{1}{p_2(k)} \sum_{i=k+1}^{L-1} ip_i. \tag{3.44}$$

The cumulative mean up to intensity level *k* is given by:

$$m(k) = m = \sum_{i=0}^{k} ip_i. \tag{3.45}$$

The global mean of the image is given by:

$$m_G = \sum_{i=0}^{L-1} ip_i. \tag{3.46}$$

The global and between-class variances σ_G^2, and $\sigma_B^2(k)$ respectively are given by:

$$\sigma_G^2 = \sum_{i=0}^{L-1}(i-m_G)^2 p_i, \tag{3.47}$$

and

$$\sigma_B^2 = \frac{[m_G p_1(k) - m]}{p_1(k)[1-p_1(k)]}. \tag{3.48}$$

Class separation metric, η is defined as follows

$$\eta = \frac{\sigma_B^2(k)}{\sigma_G^2} = \frac{[m_G p_1(k) - m(k)]^2}{p_1(k)[1-p_1(k)]}. \tag{3.49}$$

The optimum threshold is the value, k_{opt} that maximizes the between-class variance obtained from

$$\sigma_B^2(k_{opt}) = \max \sigma_B^2(k), \tag{3.50}$$

for $0 \le k_{opt} \le 1$.

In local thresholding methodologies, local image information obtained in the neighbourhood of each pixel is used to determine the threshold level for each pixel. The commonly used equation for calculating the local threshold level for a pixel at the coordinates (x, y) are given in equations (3.51) and (3.52) [2, 8].

$$T_{xy} + a\sigma_{xy} + b\mu_{xy} \tag{3.51}$$

and

$$T_{xy} + a\sigma_{xy} + b\mu_G \tag{3.52}$$

where *a* and *b* are constants, μ_{xy} and σ_{xy} denote the mean and standard deviation of the set of pixels contained in a neighbourhood that is centered at the coordinates (x, y) whereas μ_G is the global mean of the image. The thresholded image g(x, y) is obtained as:

$$g(x, y) = \begin{cases} 1 & \text{if} \quad f(x, y) > T_{xy} \\ \\ 0 & \text{if} \quad f(x, y) \leq T_{xy} \end{cases} \qquad (3.53)$$

where $f(x, y)$ is the input image.

Equation 3.53 is evaluated for all pixel locations in the image and a different threshold is computed at each location (x, y) using the pixels in the neighbourhood centered at that location. In general, local binarisation methods perform better than global binarisation methods on degraded document images but are computationally slow and are sensitive to the selection of parameter values *a* and *b* [15, 16].

3.7 Enhancement by Blur and Noise Reduction

Image degradations such as blur are modeled in terms of a degradation function $H(u, v)$ that together with additive noise $\eta(x, y)$ operates on an input image $f(x, y)$ to produce a degraded image $g(x, y)$. The degradation reduction process is implemented using filters whose objective is to obtain an estimate $\hat{f}(x, y)$ of the original image. The more $H(u, v)$ and $\eta(x, y)$ are known, the closer $\hat{f}(x, y)$ is to $f(x, y)$ [1, 2]. The image blurring and deblurring model is shown in figure 3.2.
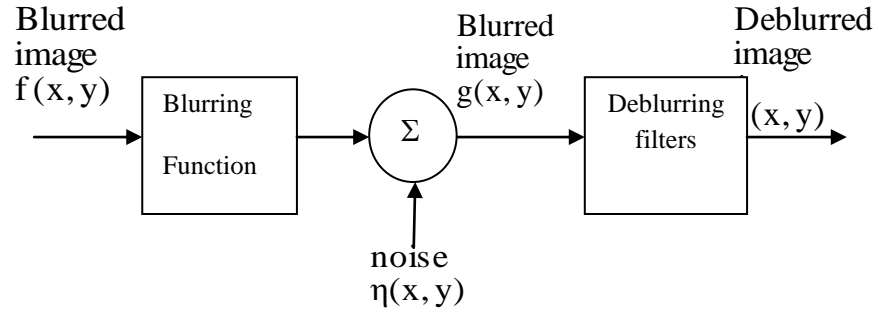
Figure 3.2: Image blurring-deblurring model.

When $H(u, v)$ is a linear position invariant process, the degraded image in spatial domain is given by:

$$g(x, y) = [h(x, y) \otimes f(x, y)] + \eta(x, y) \qquad (3.54)$$

where h(x,y) is the point spread function obtained by taking the inverse discrete Fourier transform of $H(u, v)$ and $\otimes$ denotes convolution.

The equivalent of equation 3.54 in frequency domain is:

$$G(u, v) = H(u, v)F(u, v) + N(u, v) \qquad (3.55)$$

where G(u, v), H(u, v), F(u, v) and N(u, v) are the two-dimensional discrete Fourier transforms of g (x, y), h(x, y), f(x, y) and $\eta(x, y)$ respectively.

Mean smoothing filters are useful in separating image signals from noise, particularly when the signal and the noise occupy different frequency bands. The estimated output image for a mean filter is given by equation (3.56).

$$\hat{f}_1(x, y) = \frac{1}{MN} \sum_{(s,t) \in s_{xy}} g(s, t) \qquad (3.56)$$

Where s_{xy} is a set of coordinates in a rectangular sub-image of size $M \times N$ centered at point (x, y) [2, 5].

While mean filters are useful for deterministic sinusoidal noise or for statistically uniform random noise, they do not work as well for impulsive noise known as salt-and-pepper noise. Salt-and-pepper noise is

54

characterized by large spikes in isolated pixels. A typical example is random bit errors in a communication channel used to transmit images. Only a small fraction of the pixels in the image are affected, but the error in the affected pixels is often great. Median filters are well suited for the removal of this kind of noise [2]. A median filter mask of length N replaces the intensity of the image pixel at which it is centred with the median of the N image intensities encompassed by the filter mask.

CHAPTER 4

TRANSFORM DOMAIN ENHANCEMENT TECHNIQUES

In transform domain operations the image is first transformed according to an orthogonal transformation such as Discrete Fourier transform (DFT), Discrete Cosine Transform (DCT) or Discrete Wavelet Transform (DWT). The transformed image is then multiplied by a transfer function of an appropriate filter to yield the enhanced image in frequency domain. Inverse transformation is finally carried out to produce the enhanced image in the spatial domain.

Frequency domain methods include smoothing filtering for removal of high frequency noise and bridging of gaps in text images, sharpening filtering to highlight edges and homomorphic filtering which is used to suppress the illumination component of the image and also as a shadow removal technique [2, 4, 6].

4.1 General Transform Domain Principles

A linear two-dimensional transform of an $M \times N$ digital image $f(x, y)$ takes the general form;

$$F(u, v) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x, y) r(x, y, u, v) \tag{4.1}$$

where $r(x, y, u, v)$ is the forward transform kernel, x and y are the spatial coordinates, u and v are the transform or frequency coordinates, F (u, v) is the forward transform of the image f(x,y).The variables x, y, u, and v satisfy the following inequalities;

$0 \le x \le M-1, 0 \le y \le N-1, \quad 0 \le u \le M-1$ and $0 \le v \le N-1$.

Given F(u, v), the input image is recovered using the inverse transform of F(u, v) as,

$$f(x,y) = \sum_{u=0}^{M-1}\sum_{v=0}^{N-1}F(u,v)S(x,y,u,v) \qquad (4.2)$$

where S(x, y, u, v) is the inverse transformation kernel

The equations for f(x, y) and F (u, v) form a transform pair

The forward and reverse transformation kernel are said to be separable if they satisfy the following equations.

$$r(x,y,u,v) = r_1(x,u)r_2(y,v)$$

$$s(x,y,u,v) = s_1(x,u)s_2(y,v)$$

$$(4.3)$$

The kernels are said to be symmetric if they satisfy the following equations.

$$r(x,y,u,v) = r_1(x,u)r_1(y,v)$$

$$s(x,y,u,v) = s_1(x,u)s_1(y,v)$$

$$(4.4)$$

The two dimensional Discrete Fourier Transform (DFT) has the following forward and reverse transform kernels.

$$r(x,y,u,v) = e^{-j2\pi(\frac{ux}{M}+\frac{vy}{N})}$$

$$s(x,y,u,v) = \frac{1}{MN}e^{j2\pi(\frac{ux}{M}+\frac{vy}{N})}$$

$$(4.5)$$

The kernels are separable and symmetric for square images (M=N). The two-dimensional discrete Fourier transform pair can therefore represented by:

$$F(u,v) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)e^{-j2\pi(\frac{ux}{M}+\frac{vy}{N})}$$

$$f(x,y) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1}F(u,v)e^{j2\pi(\frac{ux}{M}+\frac{vy}{N})}$$

$$(4.6)$$

Frequency domain enhancement techniques can be classified into two groups of Smoothing and Sharpening filtering techniques [1, 2, 5].

4.1.1 Smoothing Frequency Domain Filters

In smoothing filtering, high frequency components in the transform of an image are attenuated. High frequencies are an attribute of edges which are sharp intensity transitions and by noise. The basic model for filtering in the frequency domain is given by,

$$G(u, v) = H(u, v)F(u, v) \qquad (4.7)$$

where, $F(u, v)$ is the two dimensional discrete Fourier transform of the input image $f(x, y)$. The filter transfer function is $H(u, v)$ and $G(u, v)$ is the Fourier transform of the enhanced output image $g(x, y)$. The different types of low pass filters include the following: ideal low pass filter, Butterworth low pass filter and the Gaussian Low pass Filter.

The ideal low pass filter has a sharp transition from band pass to band reject frequencies while a Gaussian filter has smooth transitions. A low order Butterworth filter has smooth transitions while a higher order one has sharper transitions [1, 2].

The transfer function of the Gaussian low pass filter is given by:

$$H(u, v) = e^{-D^2(u,v)/D_0^2(u,v)} \qquad (4.8)$$

where $D(u, v)$ is the distance from point (u, v) to the origin and $D_o(u,v)$ is the cut off frequency. If the image is of size M x N pixels, the origin of the transform is at the center (M/2, N/2). The distance from any point (u, v) to the centre of the Fourier transform is given by,

$$D(u, v) = \left[(u - M/2)^2 + (v - N/2)^2 \right]^{\frac{1}{2}} \qquad (4.9)$$

The Butterworth low pass filter is used when tight control of transition between low and high frequencies about the cut off is required.

The Gaussian Low pass is used when a smooth transition between low and high frequencies about the cut off is required.

Low pass filtering is used in printing and publishing industry. The characters in a particular document may be broken or have distorted shapes. A machine recognition system will have difficulties reading such characters. Hence a low pass filter is applied onto such an image, to bridge small gaps by introducing some limited blur [1, 2].

4.1.2 Sharpening Frequency Domain Filters

Images can be blurred by attenuating the high frequency components. Inversely, image sharpening can be achieved in frequency domain by using high pass filters, which attenuate the low frequency components.

The transfer function of the high pass filters can be obtained by,

$$H_{hp}(u, v) = 1 - H_{lp}(u, v) \qquad (4.10)$$

Where $H_{lp}(u, v)$, is the transfer function of a low pass filter and $H_{hp}(u, v)$, is the transfer function of the high pass filter. High pass filters include the following: ideal high pass filter, Butterworth high pass filter, Gaussian high pass filter and the homomorphic filter.

The ideal high pass filter transfer function given by,

$$H(u, v) = \begin{cases} 0 & \text{if} \quad D(u, v) \leq D_0(u, v) \\ 1 & \text{otherwise} \end{cases} \qquad (4.11)$$

where $D_0(u, v)$ is the cut off frequency of the filter [1].

The ideal high pass filter sets all frequencies inside the circle of radius $D_0(u, v)$ to zero but passes all other frequencies outside the circle, without attenuation.

The transfer function of the Butterworth high pass filter of order *n* is given by,

$$H(u, v) = \frac{1}{1 + \left[D_0(u, v)/D(u, v)\right]^{2n}} \tag{4.12}$$

where $D_0(u, v)$ is the cut off frequency of the filter.

The Butterworth high pass filter introduces less edge distortions than the ideal high pass filter. In addition, it has a smoother transition from the stopband to the passband region [1,2].

The transfer function of the Gaussian high pass filter is given by,

$$H(u, v) = 1 - e^{-D^2(u, v)/2D_0^2(u, v)} \tag{4.13}$$

The filter has a smoother response when compared to the other two filters. In addition, the filter can be constructed as a difference of Gaussian low pass filters. This construction makes it easy to control the filter shape.

4.2 Homomorphic Filtering

An image is characterised by the amount of light incident on the scene which is referred to as the illumination $i(x, y)$ and the amount of light reflected by the objects in the scene referred to as the reflectance $r(x, y)$. The two components combine as a product to form the image $f(x, y)$ as follows:

$$f(x, y) = i(x, y)r(x, y), \tag{4.14}$$

where $0 < i(x, y) < K$ and $0 < r(x, y) < 1$.

K is an upper bound for the illumination whose practical value is approximately equal to 9×10^4 Lumens/m^2 for photographic images [1, 2]. The illumination component is characterised by low frequencies while the reflectance is characterized by higher frequencies. The homomorphic filtering algorithm reduces the contribution of the illumination component by means of a modified Gaussian high pass filter. In order to separate the two image components the logarithm of the image is taken as [5, 8]:

$$\log f(x, y) = \log[i(x, y)] + \log[r(x, y)]. \tag{4.15}$$

Let z(x,y) = log[f(x,y)], then taking the two-dimensional DFT yields:

$$Z(u, v) = DFT\{\log i(x, y)\} + DFT\{\log r(x, y)\}. \qquad (4.16)$$

The transformed image Z(u,v) is filtered using a Gaussian high pass filter having a transfer function H(u,v) to give:

$$S(u, v) = H(u, v)F_i(u, v) + H(u, v)F_r(u, v). \qquad (4.17)$$

The filtered image in spatial domain is obtained by the inverse Discrete Fourier transformation as:

$$s(x, y) = IDFT[H(u, v)F_i(u, v) + H(u, v)F_r(u, v)]. \qquad (4.18)$$

Taking the anti-logarithm of $s(x, y)$ yields the enhanced image $g(x, y)$ as follows:

$$g(x, y) = \exp[s(x, y)]$$
$$= i_o(x, y)r_o(x, y) \qquad (4.19)$$

Where $i_o(x, y) = \exp[i^{'}(x, y)]$ and $r_o(x, y) = \exp[r^{'}(x, y)]$

The transfer function of the modified Gaussian high pass filter $H(u, v)$ is given by:

$$H(u, v) = (\gamma_H - \gamma_L)[1 - \exp(-c[D^2(u, v) / D_o^2])] + \gamma_L \qquad (4.20)$$

Where γ_H, γ_L, c and D_o are constants and $D(u, v)$ is the distance between a point (u, v) in the frequency domain to the centre of the frequency rectangle.

Also $\gamma_H > 1$, $\gamma_L < 1$, $0 < c < 1$ and $10 \leq D_o \leq 300$ have been found to be suitable values [1, 7].

4.3 Hough Transform

The Hough transform is used as a technique for edge linking. It involves the transformation of a line in Cartesian coordinate space to a point in polar coordinate space. A straight line can be described parametrically as [2]

$$\rho = x\cos\theta + y\sin\theta \qquad\qquad (4.21)$$

where ρ is the normal distance of the line from the origin and θ is the angle between the normal to the line and the x axis as shown in figure 4.1(a) [2].

The Hough transform of the line is simply a point at (ρ, θ) coordinate in the polar domain as shown in the figure 4.1(b). A family of lines passing through a common point, as shown in figure 4.1(c), maps onto the connected set $\rho - \theta$ of points of figure 4.1(d). The Hough transform of the family of curves passing through three colinear points in figure 4.1e) result in the set of three parametric curves in the $\rho - \theta$ space as in figure 4.1(f). The three curves cross at a single point (ρ_0, θ_0) corresponding to the dashed line passing through the colinear points.

(a) Parametric line

(b) Hough transform of (a)

(c) Family of lines, common point

(d) Hough transform of (c)

$$\rho_0 = x \cos \theta_0 + y \sin \theta_0$$

(ρ_0, θ_0)

(e) Colinear points

(f) Hough transform of (e)
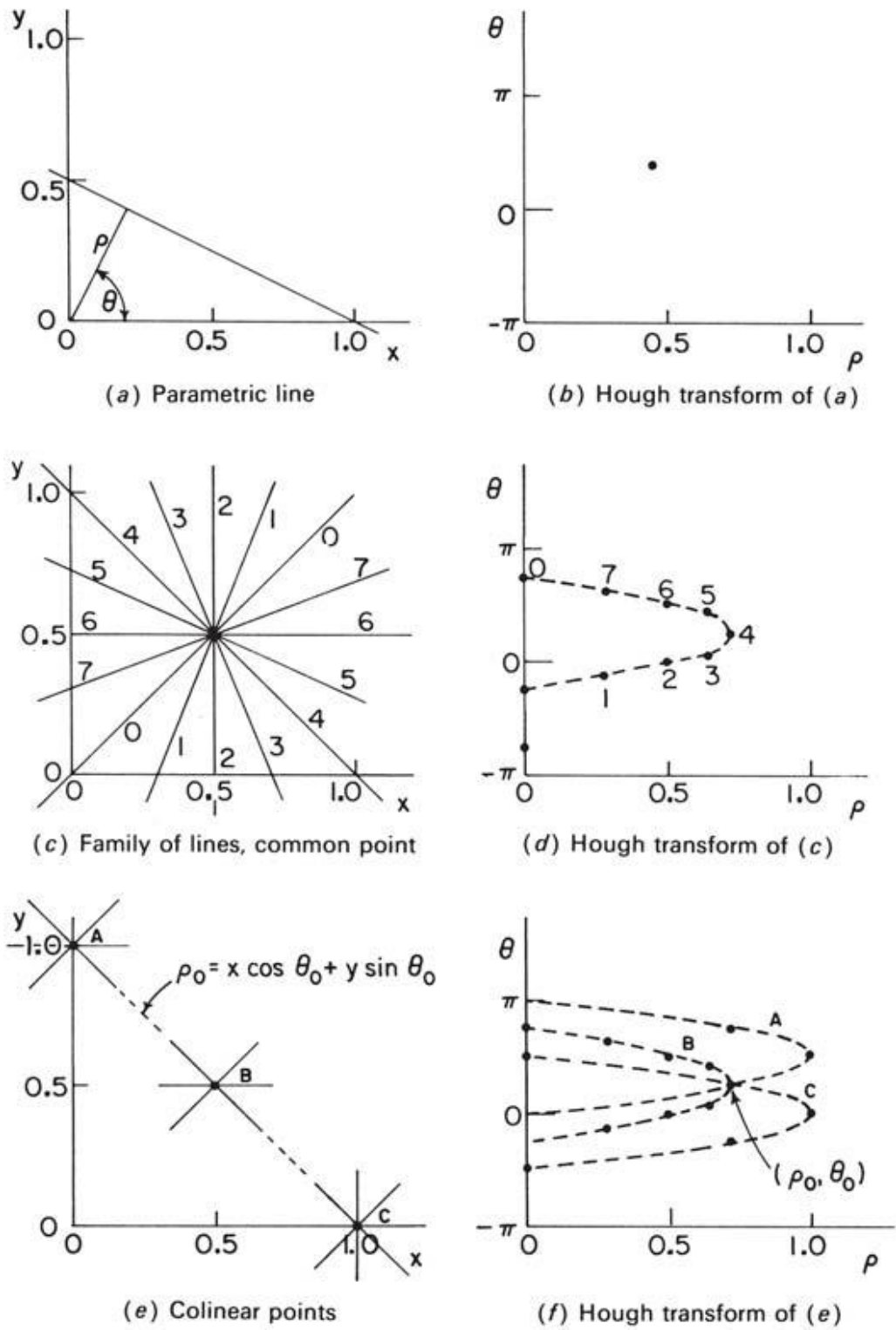
Figure 4.1: Hough transforms (after [2]).

4.4 Discrete Wavelet Transform

A wavelet is a localized signal whose energy is concentrated in time or space. Discrete Wavelets Transform (DWT) has been found to be suitable for analysis of transient signals since it provides both time (or spatial) and frequency information [1, 2].

4.4.1 One-Dimensional DWT

The DWT of a one-dimensional discrete signal f(x) of length M is obtained using the following equations.

$$W_\phi(j_0, m) = \frac{1}{\sqrt{M}} \sum_{m=0}^{M-1} f(x)\phi_{j_0,m}(x) \tag{4.22}$$

$$W_\psi(j, m) = \frac{1}{\sqrt{M}} \sum_{m=0}^{M-1} f(x)\psi_{j,m}(x) \quad \text{for } j \geq j_0 \tag{4.23}$$

Where $\phi(x)$ and $\psi(x)$ are basis functions termed the scaling function and the wavelet respectively. The functions $\phi_{j,m}(x)$ and $\psi_{j,m}(x)$ are scaled and translated versions of the basis functions which are given by the following equations.

$$\phi_{j,m}(x) = 2^{j/2}\phi(2^j x - m) \tag{4.24}$$

$$\psi_{j,m}(x) = 2^{j/2}\psi(2^j x - m) \tag{4.25}$$

The original signal $f(x)$ is obtained from the inverse discrete wavelet transform as follows:

$$f(x) = \frac{1}{\sqrt{M}} \sum_{m=0}^{M-1} W_\phi(j_0, m)\phi_{j_0,m}(x) + \frac{1}{\sqrt{M}} \sum_{j=0}^{J-1} \sum_{m=0}^{M-1} W_\psi(j, m, n)\psi_{j,m}(x) \tag{4.26}$$

where j_0 is an arbitrary starting scale which is usually set to zero,

$W_\varphi(j_0, m)$ defines an approximation of $f(x,)$ at the starting scale

$W_\psi(j, m)$ gives the details for scales greater than the starting scale.

Normally the 1D-DWT of a digital signal is obtained with the following settings:

$j_0=0$, $M=2^J$, $J=0, 1, 2, \ldots, J-1$ and $m = 0, 1, 2, \ldots, 2^{j-1}$.

Some of the commonly used wavelets are the Haar, Daubechies, Coiflet, symlet, Meyer, Morlet and the mexican hat wavelets [2]. The Haar wavelet and scaling function are the oldest and the simplest. The scaling function is obtained from the first row of a 2X2 Haar matrix **H₂** given in equation (4.27) while its wavelet function is obtained from the second row of the same matrix as follows [2, 45]:

$$H_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{4.27}$$

$$\psi(x) = \begin{cases} \dfrac{1}{\sqrt{2}} & \text{for } 0 \leq x < 0.5 \\[2mm] -\dfrac{1}{\sqrt{2}} & \text{for } 0.5 \leq x \leq 1 \\[2mm] 0 & \text{elsewhere} \end{cases} \tag{4.28}$$

and

$$\phi(x) = \begin{cases} \dfrac{1}{\sqrt{2}} & \text{for } 0 \leq x \leq 0.5 \\[2mm] 0 & \text{elsewhere} \end{cases} \tag{4.29}$$

4.4.2 Two-Dimensional DWT

DWT is also applied to two-dimensional signals such as images. A two-dimensional scaling function $\varphi(x, y)$ and three 2-D wavelet functions $\psi^H(x, y)$, $\psi^V(x, y)$ and $\psi^D(x, y)$ are used.

Each of these two-dimensional functions is a product of two one-dimensional functions. The scaling function is scalable while the wavelet functions are both scalable and directionally sensitive. These functions are given by:

$$\varphi(x, y) = \varphi(x)\varphi(y) \tag{4.30}$$

$$\psi^H(x, y) = \psi(x)\varphi(y) \tag{4.31}$$

$$\psi^V(x, y) = \varphi(x)\psi(y) \tag{4.32}$$

$$\psi^D(x, y) = \psi(x)\psi(y) \tag{4.33}$$

The wavelets measure the image intensity variations in different directions: $\psi^H(x, y)$ measures variations along the image columns, $\psi^V(x, y)$ measures variations along the image rows, and $\psi^D(x, y)$ measures variations along the diagonals.

Two-dimensional DWT of an image f(x,y) that has a size of $M \times N$ pixels is given by:

$$W_\phi(j_0, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y)\phi_{j,m,n}(x, y) \tag{4.34}$$

$$W_\psi^i(j, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y)\psi_{j,m,n}^i(x, y) \tag{4.35}$$

Where i= {H, V, D}

The scaled and translated basis functions used in equations (4.34) and (4.35) are defined as follows:

$$\varphi_{j,m,n}(x,y) = 2^{j/2}\varphi(2^j x - m, 2^j y - n) \qquad (4.36)$$

$$\psi^i_{j,m,n}(x,y) = 2^{j/2}\psi^i(2^j x - m, 2^j y - n) \qquad (4.37)$$

where i is an index that identifies the directional wavelets. It takes on the values H, V and D.

j_0 is an arbitrary starting scale.

$W_\varphi(j_0,m,n)$ defines an approximation of f(x,y) at the starting scale.

$W^i_\psi(j,m,n)$ are three sequences that give the horizontal, vertical and diagonal details for scales greater than the starting scale.

Normally the DWT of an image is obtained with the following settings:

$j_0=0$, $M=N=2^J$, $J=0, 1, 2,…, J-1$ and $m=n=0, 1, 2,…, 2^{j-1}$.

Given the DWT, the original image $f(x,y)$ is obtained from the inverse discrete wavelet transform as follows:

$$
\begin{aligned}
f(x,y) = {} & \frac{1}{\sqrt{MN}}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} W_\varphi(j_0,m,n)\varphi_{j_0,m,n}(x,y) + \\
& \frac{1}{\sqrt{MN}}\sum_{i=H,V,D}\sum_{j=j_0}^{J-1}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} W^i_\varphi(j,m,n)\psi^i_{j,m,n}(x,y)
\end{aligned}
\qquad (4.38)
$$

4.5 Blur Reduction Filtering

In many situations the acquired image is a blurred version of the original image. Image deblurring can be achieved by means of filters such as the inverse or the Wiener filters as described here.

Let $f(x,y)$ be the original image and $g(x,y)$ be the blurred image in the absence additive noise. The blurred image can be expressed in terms of the original image and the blurring function H (u, v) in both spectral and spatial domains by equations (4.39) and (4.40) respectively.

$$G(u, v) = H(u, v)F(u, v) \qquad (4.39)$$

and

$$g(x, y) = h(x, y) \otimes f(x, y) \qquad (4.40)$$

where G(u, v), H(u, v) and F(u, v) are the two-dimensional discrete Fourier transforms of g (x, y), h(x, y) and f(x, y) respectively.

A simple solution for restoring $f(x, y)$ from $g(x, y)$ is to define the inverse filter function $B(u, v)$ as follows:

$$B(u, v) = \frac{1}{H(u, v)} \qquad (4.41$$

Applying this inverse filter function on the blurred image yields the original image. The problem with this approach is that $B(u, v)$ can take on extremely high values at points where $H(u, v)$ is close to zero. Even small amounts of noise such quantisation noise or computation noise can lead to huge errors in the reconstruction of $f(x, y)$. One way of avoiding this problem is to choose a suitable threshold γ and use the inverse filter function given by:

$$B(u, v) = \begin{cases} \dfrac{1}{H(u, v)} \text{ if } \dfrac{1}{H(u, v)} < \gamma \\[4mm] \dfrac{\gamma |H(u, v)|}{H(u, v)} \text{ otherwise} \end{cases} \qquad (4.42)$$

The effect of equation (4.42) is to limit the maximum value of $|B(u, v)|$ to γ [1, 2].

Wiener filtering is used in restoration of images corrupted by both a degradation function and random noise. The objective is to find an

estimate $\hat{f}(x, y)$ of $f(x, y)$ such that the mean square error is a minimum. The mean square error, e^2 is defined as;

$$e^2 = E\{(f(x, y) - \hat{f}(x, y))^2\}$$ (4.43)

Three assumptions are made in formulating the filter. The first assumption is that the image and noise are uncorrelated. The second assumption is that, the noise has a zero mean. The third assumption is that the intensity levels in the estimated image are a linear function of the intensity levels of the degraded image [1, 2].

The estimated image that gives the least mean square error is given in frequency domain by equation (4.44).

$$\hat{F}(u, v) = \left[\frac{|H(u, v)|^2}{H(u, v)[|H(u, v)|^2 + S_\eta(u, v)/S_f(u, v)]} \right] G(u, v)$$ (4.44)

Where $H(u, v)$ and $H^*(u, v)$ are the degradation function and its complex conjugate. The two are related to the magnitude of the degradation function according to equation (4.45).

$$|H(u, v)|^2 = H(u, v)H^*(u, v)$$ (4.45)

$S_f(u, v) = |F(u, v)|^2 = $ The power spectrum of the input image

$S_\eta(u, v) = |N(u, v)|^2 = $ The power spectrum of the noise

$G(u, v) = $ The DFT of the degraded image.

The Wiener filter does not have the problem associated with the inverse filter when additive noise is present. If the image is not corrupted by noise, the Wiener filter reduces to the inverse filter.

The problem with the Wiener filter is that the degradation function and the power spectrum of the original image and noise must be known.

When these quantities are not known equation (4.44) is estimated by equation (4.46).

$$F(\hat{u}, v) = \left[\frac{|H(u, v)|^2}{H(u, v)[|H(u, v)|^2 + K]} \right] G(u, v) \qquad (4.46)$$

where K is a constant [1, 5].

CHAPTER 5

ENHANCEMENT OF GEOMETRICALLY DISTORTED IMAGES

It is usual for images of documents that have been acquired by a digital camera to exhibit some form of degradations. These degradations are due to the type of printed document such as bounded volumes of books which result in warping near the book spine, the camera setup and environmental conditions like humidity and dust that cause page shrinkage. These geometrical distortions not only reduce document legibility but also affect the performance of subsequent processes such as document layout analysis and optical character recognition (OCR). The efficiency of these image processes increase with improvement in the straightness of the document text lines [46]. The rectification of geometrically distorted document images can be classified into two main categories.

The first category is based on 3-D document shape reconstruction [47, 48, 49, 50, 51, 52]. The methods in this category require special hardware such as stereo cameras, structured lighting systems and laser scanners as well as prior knowledge of the distortion type. These methods are outside the scope of this thesis and shall not be considered here.

The second category is based on 2-D document image processing and uses only the information available in the image without requiring any special hardware or prior knowledge of the nature of the distortion [53, 54, 55, 56, 57, 58, 59] The majority of these 2-D rectification techniques are based on the detection of distorted text lines of the original document image.

In this chapter, several mathematical methods that can be employed to model geometric distortions are presented and enhancement based on them is discussed. The method discussed here is based on the

2-D document image processing and consists of four stages discussed in sections 5.1 to 5.4.

There are also rectification techniques that do not rely on the detection of distorted text lines but they aim at finding spatial transformations between the warped and dewarped document images by analysing the 2-D content such as document boundaries or known reference points. Some of these techniques include the use of document boundary interpolation to correct geometric distortions and shading artifacts present in images of art-like materials and approximation of the deformation in interlinear spaces of an image based on elements of the image skeleton that lie between the text lines [27].

5.1 Distorted document image preprocessing

This is the first stage in the enhancement of geometrically distorted document images. The objective here is to binarise the distorted image and convert its words into labeled connected components which are suitable for mathematical modeling. The preprocessing is carried out in three steps which are binarisation, smoothing and finally connected component labeling.

The process of document image binarisation involves conversion of a greyscale image into a binary image. The process is performed either globally or locally. For the global methods, a single calculated threshold value is used to classify image pixels into object or background classes, while for the local or adaptive binarisation methods, local area information guides the selection of the threshold value for each pixel.

The Run Length Smoothing Algorithm (RLSA) is frequently used in block segmentation and text discrimination. The method consists of two steps. First, a segmentation procedure subdivides the area of a document into regions or blocks each of which should contain only one type of data such as text and graphics. Next, some basic features of these blocks are

calculated [60, 61]. The basic RLSA procedure 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 rule.

The 0's in X are changed to 1's in Y if the number of adjacent 0's is less than or equal to a predefined limit C whereas 1's in X are unchanged in Y. For example, with $C = 4$ the sequence X is mapped into Y as shown in figure 5.1.

X=	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0
Y=	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1	1

Fig 5.1: Run Length Smoothing of a Binary Sequence.

When applied to pattern arrays, the RLSA has the effect of linking together neighbouring black areas that are separated by less than C pixels. With an appropriate choice of C, the linked areas will be regions of a common data type.

The RLSA is applied row-by-row as well as column-by-column to a document, yielding two distinct bit-maps. Because the spacings of the document components tend to differ horizontally and vertically, different values of C are used for row and column processing. The two bit-maps are then combined in a logical AND operation. Additional horizontal smoothing using the RLSA produces the final segmentation result.

Various algorithms for connected component labeling of images have been proposed [62, 63, 64]. These algorithms can be divided into four classes. The first class consists of methods with repeated forward and backward passes over data. The second class consists of methods with two passes over data. These algorithms maintain a separate vector or an array to store equivalence information of labels and use search algorithms to resolve equivalences. The third class consists of methods

that use a hierarchical tree structure to represent the data. Finally, the fourth class consists of methods that use parallel algorithms. The Rosenfeld and Pfalts algorithm [62] is one of the commonly used connected component labeling algorithm. This algorithm belongs to the second class according to the classification outlined above. This algorithm scans the image from left to right and top to bottom using a 4-neighbour forward raster scan mask. The document is also scanned right to left and bottom to top using the 4-neighbour backward raster scan mask. The Rosenfeld and Pfalts scan masks are given in figure 5.2.

a	b	c
e	d	

e	d
c	b	a

(a)   (b)
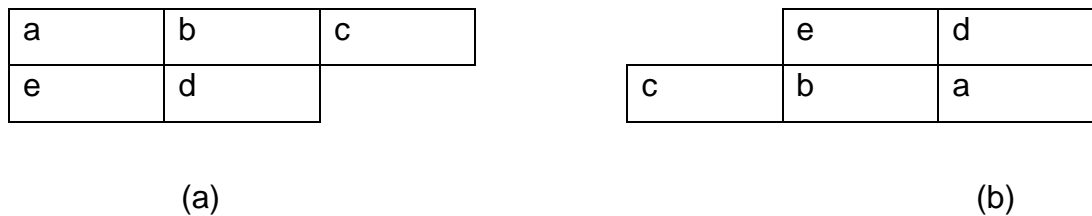
Fig 5.2: Connected component labeling masks. (a) Forward raster scan.
(b) Backward raster scan.

Current pixel position of this scan mask is considered to be position "e" (see Fig. 5.2). If the value of the current pixel position "e" is 0 then the mask is moved to the next scanning position. If current pixel "e" is 1 and all other 4-neighbour positions are 0 then a new label is assigned to the pixel "e". If two or more 4-neighbour pixels are not zero then the minimum label of the 4-neighbours is assigned to "e" and then the labels in the 4-neighbour are marked as the same component [62].

5.2 Modeling of the Curved Surface Projection

The second stage in the enhancement of geometrically distorted document images involves the formulation of a mathematical model for the

curved document surface so that it can be mapped onto a plane rectangular surface.

The modeling starts with word and text line detection which follows the same procedure as that used in the rectification using text baseline detection. Once the text lines have been detected, modeling of the curved surface projection proceeds as follows: the curved surface is considered to be delimited by the two curved lines which fit the top and bottom text lines along with the two straight lines which fit the left and right text boundaries [61]. Let A, B, C, and D denote the dominant corner points of the projection of the curved surface as shown in figure 5.3.
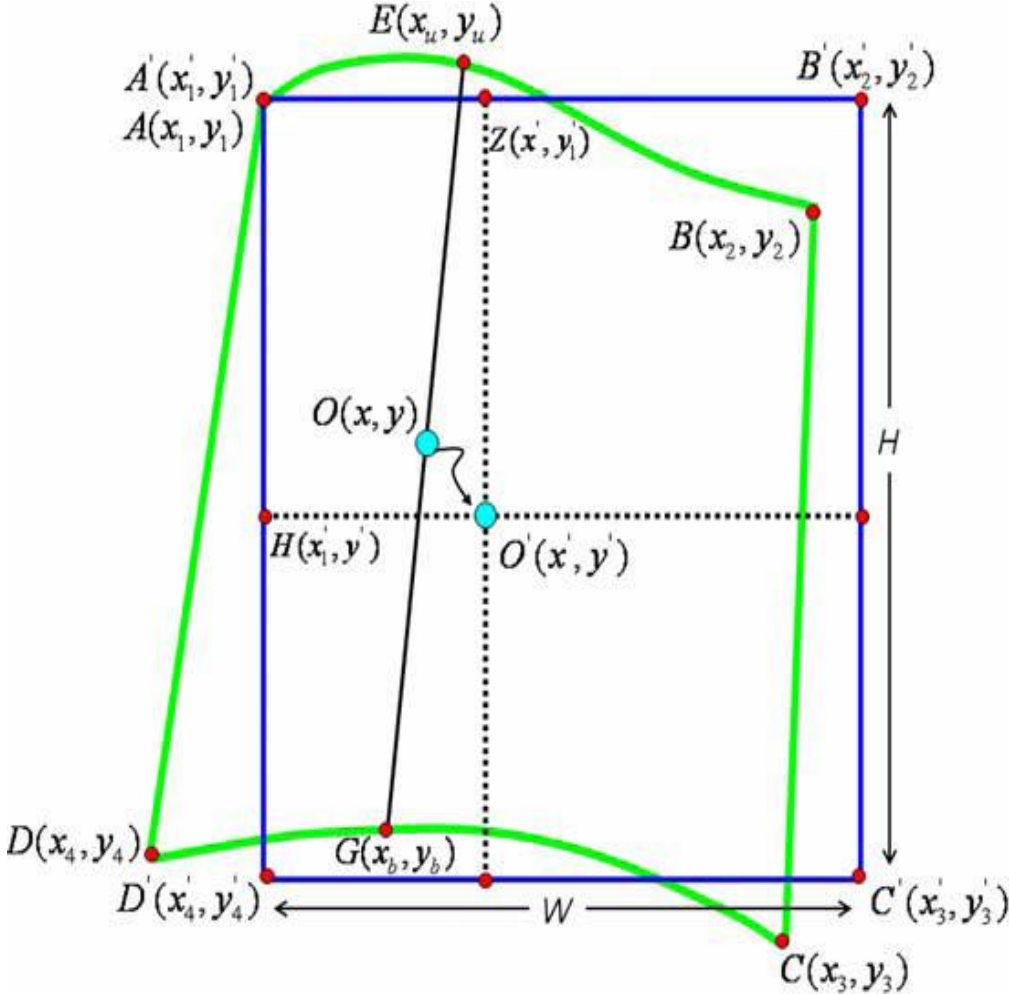
Figure 5.3: Curved surface projection (after [61]).

First, the straight lines AD and BC which correspond to the left and right text boundaries are estimated. This is done by first detecting the start and end points of each text line followed by exclusion of short text lines using the average length of all the text lines. In this way, short text lines such as titles, marginal text, graphics and mathematical expressions are eliminated leaving only the most representative text lines. Among these text lines retained, some of them do not start (or end) from the beginning (or ending) of the document frame, mainly the first and last text line of a paragraph. In order to further eliminate these text lines an iterative procedure is applied. If the deviation of the estimated straight line is greater than the dominant character height the start (or end) point with the maximum distance is excluded and the straight line is recalculated. This iteration is stopped when the above criterion is satisfied or only two text lines have remained as shown in figure 5.4.
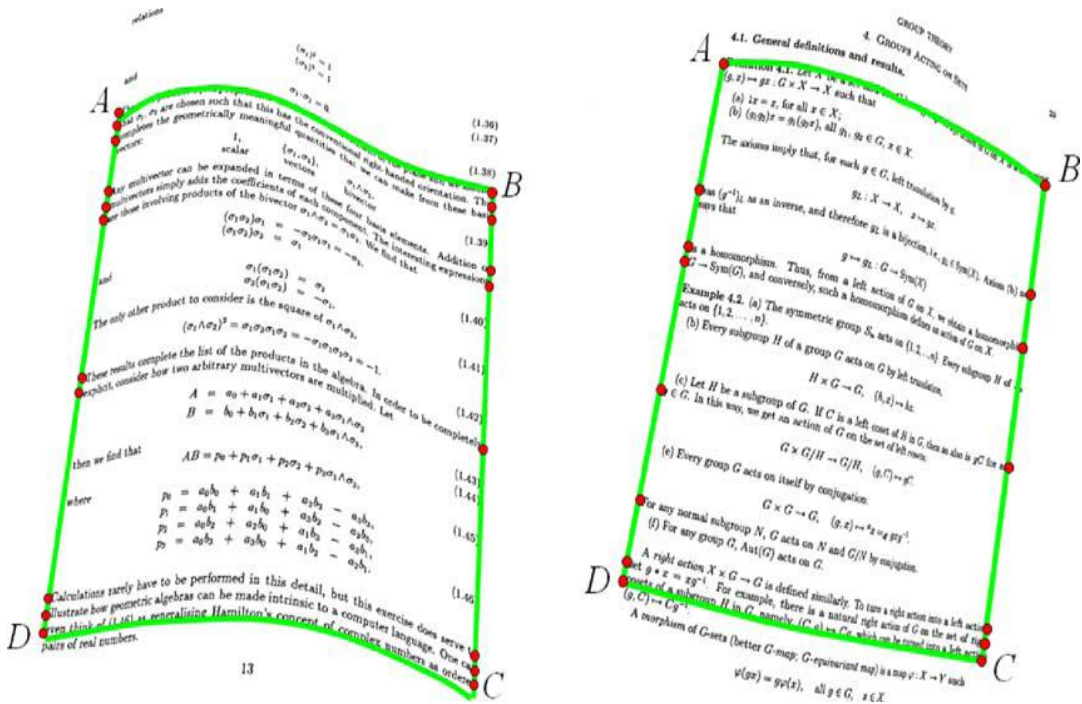
Fig 5.4: Text lines detection [after 61].

Since the iterative procedure uses only long text lines, the potential deviation of the estimated straight line will not be large.

Next, the curved lines AB and DC which correspond to the top and bottom text lines are estimated. In order to select appropriate text lines of the document with representative deformation we select the top and bottom text lines which participate in the calculation of the straight lines AD and BC as shown in figure 5.3. In this way, short text lines are excluded. Using the upper and bottom points of the selected text lines respectively, the coefficients of a third degree polynomial are calculated [59].

After modeling the projection of the curved surface on the plane delimited by the curved line segments and along with the straight line segments, the next goal is to generate a transformation that maps the projected curved surface to a 2-D rectangular area.

Let $A^{'}$, $B^{'}$, $C^{'}$, and $D^{'}$ denote the corner points of the rectangular area as shown in figure 5.3. Also, let $\overset{\frown}{AB}$ and $|AB|$ represent the arc length and the Euclidean distance, respectively, between points A and B. First, the corner points and the rectangular area are located. One of the corners coincides with one of the dominant corner points of the projection of the curved surface ($A'(x^{'}_{1}, y^{'}_{1}) \equiv A(x_{1}, y_{1})$) and the rest of them are calculated by taking into account the width W and the height H of the rectangular area which are given by equations 5.1 and 5.2 respectively.

$$W = \min\left(\overset{\frown}{AB}, \ \overset{\frown}{DC} \right)$$ 5.1

and

$$H = \min\left(|AD|, \ |BC| \right)$$ 5.2

Once the rectangular area has been defined, each point $O(x, y)$ in the projection of the curved surface is mapped to the corresponding point $O'(x', y')$ in the rectangular area. Each point is defined by two points (E and G) according to the arc length of the top and bottom curved line segments (see Figure 5.3). Points E and G satisfy the following condition:

$$\frac{\overset{\frown}{AE}}{\overset{\frown}{AB}} = \frac{\overset{\frown}{DG}}{\overset{\frown}{DC}} \tag{5.3}$$

The corresponding point $O'(x', y')$ in the rectangular area is calculated by preserving the ratio between the projection of the curved surface and the rectangular area in the x-direction as well as in the y-direction.

In order to estimate the straight lines AD and BC which correspond to the left and right text boundaries as well as the curved lines AB and DC which correspond to the top and bottom text lines is as follows:

Let the start and end points of the ith text line of a document having N text lines be represented by the coordinate pairs (x_{si}, y_{si}) and (x_{ei}, y_{ei}) respectively for $0 < i \leq N$.

The length L_i of the ith text line is calculated as in equation 5.4.

$$L_i = \sqrt{\left(\left(x_{ei} - x_{si}\right)^2 + \left(y_{ei} - y_{si}\right)^2\right)} \tag{5.4}$$

The average length L_{av} of all the text lines is determined from;

$$L_{av} = \frac{1}{N} \sum_{i=0}^{N} L_i \tag{5.5}$$

Every text line i is excluded if its length, $L_i \leq 0.8 L_{av}$.

Using the text lines that remain, a Least Squares Estimation (LSE) method is used to get the straight lines AD and BC that fit into set of points which denote the start and end points of each text line, respectively. For each straight lines AD and BC, the average Euclidean distance between

the line and the start (and end) points of each text line is determined. If the average distance is greater than the dominant character height h_d, the point with the maximum Euclidean distance is excluded and the procedure is repeated starting from the step of calculating the Euclidean distances until the aforementioned condition is satisfied.

In the top and bottom curved line segments estimation, the top and bottom text lines which take part in the calculation of the straight lines are detected. Then, all the upper points of the top text line and the bottom points of the bottom text line are detected. Finally, a least squares estimation method is used to find the coefficients of the third degree polynomial curves that fit all top and all the bottom points detected [57].

5.3 Mapping of the Curved Surface onto a Plane

After modeling the top and bottom curved ends of the document image using third degree polynomial curves, the next stage involves the mapping of the curved document surface onto a rectangular area of width W and height H. The mapping is affected by first determining the corners of the rectangular area followed by calculation of the positions onto which each pixel in the degraded image are to be mapped on the rectangular area as described in the rest of this section.

In order to locate the corner points of the rectangular area, one of them is chosen to coincide with one of the dominant corner points of the geometrically distorted image. In figure 5.3, this corner is selected to be $A(x'_1, y'_1) \equiv A(x_1, y_1)$. The rest of the corners are calculated by taking into account the width and the height of the rectangular area as follows:

$$(x'_2, y'_2) = (x'_1 + W, y'_1)$$

(5.6)

$$(x_3^{'}, y_3^{'}) = (x_1^{'} + W, y_1^{'} + H)$$

$$(x_4^{'}, y_4^{'}) = (x_1^{'}, y_1^{'} + H)$$

A correspondence is then created between the points of the curved

line segments $\overset{\frown}{AB}$ and $\overset{\frown}{DC}$ which is expressed by a function as follows:

$$G(x_b, y_b) = \zeta(E(x_u, y_u)) \qquad (5.7)$$

where $E(x_u, y_u)$ represents a point on the top curved line segment $\overset{\frown}{AB}$

and $G(x_b, y_b)$ represents a point on the bottom curved line segment $\overset{\frown}{DC}$.

Let O(x, y) represent a point on the projection of the curved surface. The goal is to calculate the new position O'(x',y') of O(x, y) on the rectangular area.

First, a straight line is defined which satisfies the following criteria:

The first criterion is that the line intersects the curved lines $\overset{\frown}{AB}$ and $\overset{\frown}{DC}$

at the points $E(x_u, y_u)$ and $G(x_b, y_b)$, respectively.

The second criterion is given by equation (5.7) as:

The third criterion is that: $O(x, y) \in \overline{EG}$

where $\overline{EG}$ is the straight joining points $E(x_u, y_u)$ and $G(x_b, y_b)$.

The position of $O'(x', y')$ is calculated from:

$$x' = x'_1 + |A'Z|$$
$$y' = y'_1 + |A'H| \qquad (5.8)$$

where H is the point $H(x'_1, y')$ which belongs to the left side of the rectangular area, and Z is the point $Z(x', y'_1)$ which belongs to the top side of the rectangular area. Finally, $|A'Z|$ and $|A'H|$ are calculated using the relationship as given in equations (5.9) and (5.10).

$$\frac{\overset{\cap}{AB}}{\overset{\cap}{AE}} = \frac{W}{|A'Z|} \Rightarrow |A'Z| = \frac{W}{\overset{\cap}{AB}} \overset{\cap}{AE} \qquad (5.9)$$

and

$$\frac{|EG|}{|EO|} = \frac{H}{|A'H|} \Rightarrow |A'H| = \frac{H}{|EG|}|EO|. \qquad (5.10)$$

The mapping is then performed for all points which are inside the projection area borders. All points which are not included at the projection area of the curved surface inherit the transformation of the nearest point [61].

5.4 Text Line Detection and Dewarping

The final stage of the enhancement of distorted document images entails the detection of the document text lines followed by modeling of the text lines using linear equations. Rectification of the distortion is finally performed using rotations and translation of the document words.

Text line detection starts with the estimation of the dominant character height h_d in order to temporarily remove non-text components such as pictures, graphs and noise from the image. This removal leaves only the text which will be used in the detection of text lines. For the calculation of the dominant character height, connected component labeling is applied to the binarised image and then the height of each bounding box of the connected components is calculated followed by height histogram construction. The dominant character height is given by the maximum value of the histogram.

Connected components which satisfy the following conditions given in equations 5.11 are removed.

$$h > 3h_d \quad \text{or} \quad h < \frac{h_d}{4} \quad \text{or} \quad t < \frac{h_d}{4} \qquad (5.11)$$

where *h* and *t* denote the height and width of the connected components respectively. The connected components that satisfy inequalities (5.11) represent noise and non text information in the image [18].

For the word detection, horizontal smoothing using the RLSA procedure is first applied with a threshold $\mathrm{Th} = 0.5\mathrm{h_d}$ followed by connected component labeling in order to detect words.

Text lines detection is then performed after performing word detection. In a left-right and top-down raster scanning, the first word is detected and assigned as the first word of the first text line. Following that, horizontally neighbouring words in the left and right directions are consecutively linked in order to detect all the words of the first text line. Finally, the scanning is continued and repeated for the remaining text lines until all the words have been assigned to text lines. The process of linking two neighbouring words is addressed as follows: Let $(x_1, y_1), (x_2, y_2)$ denote the bounding box coordinates of an assigned word and $(x'_1, y'_1), (x'_2, y'_2)$ denote the bounding box coordinates of a candidate neighbour word. For all the words in the right side of the assigned word which satisfy the condition $[y_1, y_2] \cap [y'_1, y'_2] \neq 0,$ the one with the smallest distance $D = x'_1 - x_2$ only is selected if $0 < D < 6h_d$. This condition ensures selection of the immediate neighbour word of the same text line since many words may satisfy the condition of horizontal overlapping [61]. Next, this word is assigned as processed and the search continues on the right side for a neighbour word till the last word of the text line is assigned as processed. A similar treatment is applied to the left side.

The lower and upper baselines which delimit the main body of each word are then detected. Linear regression is applied on the set of points that are the lowest and topmost black pixels for each text line column to

model the upper and lower baselines. The upper baseline of word w_{ij} in the ith row and j[th] column of the text document is defined as:

$$y = a_{ij}x + b_{ij} \qquad (5.12)$$

Similarly, the lower baseline of word w_{ij} is defined as:

$$y = a'_{ij}x + b'_{ij} \qquad (5.13)$$

Also a_{ij} is the slope of the upper baseline and a'_{ij} is the slope of the lower baseline. All the detected words are then rotated and shifted in order to obtain a dewarped binary image. The slope of each word is derived from the corresponding baseline slopes. The upper and lower baseline slope angles of word w_{ij} are respectively given by:

$$\theta_{ij}^u = \arctan(a_{ij})$$

$$\text{and} \qquad (5.14)$$

$$\theta_{ij}^l = \arctan(a'_{ij})$$

Since the smaller slope is usually the most representative, the word's slope can be defined as:

$$\theta_{ij} = \begin{cases} \theta_{ij}^u, & \text{if } \left|\theta_{ij}^u\right| < \left|\theta_{ij}^l\right| \\ \theta_{ij}^l & \text{otherwise} \end{cases} \qquad (5.15)$$

The rotation of the word $w_{ij}(x,y)$ yields a rotated word $w_{ij}^r(x^r, y^r)$ where:

$$y^r = (x - x_{min})\sin(-\theta_{ij}) + y\cos(\theta_{ij}),$$

$$x^r = x, \qquad (5.16)$$

and x_{min} is the left end of the bounding box of the word w_{ij}

After word rotation, all the words of every text line, except for the leftmost, are vertically shifted in order to restore horizontal alignment. The rotation and shifting of the word $w_{ij}(x,y)$ is done as follows:

$$y^{rs} = (x - x_{min})\sin(-\theta_{ij}) + y\cos(\theta_{ij}) + d_{ij}$$

(5.17)

$$x^{rs} = x$$

where $w_{ij}^{rs}(x^{rs}, y^{rs})$ is the rotated and shifted word. The factor d_{ij} is the vertical shift which is given by the following relationship:

$$d_{ij} = \begin{cases} y_{i0}^{rl} - y_{ij}^{rl}, & \text{if } \left|\theta_{ij}^{l}\right| < \left|\theta_{ij}^{u}\right| \\ y_{i0}^{ru} - y_{ij}^{ru}, & \text{otherwise} \end{cases}$$

(5.18)

where:

$$y_{ij}^{rl} = \left(a_{ij}x_{min} + b_{ij}\right)\cos\left(\theta_{ij}\right)$$

(5.19)

$$y_{ij}^{ru} = \left(a'_{ij}x_{min} + b'_{ij}\right)\cos\left(\theta_{ij}\right)$$

The reason for defining two shifting levels for each text line is that each word may be rotated either according to its lower baseline or upper baseline slope. Hence, it has to be shifted so that its lower or upper baseline is aligned with the lower or upper baseline of the leftmost word of the text line [18, 61].

CHAPTER 6

MATERIALS AND METHODS

This chapter is divided into three sections. Section 6.1 gives a summary of the materials that were used in this research. The proposed enhancement method is presented in section 6.2. In Section 6.3, an outline of published image quality measures is presented. Some of these quality metrics are used to test the quality of the experimental results presented in chapter 7.

6.1 Materials

The following materials were used to generate and process the test images.

(i)   A Nokia 1680 cell phone [65] together with its computer interface USB cable and Nokia PC suite version 7.1 drivers to facilitate loading of the pictures to the computer memory. This particular model of phone was selected because it is cheap and widely available.

    The cell phone camera specifications are: A maximum image size of 480×640 pixels, minimum image size of 120×160 pixels, 1.85 inches LCD display and uses JPEG image format.

(ii)   A Fujifilm model 4800Z digital camera [66] was used to capture the images that were used as the ground-truth images in testing of the enhanced images using the objective quality test methods. The camera is capable of capturing both greyscale and colour images at resolutions of 8 bits and 24 bits respectively.

    The camera specifications are: A maximum image size of 2400×1800 pixels, minimum image size of 640×480 pixels, 2.0 inches LCD display, maximum shutter closure time of 3 seconds, minimum shutter closure time of 0.5miliseconds and a zoom factor

of 36. It also uses the JPEG image format.

(iii)   Math works MATLAB version 7.14 [67]. This is a numeric computation, data analysis and visualization software that is used as a programming language for many engineering and scientific applications. The image processing section of the software was used to code and test the algorithms that were developed.

(iv)   Optical Character Recognition (OCR) software [68, 69, 70].

(v)   A desktop personal computer was used as the workstation. The PC specifications are: Intel Pentium 4 processor, 3GHz clock speed, 2GB RAM and a 64-bit word size. The PC is loaded with the Microsoft windows 7 operating system.

6.2 Methods

In this section, the proposed enhancement methods are presented. Sub section 6.2.1 presents a complete enhancement algorithm for images that suffer one or more types of degradations which do not include geometrical distortions on text document images. Images that are degraded by geometrical distortion in addition to other types of degradations such as shadows and fading are also subjected to the same enhancement procedure as a pre-processing stage.

The geometrically distorted images are processed using the algorithm described in sub section 6.2.2. This processing procedure is aimed at reducing the geometrical distortion as well as other degradations. The procedure is based on mathematical modeling, mapping and affine transformations.

The test images used in the experiments carried out during this thesis were captured as 120x160 pixels 24-bit colour images using a Nokia 1680 cellphone camera. Also, all the enhancement experiments were carried out using computer simulations based on MATLAB version 7.14.

6.2.1 Non-geometrically Distorted Images.

The proposed algorithm used to enhance degraded images that have no geometrical distortion is summarised in form of a block diagram in figure 6.1.
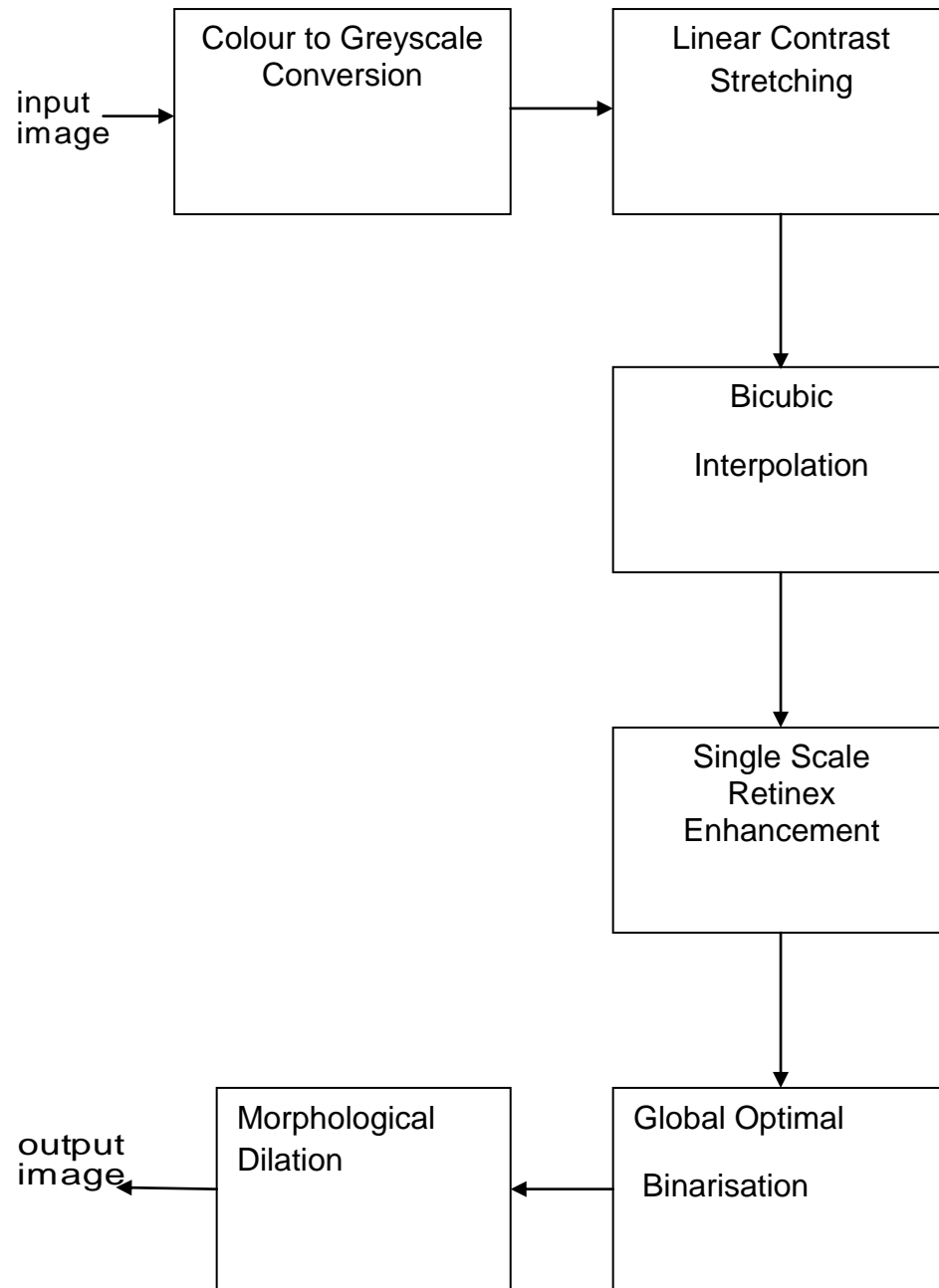

Figure 6.1: Proposed enhancement procedure.

The procedure is as follows:

(i)   The first processing stage involves conversion of the colour images to greyscale images. The conversion was achieved by removing the hue and saturation components in the colour images using the 'im2gray' function in MATLAB 7.14.The outputs of the conversion were 120x160 pixels, 8-bit greyscale images.

(ii)   The greyscale images were then enhanced using contrast stretching in order to expand the range of their greylevels to span the full intensity range (0 to 255) of the cellphone camera. The linear contrast stretching method described in section 2.1 was used.

(iii) The images were then interpolated to increase their sizes from 120x160 pixels to 480x640 pixels. Interpolation of the images led to improved interpretation of the image information as well as better legibility of text document images. The bicubic interpolation method described in section 2.5 was used due to its relatively low computational complexity compared to interpolation methods of higher degrees. It also results in less image distortion when compared with both nearest neighbour and bilinear interpolation methods.

(iv)   The interpolated images were enhanced using the SSR algorithm described in section 3.5. The SSR enhancement has the effects of improving the brightness perception, contrast and sharpness of an image. These effects resulted in suppression of shadows, compensation for poor illumination and noise reduction.

(v)   The output images of the SSR enhancement were then binarised using the Otsu's global thresholding. The method is based on maximisation of a class separation measure that leads to the determination of an optimum global threshold level as explained in section 3.6.

(vi)   The binarised images were finally enhanced using morphological dilation to bridge any gaps in the text resulting from the binarisation process. Morphological dilation achieves the gap bridging effect by growing the characters in the document image uniformly in spatial extent as explained in section 3.4.

6.2.2 Geometrically Distorted Images.

Geometrically distorted document images were enhanced using the procedure outlined in form of a block diagram in figure 6.2. The procedure is as follows:

(i)   The pre-processing stage is implemented using the enhancement procedure described in section 6.2.1. This stage reduces the degradation effects caused by factors such as shadows, fading, low illumination levels and noise. Document legibility is improved by global thresholding. Legibility is further enhanced by the character stroke connectivity improvement effect of morphological dilation.

(ii)   The pre-processed image is subjecting to the run length smoothing algorithm described in section 5.1 so as to convert the text words to connected components or blocks.

(iii)   Connected component labeling is then performed on the smoothed image using the Rosenfeld and Pfalts Algorithm also described in section 5.1.

(iv)   The next step involves word and text lines detection. Word detection is done by first removing the non-text connected components so as to be left with the connected components that are derived from the text words only. Removal of non-text components and detection of text lines were carried out as explained in section 5.4.

(v)   The geometrically distorted image surface is then mathematically modeled. Linear equations are used to represent the right and left edges while the top and bottom edges are modeled using third degree polynomials as explained in section 5.2.

(vi)   The modeled surface is then mapped onto a two-dimensional rectangular surface as described in section 5.2.

(vii)   The words and text lines of the transformed image are then detected using the procedures given in section 5.4.

(viii)   Following the detection of the words and text lines, both upper and lower text baselines of each word are approximated with straight line equations using linear regression as described in section 5.4.

(ix)   The last step involves reduction of the geometrical distortion. Using the slopes of the baselines, each word is rotated to a horizontal orientation. Next, character slants and skews are corrected by rotating them to vertical positions. Finally, for each text line, the first word is detected and all the other words belonging to that text line are aligned with the first word. The alignment is achieved by vertical translations of the words of the text line to the same horizontal line as the first word as described in section 5.4.
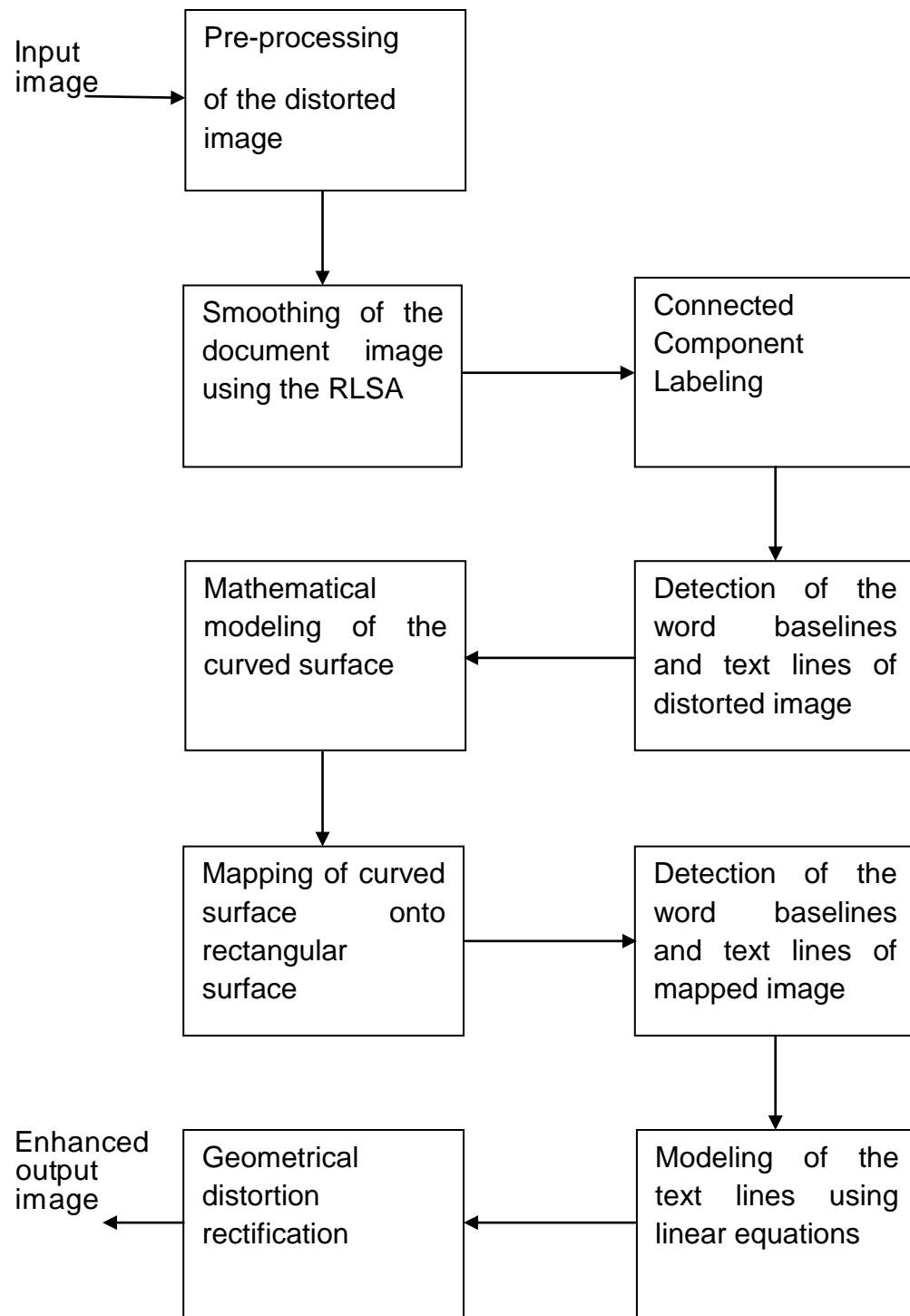
```
                  ┌─────────────────────┐
Input             │ Pre-processing      │
image ──────────▶ │ of the distorted    │
                  │ image               │
                  └─────────────────────┘
                             │
                             ▼
         ┌─────────────────────┐      ┌─────────────────────┐
         │ Smoothing of the    │      │ Connected           │
         │ document image      │ ───▶ │ Component           │
         │ using the RLSA      │      │ Labeling            │
         └─────────────────────┘      └─────────────────────┘
                                                 │
                                                 ▼
         ┌─────────────────────┐      ┌─────────────────────┐
         │ Mathematical        │      │ Detection of the    │
         │ modeling of the     │ ◀─── │ word baselines      │
         │ curved surface      │      │ and text lines of   │
         │                     │      │ distorted image     │
         └─────────────────────┘      └─────────────────────┘
                    │
                    ▼
         ┌─────────────────────┐      ┌─────────────────────┐
         │ Mapping of curved   │      │ Detection of the    │
         │ surface onto        │ ───▶ │ word baselines      │
         │ rectangular         │      │ and text lines of   │
         │ surface             │      │ mapped image        │
         └─────────────────────┘      └─────────────────────┘
                                                 │
                                                 ▼
Enhanced  ┌─────────────────────┐      ┌─────────────────────┐
output    │ Geometrical         │      │ Modeling of the     │
image  ◀─ │ distortion          │ ◀─── │ text lines using    │
          │ rectification       │      │ linear equations    │
          └─────────────────────┘      └─────────────────────┘
```

Figure 6.2: Post processing for distorted document images.

6.3 Image Quality Measures

Image quality measures are categorised into either subjective or objective types. Subjective tests rely upon the judgement of a human observer whereas objective tests are based on mathematical quality measures. Subjective tests are more trusted and reliable than objective metrics. However these tests are expensive to carry out since they involve interviewing of subjects. They are also difficult to reproduce and verify. Due to these reasons, objective quality metrics are also commonly used to determine the quality of enhanced images.

Various objective measures have been proposed. These metrics can be classified into two categories. The first category consists of those measures that require ground-truth images for comparison with the enhanced output images. These measures include: Root Mean Square Error (RMSE), Normalised Least Square Error (NLSE), Peak Signal to Noise Ratio (PSNR), Correlation (CORR) and Mutual Information (MI) [1]. For an image of size $M \times N$, these measures are defined as follows,

$$RMSE = \sqrt{\frac{\sum_{x=1}^{M}\sum_{y=1}^{N}[r(x,y)-f(x,y)]^2}{MN}} \qquad (6.1)$$

$$NLSE = \sqrt{\frac{\sum_{x=1}^{M}\sum_{y=1}^{N}[r(x,y)-f(x,y)]^2}{\sum_{x=1}^{M}\sum_{y=1}^{N}[r(x,y)]^2}} \qquad (6.2)$$

$$PSNR = 20\log_{10}\left(\frac{L}{RMSE}\right)dB \qquad (6.3)$$

$$CORR = \frac{2\sum_{x=1}^{M}\sum_{y=1}^{N}r(x,y)f(x,y)}{\sum_{x=1}^{M}\sum_{y=1}^{N}[r(x,y)]^2 + \sum_{x=1}^{M}\sum_{y=1}^{N}[f(x,y)]^2} \qquad (6.4)$$

$$MI = \sum_{x=1}^{M} \sum_{y=1}^{N} h_{rf}(x,y) \log_2\left(\frac{h_{rf}(x,y)}{h_r(x,y)h_f(x,y)}\right)$$ (6.5)

where;

$r(x,y)$ is the reference image

$f(x,y)$ is the output image

L is the maximum pixel intensity value

$h_{rf}(x,y)$ is the normalized joint histogram

$h_r(x,y)$ is the normalized histogram of the reference image

$h_f(x,y)$ is the normalized histogram of the output image.

The second category consists of the quality metrics that do not require a ground-truth reference images for their evaluation [71, 72, 73]. The method proposed by Petrovic and Xydeas [71] is a popular measure that is based on edge information transfer. It is outlined below:

Let a(x, y) and b(x, y) be the two images that are fused to form image f(x, y). Two Sobel filters kernels k_x and k_y are used to yield edge strengths $g(x,y)$ for each image and the orientation information $\alpha(x,y)$ as follows for the input image a(x, y):

$$k_x = \begin{array}{|c|c|c|} \hline -1 & 0 & 1 \\ \hline -2 & 0 & 2 \\ \hline -1 & 0 & 1 \\ \hline \end{array}$$ (6.6)

$$k_y = \begin{array}{|c|c|c|} \hline -1 & -2 & -1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 2 & 1 \\ \hline \end{array}$$ (6.7)

$$g_a(x,y) = \sqrt{\left([S_a^x(x,y)]^2 - [S_a^2(x,y)]^2\right)}$$ (6.8)

and

93

$$\alpha_a(x, y) = \tan^{-1}\left\{\frac{S_a^y(x, y)}{S_a^x(x, y)}\right\} \qquad (6.9)$$

where;

$S_a^x(x, y)$ is the convolution of $a(x, y)$ with k_x

$S_a^y(x, y)$ is the convolution of $a(x, y)$ with k_y

The relative strength and orientation of image a(x, y) in the fused image are respectively given by:

$$G^{af}(x, y) = \begin{cases} \dfrac{g_f(x, y)}{g_a(x, y)} & \text{for} \quad g_a(x, y) > g_f(x, y) \\[4mm] \dfrac{g_a(x, y)}{g_f(x, y)} & \text{otherwise} \end{cases} \qquad (6.10)$$

and

$$\alpha^{af}(x, y) = \frac{\left|\left\{\left|\alpha_a(x, y) - \alpha_f(x, y)\right|\right\} - \pi/2\right|}{\pi/2} \qquad (6.11)$$

The edge strength and orientation preservation values are respectively obtained as:

$$Q_g^{af}(x, y) = \frac{\Gamma_g}{1 + e^{k_g(G^{af}(x, y) - \sigma_g)}} \qquad (6.12)$$

and

$$Q_\alpha^{af}(x, y) = \frac{\Gamma_\alpha}{1 + e^{k_\alpha(G^{af}(x, y) - \sigma_\alpha)}} \qquad (6.13)$$

where $\Gamma_g, k_g, \sigma_g, \Gamma_\alpha, k_\alpha$ and σ_α are constants.

Edge information preservation value for image A(x, y) is then defined by:

$$Q^{af}(x, y) = Q_g^{af}(x, y) Q_\alpha^{af}(x, y) \qquad (6.14)$$

Where $0 \le Q^{af}(x, y) \le 1$.

A normalised weighted performance metric value for the edge preservation of the two fused images is given by:

$$Q_P^{ab/f} = \frac{\sum\limits_{x=1}^{M}\sum\limits_{y=1}^{N}[Q^{af}(x,y)w^a(x,y) + Q^{bf}(x,y)w^b(x,y)]}{\sum\limits_{x=1}^{M}\sum\limits_{y=1}^{N}\left(w^a(x,y) + w^b(x,y)\right)} \tag{6.15}$$

where

$$w^a(x,y) = \{g_a(x,y)\}^L$$

$$w^b(x,y) = \{g_b(x,y)\}^L$$

L is a constant whose value is usually chosen to be 1.5 to satisfy the condition $0 \leq Q_P^{ab/f} \leq 1$.

Stamatopoulos et al. [61] have proposed two quality measures based on the Optical Character Recognition results of typed document images. The measures are character accuracy and word accuracy which are defined as follows:

Character accuracy is the ratio of the correctly recognized characters to the total number of characters in the correct document transcription. Word accuracy is the ratio of the correctly recognised words to the total number of words in the correct document transcription.

Both subjective and objective test methods were used in this study to evaluate the performance of the proposed algorithm. Images captured by a high resolution Fujifilm digital camera were used as the standard ground-truth images for testing the enhanced image.

CHAPTER 7

COMPUTER SIMULATION RESULTS AND DISCUSSIONS

In this chapter, some of the results obtained from a series of experiments conducted using computer simulation based on MATLAB are presented. Section 7.1 gives a presentation of the results obtained using the already published spatial domain image enhancement approaches. The results presented in Section 7.2 were obtained using various transform domain enhancement techniques. Results of geometrical distortion reduction based on two-dimensional processing of the image are given in section 7.3. The results presented in section 7.4 were obtained using the proposed algorithm. Comparison of some of the proposed method results with the results obtained using some state-of-the art enhancement methods as well as quality measure results are presented in this section.

7.1 Spatial Domain Enhancement Results

In this section, computer simulation results of experiments performed in spatial domain are presented. The input images presented here have been bicubic interpolated to a size of 480x 640 pixels.

7.1.1 Contrast stretching

Figure 7.1 shows the result of contrast stretching enhancement of a greyscale image. The linear contrast stretching method discussed in section 2.1 was used. The input image shown in figure 7.1(a) was captured under poor illumination conditions and has parameter values: $f_{max} = 207$ and $f_{min} = 5$. The image shown in figure 7.1(b) is the result of applying the linear contrast stretching enhancement on the input image. The enhanced image has a higher contrast than the input image. This improvement in contrast is indicated by the ease with which the objects in the image are recognised. The improvement in contrast is as a result of

the darkening of the intensity levels below 24 and brightening of the levels above this threshold value by the contrast stretching function used.


<center>(a)   (b)</center>

Figure 7.1: Contrast stretching. (a) Input image. (b) Contrast stretched output image.

7.1.2 Linear Negative Transformation

Figure 7.2 shows the result of using linear negative transformation on a greyscale document image captured under conditions of low illumination. The effect of this transformation is to enhance fine dark text details that are embedded in the dominant grey background regions of the image. A Subjective test performed on the two images resulted in 65% of 20 subjects interviewed choosing the negatively transformed image as having better legibility.
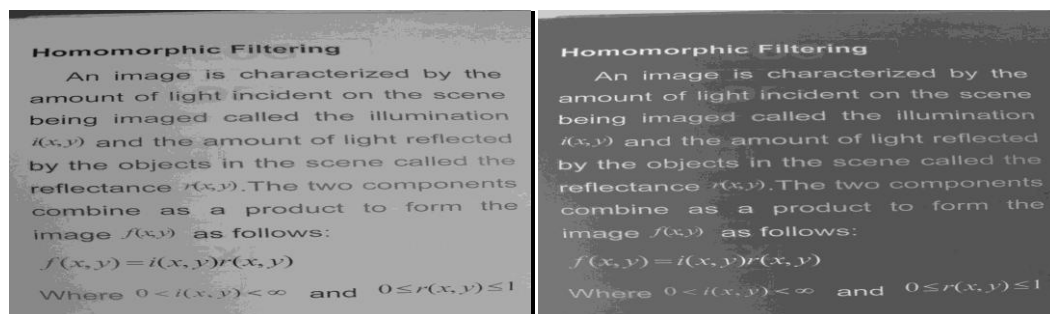
<center>(a)   (b)</center>

Figure 7.2: Linear negative transformation. (a) Input image. (b) Enhanced image.

7.1.3 Spatial Domain Smoothing

Figure 7.3 shows the result of spatial domain smoothing filtering of a document image using a 3x3 box mask. The input image in figure 7.3(a) has gaps in some of the text characters such as in the last word of the first row. One of the effects of smoothing filtering is to bridge these gaps in the characters as in figure 7.3(b). Another effect is to enlarge small foreground details such as punctuation marks and noise artifacts as is evident at the beginnings of the first two lines of the image in figure7.3 (b).

the scene called the   the scene called the

The two components   .The two components

product to form the   product to form the

(a)   (b)
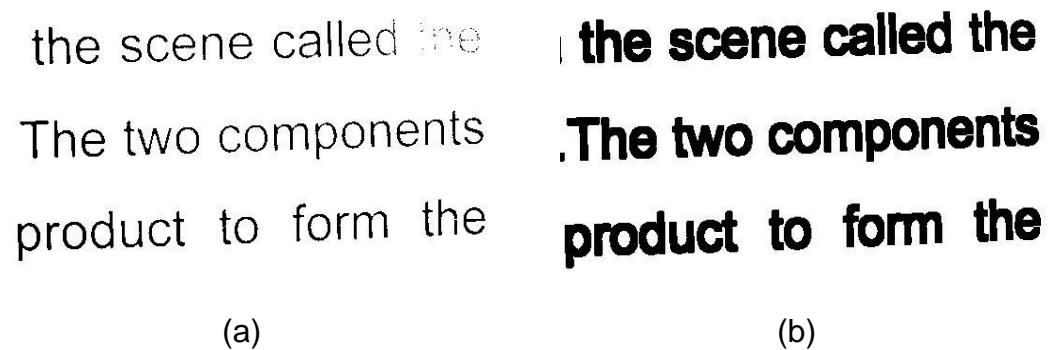
Figure 7.3: Spatial domain smoothing. (a) Input image. (b) Output image.

7.1.4 Unsharp Mask Sharpening

The result of applying a spatial domain unsharp mask sharpening to a greyscale image that has both poor and well illuminated regions is shown in Figure 7.4. The image has a well illuminated part at the bottom left corner unlike the rest of the image. The effects of the sharpening are highlighting of points, thin lines and edges in the image .These effects can be seen in the tusks of the animals, edges in the building and the cable hanging along the wall in figure 7.4 (b).

(a)   (b)

Figure 7.4: Image Sharpening. (a) Input image. (b) Sharpened image.

7.1.5 Image Interpolation

Figure 7.5 shows the results of different image interpolation operations on a greyscale image. The interpolations were performed using the in-built functions in MATLAB 7.14 that are based on the procedures outlined in section 2.5 of this thesis.

The image in figure 7.5 (a) is the input of size 480x 640 at a resolution of 8 bits. The second image is the result of applying optimal global thresholding on the input image using a threshold value of $k_{opt} = 116$. The image in part (c) shows the result of shrinking the image in part (b) to a size 120x160 followed by zooming it back to size 480x 640 using the nearest neighbour interpolation. The image in figure7.5 (d) shows the result of shrinking the image in part (b) to size 120x160 followed by zooming it back to size 480x 640 using the bilinear interpolation. The image in figure 7.5 (e) shows the result of shrinking the image in part (b) to size 120x160 and then zooming it back to size 480x 640 using the bicubic interpolation. Among the three interpolation results, the bicubic interpolation result compares best with the thresholded input

image with a cross correlation of 0.9624. The bilinear interpolated image gave a cross correlation coefficient of 0.9273 with the thresholded input image whereas the nearest neighbour interpolated image and the thresholded input image have a cross correlation of 0.8871.

automated image processing techniques. The principal objective of image enhancement is to modify attributes of an image to make it more suitable for a given

(a)

automated image processing techniques. The principal objective of image enhancement is to modify attributes of an image to make it more suitable for a given

(b)

automated image processing techniques. The principal objective of image enhancement is to modify attributes of an image to make it more suitable for a given

(c)

automated image processing techniques. The principal objective of image enhancement is to modify attributes of an image to make it more suitable for a given

(d)

automated image processing techniques. The principal objective of image enhancement is to modify attributes of an image to make it more suitable for a given

(e)

Figure 7.5: Image interpolation. (a) Input image. (b) Thresholded input image. (c) Nearest neighbour interpolated image. (d) Bilinear interpolated image. (e) Bicubic interpolated image.

7.1.6 Histogram Equalisation

Figure 7.6 shows the result of applying the histogram equalisation function in MATLAB 7.14 on a greyscale image. Figure 7.6 (a) shows the input image that is corrupted by shadows. One effect of histogram equalisation is to suppress the shadows as is evident on the top of the vehicle in figure 7.6 (b). A second effect is to increase the contrast as can be seen in the background objects and the number plate of the car. The third effect of histogram equalization is dynamic range compression this effect can be seen in the brightening of the grey regions of the vehicle and the person standing beside it. The histogram of the input and the enhanced images are given in parts (c) and (d) respectively. The histogram in part (d) has a better uniformity compared to that in part (c).

(a)   (b)
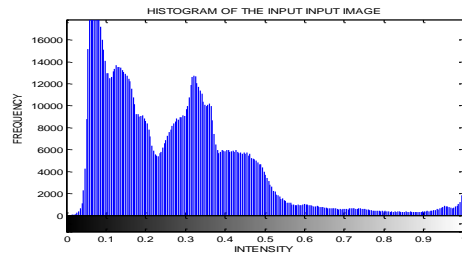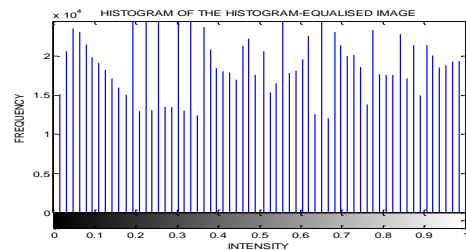
(c)   (d)

Figure 7.6: Histogram equalisation. (a) Input image. (b) Histogram equalised image. (c) Histogram of (a). (d) Histogram of (b).

7.1.7 Morphological Processing

Figure 7.7 shows the result of applying morphological processes on a greyscale image of a historical handwritten document (see appendix A3). In figure 7.7 (a), a thresholded image of the document is presented. The effects of dilation are thickening of the characters and bridging of small gaps in the characters as shown in figure 7.7 (b). The effects of erosion are shrinking of the characters as well as removal of thin lines as is evident in figure 7.7 (c). Opening has the effects of smoothing of the character boundaries, breaking of narrow isthmuses and elimination of small islands as can be seen in figure 7.7 (d). The effects of closing are fusion of narrow breaks in the image which is evident in figure 7.7(e).

[stamp: 15/11/36 / 7 OCT 1961 / Northern Province, Kenya]

Bedan M. Machuo.
To Mutt-u-ini Cottage
Dagretti P.o. Kikuyu.
7th October 1961

ag. Provicial Commissioner
Northern Province
P.o-Box 11 Isiolo.

Dear sir,
   I have the honour most humble
to Inform you that I received your letter
which, asked me to tell you where I want
to buy sheeps & goats and when.
   (A) I want to buy them at Isiolo auction
   (B) I want to come there during auctions

7.7 (a)

ag. Provincial Commissioner
Mohan Province
P.O. Box 11 Isiolo.
Dear sir,
I have the honour most humble
to inform you that I received your letter
which asked me to tell you where I want
to buy sheeps & goats and when.
(A) I want to buy them at ....
(B) I want to come there during auction

7.7 (c)

ag. Provincial Commissioner
Mohan Province
P.O. Box 11 Isiolo.
Dear sir,
I have the honour most humble
to inform you that I received your letter
which asked me to tell you where I want
to buy sheeps & goats and when.
(A) I want to buy them at ....
(B) I want to come there during auction
time only.

7.7 (d)

(e)

Figure 7.7: Morphological processing. (a) Input image. (b) Dilated image.
(c) Eroded image. (d) Opened image. (e) Closed image.

7.1.8 Single Scale Retinex

The effects of applying the single scale retinex algorithm on an image are shown in   figure 7.8. Figure 7.8 (a) shows the input image which is degraded by shadow effects. The SSR enhancement result given in figure 7.8 (b) shows a suppression of shadows as well as dynamic range compression which evident in the greying of the image. Figure 7.8 (c) shows the histogram of the input image which has a global variance of $\sigma_G^2 = 0.0654$. Figure 7.8 (d) shows the histogram of the SSR enhanced image. The histogram has a global variance of $\sigma_G^2 = 0.0063$ thereby showing a global variance reduction as a result of SSR enhancement.

(a)   (b)

(c)   (d)
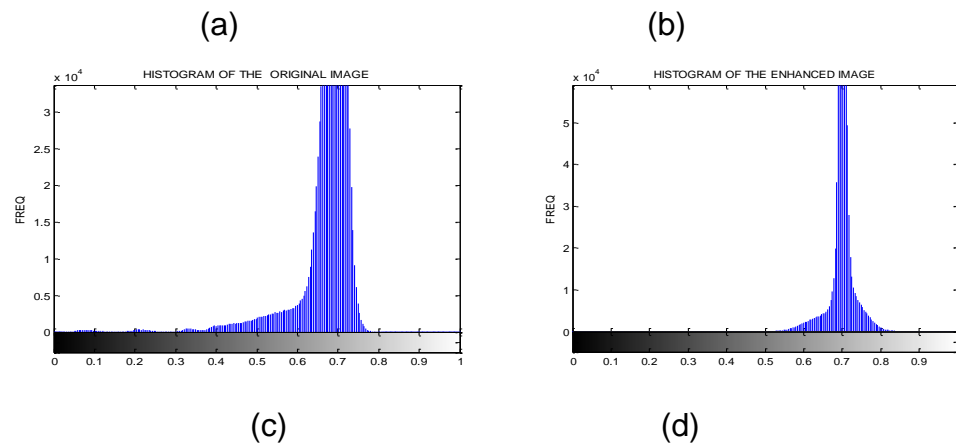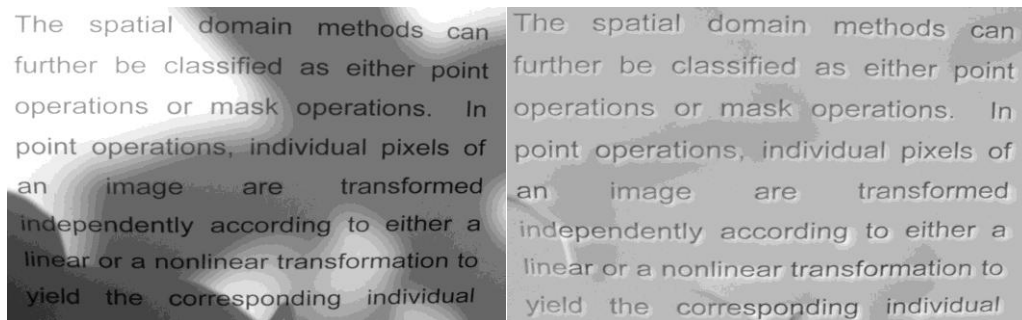
Figure 7.8: SSR enhancement. (a) Original image. (b) SSR enhanced image. (c) Histogram of the original image. (d) Histogram of the SSR enhanced image.

7.1.9 Optimal Global Thresholding

Figure 7.9 shows the results of the optimal global thresholding on two degraded input images. The greyscale document image in figure7.9 (a) was captured under low, uniform illumination conditions. The result of applying the Otsu's global thresholding on this image is shown in figure 7.9 (b). The threshold value used was $k_{opt} = 77.01$. The thresholding did not cause any loss in information.

Figure 7.9 (c) shows an input greyscale document image captured under conditions of non-uniform illumination. The result of applying the optimum global thresholding on this image using $k_{opt} = 128.0$ is shown in figure 7.9 (d). This result shows loss of information in the regions of the image covered by the shadows.

(a)   (b)

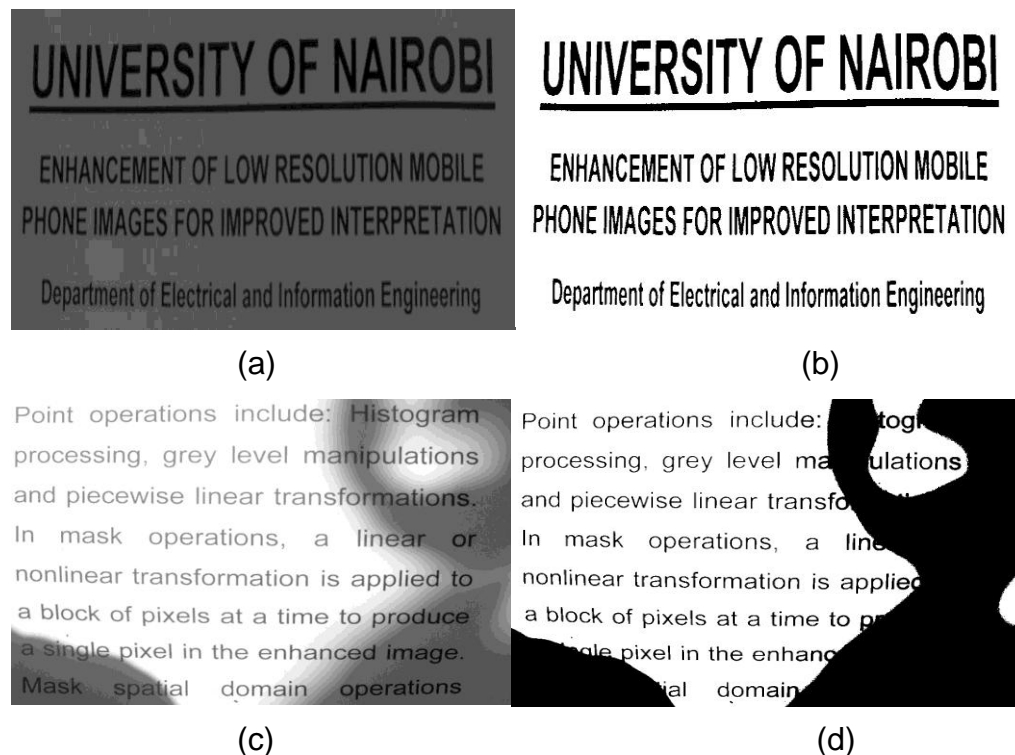(c)   (d)

Figure 7.9: Otsu's Thresholding. (a) Lowly illuminated image. (b) Thresholding result of (a). (c) Non-uniformly illuminated image. (d) Thresholding result of (c).

7.1.10 Local Thresholding

Figure 7.10 shows the results of applying the Sauvola's and the Niblack's local thresholding methods on a grey scale image degraded by shadows. The results were obtained using computer simulation of the methods discussed in section 2.7 using MATLAB 7.14. The result of the Sauvola's method presented in figure 7.10(b) has less background noise and better legibility than the Niblack's method result shown in figure 7.10 (c).
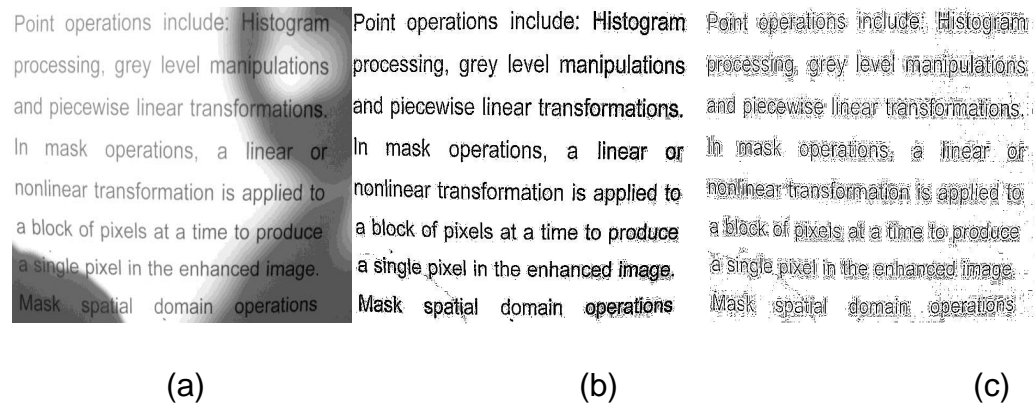

| (a) | (b) | (c) |

Figure 7.10: Local thresholding. (a) Input image (b) Thresholding result of the Sauvola's method, (c) Thresholding result of the Niblack's method.

7.2 Transform Domain Enhancement Results

In this section, computer simulation results of experiments performed in transform domain are presented. All the input test images used here have been interpolated to a size of 480x 640 pixels.

7.2.1 Homomorphic Filtering

The effects of applying homomorphic filtering on a greyscale image are shown in figure 7.11. The input image shown in figure 7.11 (a) was taken under conditions of low and non-uniform illumination conditions. The input image is homomorphic filtered using a MATLAB based computer simulation of the procedure outlined in section 4.2 to yield the image in part (b). The parameter values selected based on subjective tests are:

$\gamma_H = 1.6, \gamma_1 = 0.6, c = 0.8$ and $D_0 = 100.$

The filtered image has a higher contrast than the input as evidenced by the improvement in clarity of the objects in part (b). It also has a narrower dynamic range than the input due to suppression of the illumination component by the homomorphic filtering operation. Dynamic range compression is indicated by the overall brightening of the image in part (b).

(a)   (b)

Figure 7.11: Homomorphic filtering. (a) Input image. (b) Homomorphic filtered image.

7.2.2 Edge Sharpening

Figure 7.12 shows the results of performing edge-sharpening enhancement of an image in transform domain. The image in part (a) image is the blurred input, the one in part (b) shows the result of sharpening the input image using the unsharp mask filter while the image in part (c) shows the result of using the Laplacian filter on the input image. The result in part (b) shows that unsharp mask filtering sharpens the image edges without greatly affecting the texture (low frequency components) of the image. Laplacian filter sharpening yields an image that has highlighted edges although some low frequency information is lost.

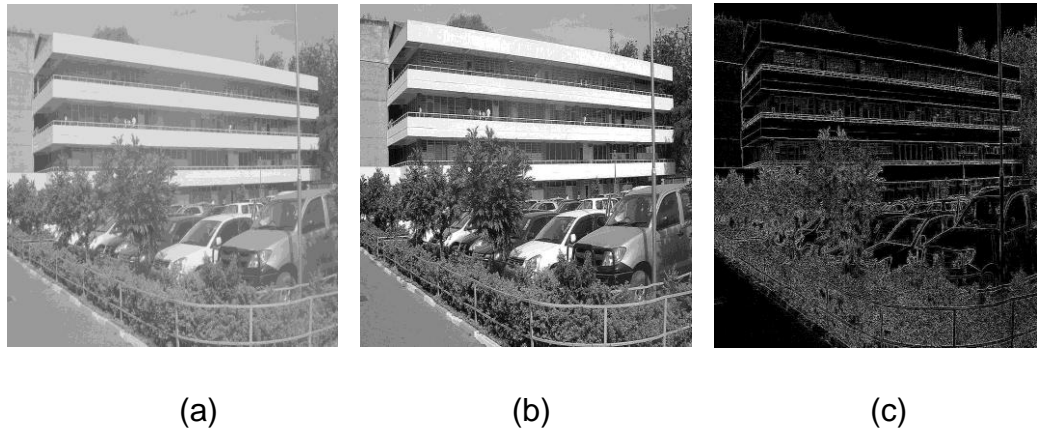|		
(a)	(b)	(c)

Figure 7.12: Edge sharpening. (a) Input image. (b) Unsharp mask filter sharpening result. (c) Laplacian filter sharpening result.
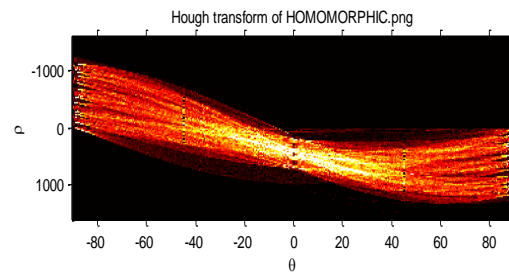
7.2.3 Hough Transform

The application of the Hough transform in detection of the longest lines in a faded greyscale document image is illustrated in figure 7.13. The document is a computer flow chart that has geometrical shapes as well as text. The image in part (a) is the input while the one in part (b) is the Hough transform of the input image. The image in part (c) shows the points that represent the five longest lines on the Hough transformed image. The image in figure 7.13 (d) shows the detected longest lines highlighted on the input image.
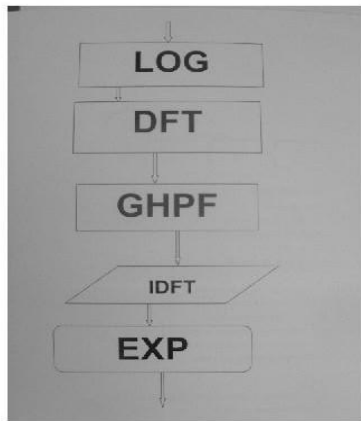
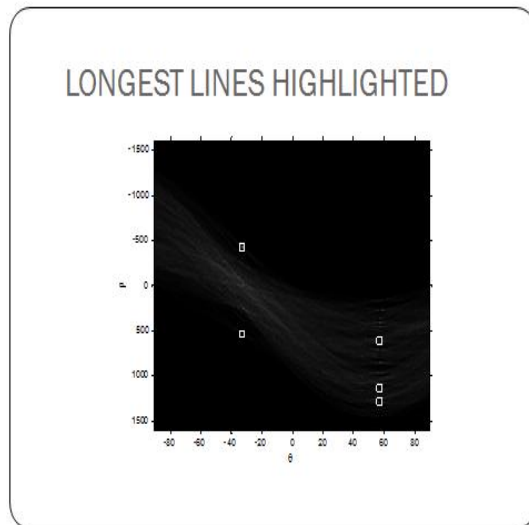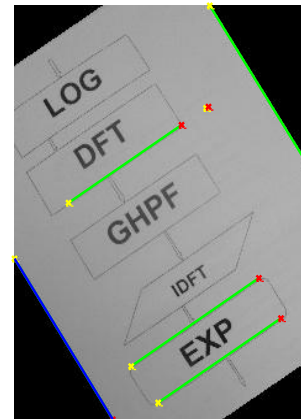(a)   (b)

(c)   (d)

Figure 7.13: Hough transform. (a) Input image. (b) Hough transform.
(c) Points representing longest lines highlighted on the transform.
(d) Longest lines detected.

7.2.4 Edge Detection

Figure 7.14 shows the result of applying the Sobel's edge detection algorithm on a grey scale image. The image in fig. 7.14 (a) is the input while the one in fig. 7.14 (b) is the output image showing the edges of the objects in the image.

(a)   (b)

Figure 7.14: Sobel filtering. (a) Input image. (b) Output image.

7.2.5 Image De-blurring

Figure 7.15 shows the results of image de-blurring on an image blurred by a uniform relative motion between the imaging device and the imaged scene. The blurring function was simulated in MATLAB using the model discussed in appendix A4. The parameters used in the model are: $a=b=8 \times 10^{-4}$ metres and $T=1$ second. These values were arrived at through experimentation. The de-blurring was carried out using a MATLAB computer simulation of the inverse filtering using the same parameter values used in the blurring function. Figure 7.15 (a) shows the ground-truth image while the one in part (b) is the blurred image. The cross correlation coefficient between the ground-truth image and the blurred image is 0.63 the image in figure 7.15 (c) shows the de-blurred output image. The de-blurring operation produces a reduction in image blurring degradation. The improvement is indicated by the increase in clarity and a

cross correlation coefficient of 0.94 between the ground-truth image and de-blurred image.
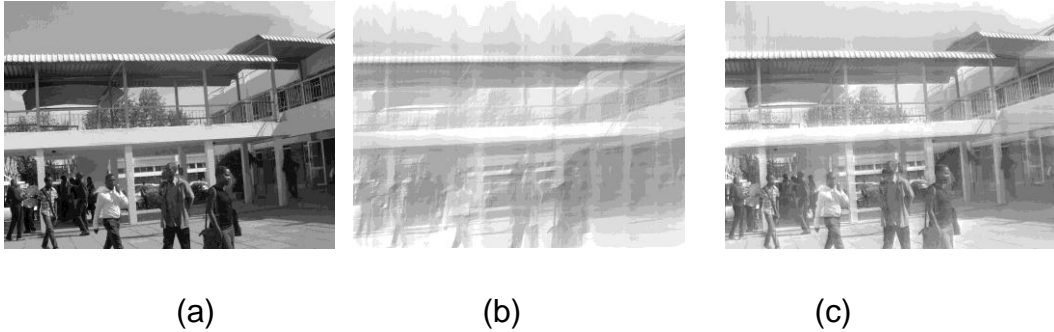

(a)   (b)   (c)

Figure 7.15: Image deblurring. (a) Original image. (b) Blurred input image. (c) Deblurred output image.

7.3 Geometrical Distortion Reduction

Figure 7.16 shows the result of geometrical distortion reduction on a warped document image. The enhancement is based on the 2-D document image processing method proposed by Stamatopolous et al. [61]. Figure 7.16 (a) shows the geometrically distorted input image. The image in figure 7.16 (b) is the result of distortion reduction using a computer simulation of the procedure described in chapter 5. The values of the parameters used are: L_{av}= 614pixels and h_d=36 pixels. The enhancement result in part (b) shows an improvement in the straightness of the text lines.
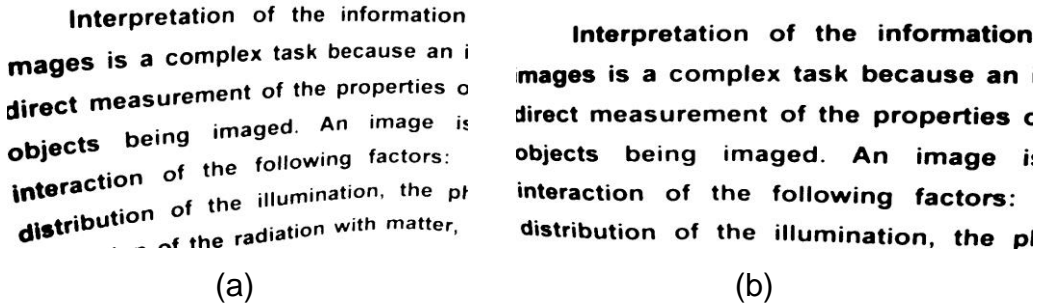

(a)   (b)

Figure 7.16: Geometrical degradation reduction. (a) Degraded input image. (b) Enhanced output image.

7.4 Proposed Method Results

A series of experiments were conducted using MATLAB simulation. The Otsu's binarisation function (greythresh) in MATLAB version 7 was used. Linear contrast stretching was used to expand the intensity range of the input image so that it spans over the entire range from 0 to 255. A 4x4 pixels square structuring element was used in morphological dilation. The parameters used in the single scale retinex algorithm are: $\alpha = 0.68, \beta = -0.72$ and $\sigma = 15$. These parameter values and structuring element were selected based on the perceived quality of the output image after performing a series of experiments with different parameters values and structuring element sizes and shapes.

7.4.1 Shadow Degradations

The results in figures 7.17 to 7.20 show the performance of the proposed algorithm on document images that are degraded by shadow components of different frequencies in comparison with the Otsu's method results. In each of the figures, the first image is the degraded input, the second shows the result of enhancing the input image using the proposed algorithm while the third shows the result of using Otsu's method on the same input. The results of the proposed algorithm show a significant improvement over the results of the Otsu's method in terms of shadow removal capability. For example in figure 7.17, the input image in part (a) has high frequency shadows which is indicated by multiple small shadows having step edge profiles. The result of the proposed algorithm in part (b) shows almost complete removal of these shadows. The result of using the Otsu's method given in part (c) shows a complete loss of the image information that is covered by the shadows.
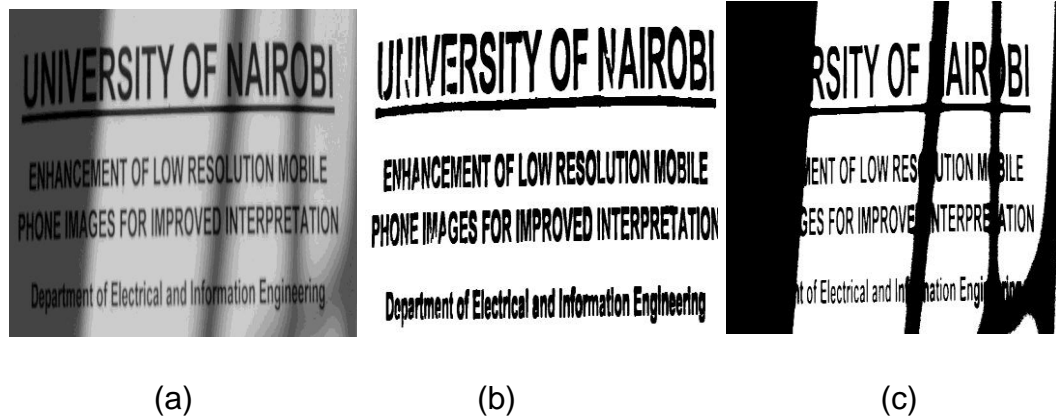
| (a) | (b) | (c) |

Figure 7.17: Enhancement of image with high frequency shadows. (a) Input image. (b) Proposed algorithm result. (c) Otsu's method result.
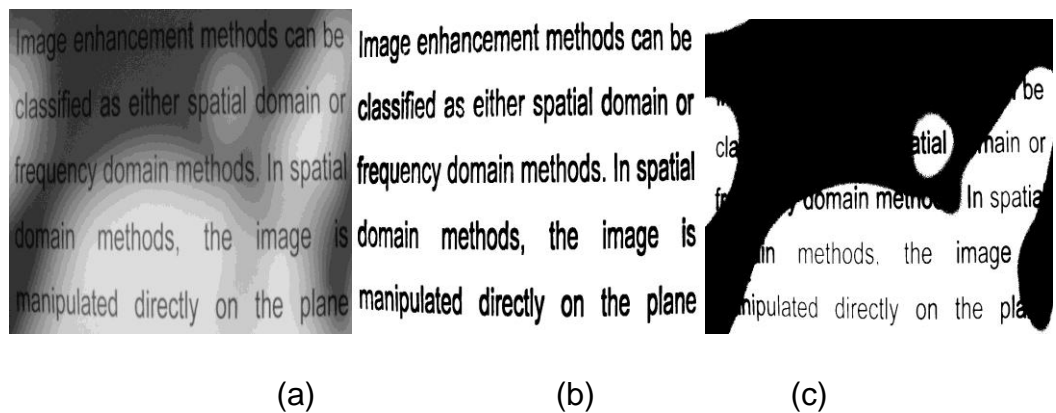

| (a) | (b) | (c) |

Figure 7.18: Enhancement of an image with low frequency shadows. (a) Input image. (b) Proposed algorithm result. (c) Otsu's method result.
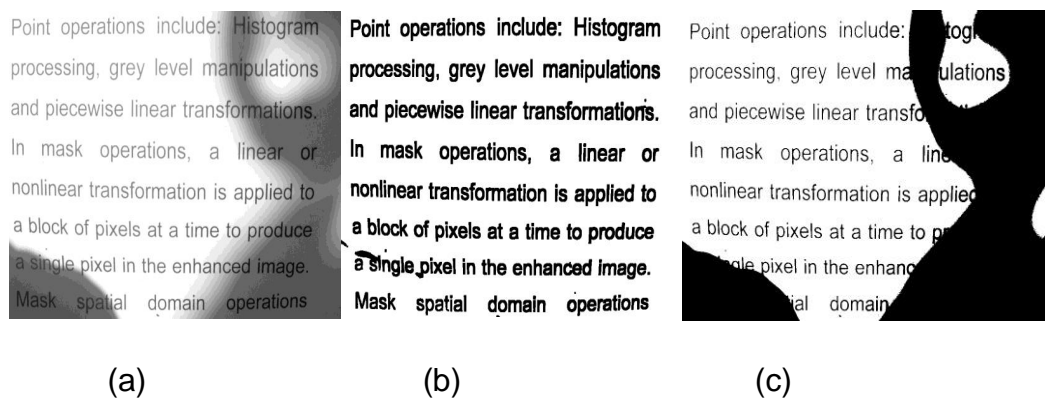

| (a) | (b) | (c) |

Figure 7.19: Enhancement of an image with medium frequency shadows. (a) Input image. (b) Proposed algorithm result. (c) Otsu's method result.

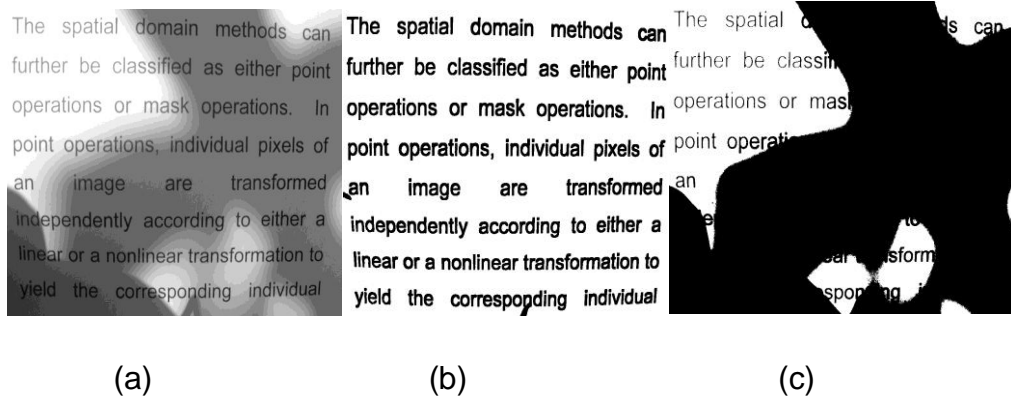(a)   (b)   (c)

Figure 7.20: Enhancement of an image with mixed shadows. (a) Input image. (b) Proposed algorithm result. (c) Otsu's method result.

7.4.2 Poorly Illuminated Images

The results in figure 7.21 show the performance of the proposed algorithm on a document image that is degraded by low level of illumination in comparison with that of the Otsu's method. The image in part (a) is the degraded input, the one in part(b) shows the result of enhancement of the input using the proposed algorithm while the image in part (c) shows the result of using Otsu's method on the input image. The results of the proposed algorithm show a slight improvement over the results of the Otsu's method in terms of processed document legibility as well as character stroke connectivity.
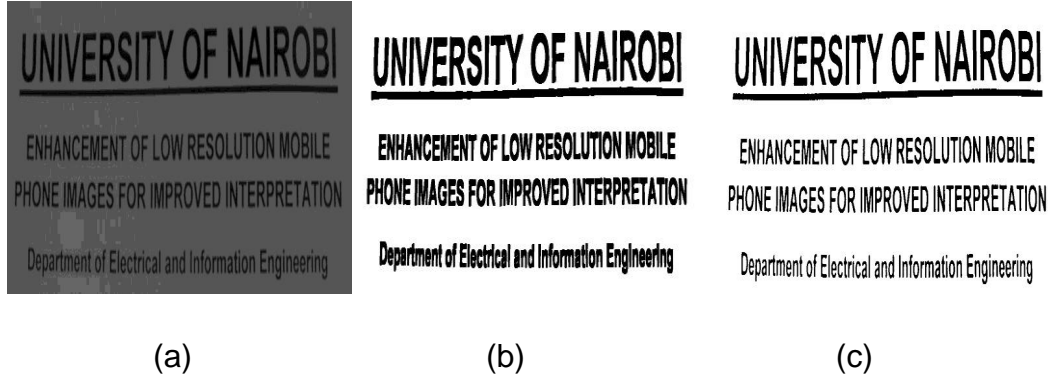

(a)   (b)   (c)

Figure 7.21: Enhancement of a poorly illuminated image. (a) Input image. (b) Proposed method result. (c) Otsu's method result.

116

7.4.3 Faded Images

The results in figures 7.22 to 7.25 show the performance of the proposed algorithm on document images that are degraded by fading in comparison with the Otsu's method.

In each of the figures, the image in part (a) is the degraded input, part (b) shows the result of enhancement of the input image using the proposed algorithm while the image in part (c) shows the result of using Otsu's method on the same input. The results of the proposed algorithm generally show a significant improvement over the results of the Otsu's method in terms of legibility of the processed image.

In figure 7.22, the input image in part (a) is that of a faded handwritten document with ink smears. The result of the proposed algorithm given in part (b) shows an improvement in the document legibility. Otsu's method result given in part (c) shows loss of most of the document information.

In figure 7.23, the input image given in part (a) is that of a handwritten historical document without ink smears. The proposed algorithm result given in part (b) shows an improvement over the Otsu's method result given in part (c) in terms of legibility.

The input images in parts (a) of figures 7.24 and 7.25 are those of heavily faded historical document. The results of the proposed method given in parts (b) show an improvement in document legibility. The results of using the Otsu's method shown in parts (c) indicate loss of most of the visual information.
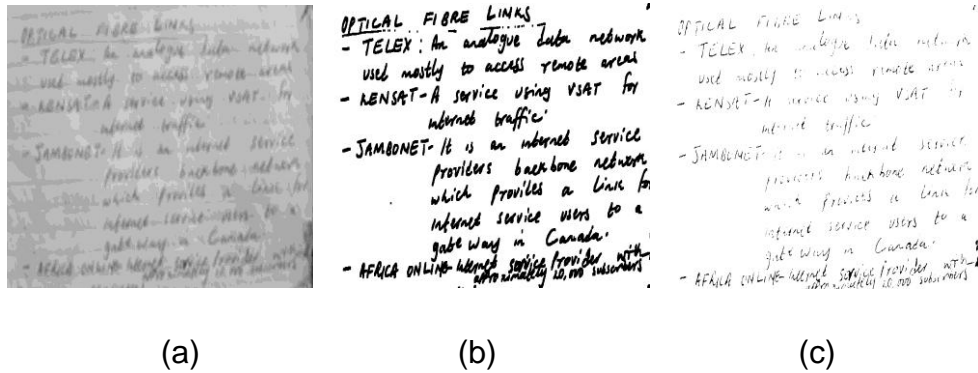
(a)   (b)   (c)

Figure 7.22: Enhancement of faded document image. (a) Input image. (b) Proposed algorithm result. (c) Otsu's method result.
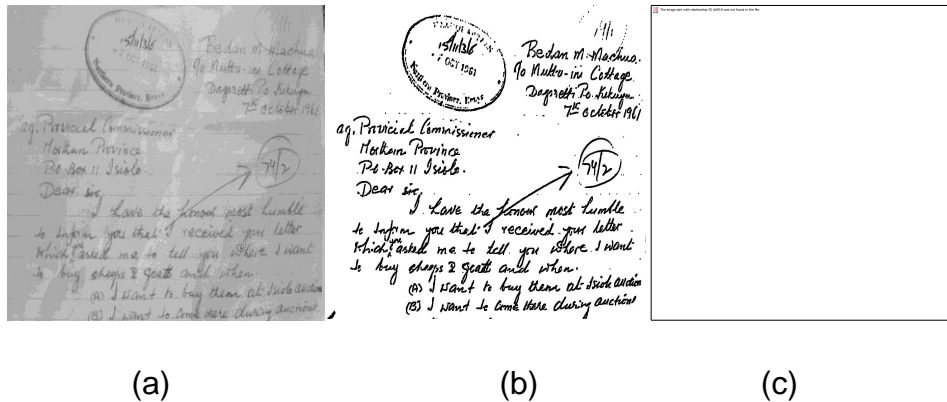

(a)   (b)   (c)

Figure 7.23: Enhancement of a historical handwritten document image. (a) Input image. (b) Proposed algorithm result. (c) Otsu's method result.
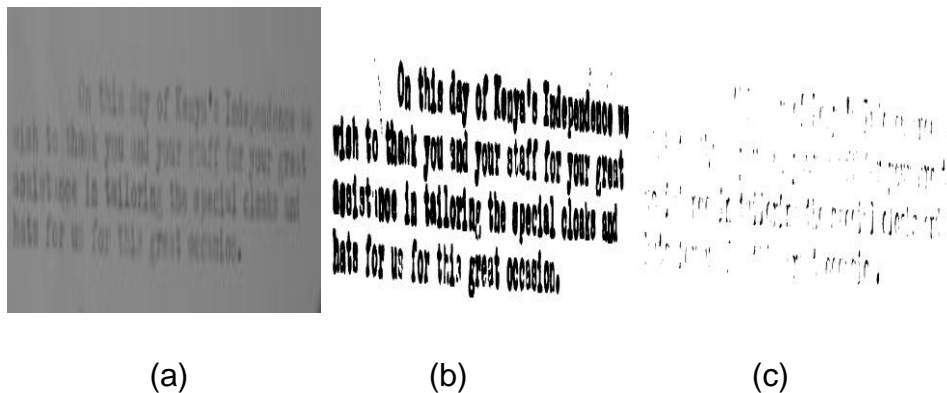

(a)   (b)   (c)

Figure 7.24: Enhancement of a historical typed document image. (a) Input image. (b) Proposed algorithm result. (c) Otsu's method result.
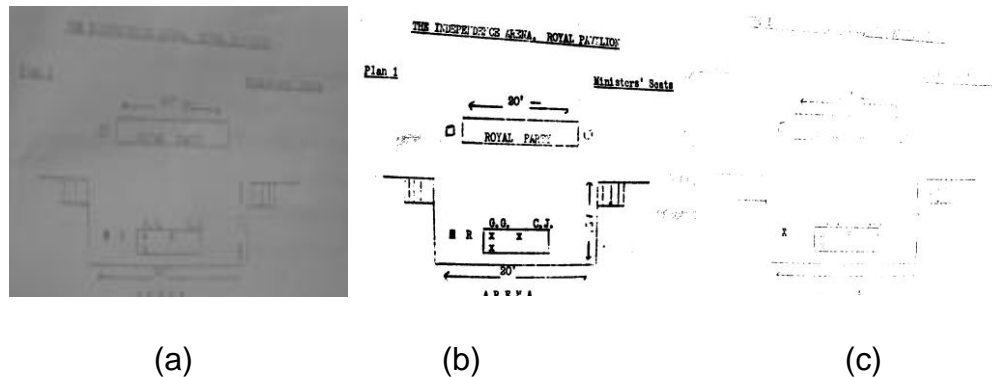
(a) (b) (c)

Figure 7.25: Enhancement of a document image with diagrams. (a) Input image. (b) Proposed algorithm result. (c) Otsu's method result.

7.4.4 Geometrically Distorted Document Images

The results in figure 7.26 show the performance of the proposed algorithm on geometrically distorted document images in comparison with the result obtained using the method proposed by Stamatopolous et al. [61]. Part (a) shows the geometrically distorted input image. The result of using the proposed method is shown in part (b). Using the proposed algorithm leads to an improvement in the straightness of the document text lines compared to the Stamatopolous method result shown in part (c).
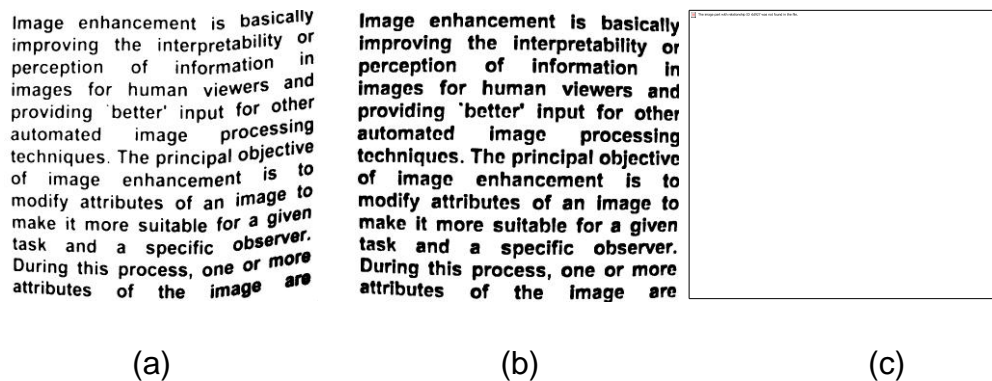

(a) (b) (c)

Figure 7.26: Geometrical degradation reduction. (a) Degraded input image. (b) Result of using the proposed method. (c) Result of using the Stamatopolous method.

7. 5 Performance Test Results

The results presented in this section demonstrate the effects of the proposed algorithm on the global variance and quality of images as well as on the performance of OCR software on the images.

7.5.1 Global Variance Tests

The results presented in this subsection demonstrate the effects of the SSR enhancement on the global variances of images degraded by shadows.

In all the figures 7.27 to 7.29, the image in part (a) is the degraded input image while the one in part (b) is the single scale retinex enhanced image. The results generally show that single scale retinex enhancement improves the background uniformity of such images which leads to improvement in thresholding. The input image in figure 7.27 (a) is corrupted by low frequency shadows. This is indicated by presence of shadow portions that are relatively larger than the image portions that are not covered by the shadows as well as the presence of ramp edge profiles. The enhanced image in part (b) shows suppression of most of the shadow components.

The input image in figure 7.28 (a) is corrupted by a mixture of low and high frequency shadows. The presence of both low and high frequencies in the shadows is indicated by the presence of both large and small shadow portions as well as step, ramp and roof-top edge profiles. The enhanced image in part (b) shows complete removal of most of the shadow components apart from very sharp edges as can be seen at the bottom left corner of figure 7.28 (b).

The input image in figure 7.29 (a) is corrupted by a high frequency shadow components. The presence of high frequencies in the shadows is indicated by the presence of relatively small shadow portions as well as step and sharp roof-top edge profiles. The enhanced image in part (b)

shows removal of most of the shadow components apart from the sharp edges.
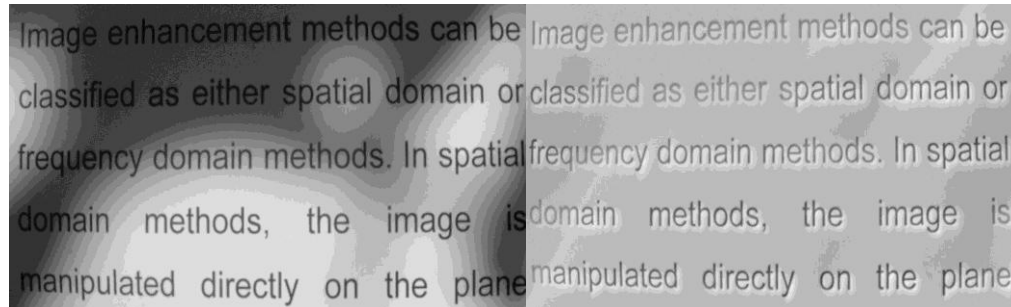
(a)   (b)

Figure 7.27: Low frequency shadow removal using SSR. (a) Input image. (b) SSR enhanced image.
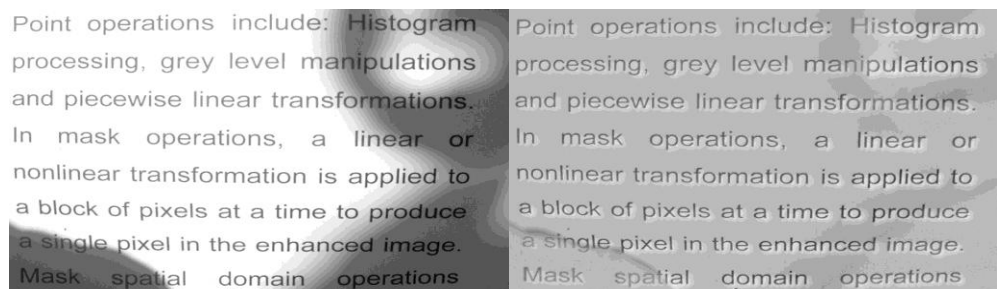
(a)   (b)

Figure 7.28: Mixed shadows removal using SSR. (a) Input image. (b) SSR enhanced image.

(a)   (b)

Figure 7.29: High frequency shadow removal using SSR. (a) Input image (b) SSR enhanced image.

The results presented in table 7.1 shows the global variances of the images in figures 7.27 to 7.29. The second column shows the global variances of the input images. The global variance values presented in the third column are those of the single scale retinex enhanced images. The results show a reduction in global variances as a result of the SSR enhancement which leads to a better selection of the global thresholding level.

Table 7.1. Reduction of global variance by SSR enhancement

Image	Global variance (σ_G^2)	Global variance (σ_G^2)
	Input image (a)	SSR enhanced image (b)
Fig. 7.27	0.0490	0.0073
Fig. 7.28	0.0730	0.0055
Fig. 7.29	0. 0464	0.0082

7.5.2 OCR Tests
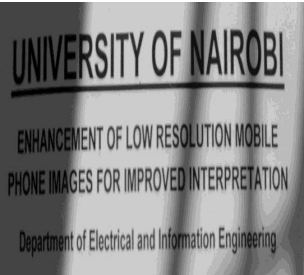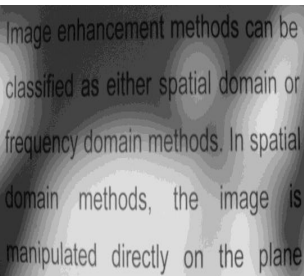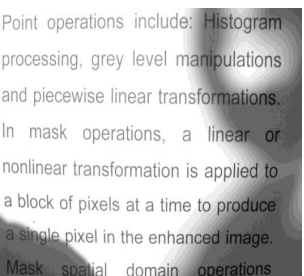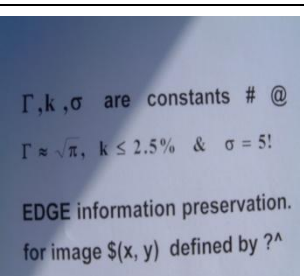
Table 7.2 shows the results of applying the ABBYY finereader optical character recognition software [68] on the results of the proposed method and those of the Otsu's method for images with shadows. From these results, a higher proportion of characters were recognised when the image is enhanced using the proposed method than when preprocessed using Otsu's method.

The results in the first row of the table show that, processing the input image using the proposed algorithm led to correct recognition of 132 out of the 133 characters while using Otsu's method, 78 characters were correctly recognised. For the image in the second row, the proposed enhancement method resulted in correct recognition of all 147 characters

while Otsu's method resulted in correct recognition of 57 characters. Results shown in the third row indicate that, the proposed algorithm resulted in correct recognition of 232 out of 262 characters while Otsu's method resulted in correct recognition of 164 characters of the 232 characters. The results presented in the fourth row of the table indicate that, the proposed algorithm resulted in correct recognition of 77 out of 88 characters while Otsu's method resulted in correct recognition of 63 of the 88 characters. In all the images, the OCR used did not recognise characters that are not used by the English language as well as image components such as lines.

Table 7.2. Optical character recognition results

Input image	Proposed method	Otsu's method

	results	results
	UNIVERSITY OF NAIROBI ENHANCEMENT OF LOW RESOLUTION MOBIL PHONE IMAGES FOR IMPROVED INTERPRETATION Department of Electrical and Information Engineering	SITY OF AIR BI ENT OF LOW RE$ TION M ILE ES FOR IMPROV A TERPR TION tof Electrical and In { ation Engi » =
	Image enhancement methods canbe classified as either spatial domain or frequency domain methods. In spatial domain methods, the image is manipulated directly on the plane	é ¢| M éin ori min - ; _ln spati A methods. the image ipulated directly on the pl
	Point operations include: Histogram processing, grey level manipulations and piecewise linear transformations. In mask operations, a linear or nonlinear transformation is applied to \ a block of pixels at a time to produce afilngleplxel in the enhanced image. Mask spatial d_omaln opemtivfls	Point operations include: 0g processing, grey level m latiorls and piecewise linear transf .In mask operations, a lin nonlinear transformation is applie A a block of pixels at a time to ' . ixel in the enhan 33' ma'
	F,k are constants # @ k < 2.5% & = 5! EDGE information preservation. for image $(x, y) defined by ?^	, constants # @ < 2.5% & = 5! Information preservation. image $(x, y) defined by ?^

7.5.3 Quality Tests

The results in figure 7.30 shows the performance of the proposed algorithm on document images that are degraded by shadows in comparison with the Otsu's, Sauvola's and Niblack's methods. The image in part (a) is the input degraded image while that in part (b) is the result of enhancement using the proposed method. The result of applying the Otsu's, Niblack's and Sauvola's binarisation methods on the input image are presented in parts (c), (d) and (e) respectively.

Applying quality measures on these images yielded the results summarised in table 7.3. The subjective test was performed using twenty subjects who were asked to choose the best output image from amongst the outputs of the four enhancement methods. None of the subjects chose the result of Otsu's method. Four of the subjects chose the result of Niblack's method while seven subjects voted for the Sauvola's method result. The output image of the proposed algorithm was selected as the best one by nine subjects. These subjective results are included in table 7.3 as percentages of the total number of subjects interviewed.

All the quality metrics used show that the proposed method produced improved results when compared to the other methods. It was followed by the Sauvola's, the Niblack's and the Otsu's methods respectively.

Point operations include: Histogram processing, grey level manipulations and piecewise linear transformations. In mask operations, a linear or nonlinear transformation is applied to a block of pixels at a time to produce a single pixel in the enhanced image. Mask spatial domain operations

(a)

Point operations include: Histogram processing, grey level manipulations and piecewise linear transformations. In mask operations, a linear or nonlinear transformation is applied to a block of pixels at a time to produce a single pixel in the enhanced image. Mask spatial domain operations

(b)

Point operations include: tog processing, grey level ma ulations and piecewise linear transf In mask operations, a lin nonlinear transformation is applie a block of pixels at a time to p le pixel in the enhan ial domain

(c)

Point operations include: Histogram processing, grey level manipulations and piecewise linear transformations. In mask operations, a linear or nonlinear transformation is applied to a block of pixels at a time to produce a single pixel in the enhanced image. Mask spatial domain operations

(d)

Point operations include: Histogram processing, grey level manipulations and piecewise linear transformations. In mask operations, a linear or nonlinear transformation is applied to a block of pixels at a time to produce a single pixel in the enhanced image. Mask spatial domain operations

(e)

Figure 7.30: Comparison between enhancement methods. (a) Input image (b) Proposed algorithm result. (c) Otsu's method result. (d) Niblack's method result. (e) Sauvola's method result.

Table 7.3. Image quality measurement results

Quality metric	Otsu's Method	Niblack's Method	Sauvola's Method	Proposed Method
PSNR	12.4390	16.3325	17.3792	18.8432
NSLE	0.2843	0.2793	0.1982	0.1655
CORR	0.9027	0.9428	0.9533	0.9621
MI	1.5361	1.6314	1.7722	1.8626
PETROVIC-XYDEAS	0.2938	0.3107	0.3847	0.4018
WANG-BOVIK	0.3147	0.4216	0.4817	0.5382
SUBJECTIVE	0.0000	20.00	35.00	45.00
RMSE	26.3915	25.3295	23.7390	22.3616

CHAPTER 8

CONCLUSIONS AND RECOMMENDATIONS

8.1 Conclusion

In this thesis report, an image enhancement procedure has been proposed which addresses the effects of poor resolution, fading, noise, shadows, poor illumination, geometrical distortions and ink smears simultaneously. The effectiveness of the algorithm has also been illustrated with some test results for each of these degradations. Both subjective and objective image quality measurement results show that the proposed enhancement method produces higher quality output images than the other enhancement methods considered here.

Results of optical character recognition experiments show that the proposed method performs better in enhancing degraded document images than the other enhancement procedures considered here. The OCR results also compare quite favourably with those obtained using the state-of-the-art Sauvola's method.

Test results on faded handwritten document images show that the proposed algorithm improves the legibility of the documents and can therefore be used as a pre-process in handwriting recognition as well as in other image analysis and understanding processes.

Global variance test results show that the single scale retinex algorithm reduces the global variance of the images leading to better binarisation results for images degraded by shadows than when the Otsu's method is applied on the original image.

Results on geometrically distorted document images show that the proposed method leads to improved text lines straightness which leads to better legibility. However, the proposed method does not perform very well on images degraded by severe geometrical distortions as well as heavy fading and very low intensity shadows.

8.2 Recommendations

Further research is recommended in order to improve the performance of the proposed algorithm especially on images that are heavily degraded by shadows, fading and document warping.

Optimisation in the selection of the parameters used in the SSR algorithm is recommended. This may improve the shadow reduction effect of the algorithm leading to better performance on heavily shadowed images.

Research on better selection of the structuring element used in the morphological dilation stage is also recommended. Better selection of the SE shall result in better bridging of the character gaps resulting from the thresholding of heavily faded images.

Further research work is also recommended on rectification of heavily warped document images. Rectification based on 3-D shape reconstruction can be used to improve the results of geometrical distortion reduction. These techniques rely upon the extraction of 3-D information of the document images and can be divided into two sub-categories. Techniques in the first sub-category obtain the 3-D shape of the document image using special equipment such as laser scanners, stereo cameras, and structured light setups. On the other hand, techniques in the second subcategory reconstruct the 3-D document model from information existing in the document image. Both sub-categories are recommended for further investigation.

To improve the results of the thresholding step, use the watershed segmentation algorithm as well as Markov Random Fields (MRF) to model the correlations between intensities of neighbouring pixels is recommended.

REFERENCES

1. R. C. Gonzalez and R.E. Woods, *Digital Image Processing. 3rd edition*, New Jersey, USA: Addison-Wesley, 2008.

2. W. K. Pratt, *Digital Image Processing, 4th edition.* New Jersey, USA: John Wiley & sons, 2007.

3. N. P. Galatsanos, C. A. Segall, and A.K. Katsaggelos "Digital image enhancement," *Encyclopedia of optical engineering,* pp. 388-402, 2003.

4. B. K. Gunturk et al., "Demosaicking: color filter array interpolation," *IEEE signal processing magazine*, pp. 44-54, January 2005.

5. R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image processing using MATLAB*, Englewood cliff, USA: Prentice-Hall, 2003.

6. A. K Jain, *Fundamentals of Digital Image Processing*, Englewood cliff, New Jersey, USA: Prentice Hall, 1980.

7. H. Kiragu and E. Mwangi, "An improved enhancement of degraded binary text document images using morphological and single scale retinex operations," *The IET image processing conference 2012*, University of Westminster, London, UK, July 2012.

8. B. Gatos, I. Pratikakis, and S.J. Perantonis, "Adaptive degraded document image binarization", *Pattern Recognition Journal, Vol.* 39, No.3, pp. 317-327, 2006.

9. N. Mokhtar et al., "Image enhancement techniques using local, global, bright, dark and partial contrast stretching for acute leukemia images," *Proceedings of the World Congress on Engineering*, Volume 1, WCE, 2009, London, U.K., 2009.

10. D. Vendhan, and R. Selvakumar, "Color contrast enhancement based on fuzzy model for thinprep cervical cell images," *IEEE international conference on computational intelligence and computing research*, 2010.

11. F. Bin, C. Jianwen, and X. Xinyan, "A novel adaptive vein image contrast enhancement method based on fuzzy and Retinex theory," *Proceedings of the IEEE International Conference on Information and Automation*, Harbin, China, pp 2447-2450, June, 2010.

12. O. N. Ucan and E Oguslu, "A non-linear technique for the enhancement of extremely non-uniform lighting images," *Journal of Aeronautics and Space Technologies,* volume 3, number 2, pp.37-47, 2007.

13. Z. Mei, H. Ren, and B. Dong, "An edge detection algorithm for the image of log end based on homomorphic enhancement," *International Conference on Computer Applications and System Modeling,* Northeast Forestry University, China, pp V2-54-V2-56, 2010.

14. N. Otsu, "A threshold selection method from gray-level histograms," IEEE Transactions. Syst., Man, Cybern, vol. SMC-9, pp. 62–66, 1979.

15. B. Gatos, I. Pratikakis, and S.J. Perantonis , "Improved document image binarization by using a combination of multiple binarization techniques and adapted edge information," *Computational Intelligence Laboratory, Institute of Informatics and Telecommunications, National Center for Scientific Research "Demokritos"*, Athens, Greece, 2008.

16. S. S. Bukhari, F. Shafait, and T. M. Breuel "Adaptive binarization of unconstrained hand-held camera-captured

document images," *Journal of Universal Computer Science, vol. 15, no. 18*, 2009.

17. J. Wu and C.Tang, "PDE-based random-valued impulse noise removal based on new class of controlling functions," *IEEE transactions on image processing*, vol. 20, no. 9, September 2011.

18. B. Gatos, I. Pratikakis, and I. Ntirogiannis, "Segmentation based recovery of arbitrarily warped document images," *International Conference on Document Analysis and Recognition,* Curitiba, Brazil, pp. 989-993, September, 2007.

19. A. Masalovitch, and L. Mestetskiy, "Usage of continuous skeletal image representation for document images dewarping," *In 2nd Int. Workshop on Camera-Based Document Analysis and Recognition*, Curitiba, Brazil, pp. 45-53, September, 2007.

20. S. V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction, Southern Gate*, Chichester, West Sussex PO19 8SQ, England: John Wiley & Sons Ltd, 2006.

21. N. Stamatopoulos, B. Gatos, I. Pratikakis and S. J.Perantonis, "A two-step dewarping of camera document images," *8th International Workshop on Document Analysis Systems*, Nara, Japan, pp. 209-216, September,2008*.*

22. O. Lavialle, X. Molines, F. Angella, and P. Baylou, "Active contours network to straighten distorted text lines," *in Proc. Int. Conf. Image Processing,* Thessaloniki, Greece, pp. 748–751, October, 2001.

23. C. Wu and G. Agam, "Document image de-warping for text/ graphics recognition," *in Joint IAPR Int. Workshop Structural,*

Syntactic and Statistical Pattern Recognition, Windsor, Canada, pp. 348–357, August, 2002.

24. L. Zhang, and C.L. Tan, "Warped image restoration with applications to digital libraries," *In Proc. Eighth Int. Conf. on Document Analysis and Recognition*, Seoul, South Korea, August, 2005.

25. H. Ezaki, S. Uchida, A. Asano, and H. Sakoe, "Dewarping of document image by global optimization," *in Proc. 8th Int. Conf. Document Analysis Recognition*, Seoul, Korea, pp. 302–306, September, 2005.

26. L. Mischke and W. Luther, "Document image de-warping based on detection of distorted text lines," *in Proc. Int. Conf. Image Analysis and Process*ing, Cagliari, Italy, pp. 1068–1075*, September, 2005.

27. A. Ulges and, C.H. Lampert, and T.M. Breuel, "Document image dewarping using robust estimation of curled text lines," *Proc. ICDAR'05,* 1001-1005, Seoul, South Korea, August, 2005.

28. S. J. Lu, B. M. Chen, and C. C. Ko, "A partition approach for the restoration of camera images of planar and curled document," *Image and Vision Computing, vol. 24, no. 8*, pp. 837–848, 2006.

29. D. C. Schneider, M. Block, and R. Rojas, "Robust document warping with interpolated vector fields," *in Proc. 9th Int. Conf. Document Analysis and Recognition,* Curitiba, Brazil, pp. 113–117*, September, 2007.

30. S. S. Bukhari, F. Shafait, and T. M. Breuel, "Dewarping of document images using coupled-snakes," *in Proc. Int.*

Workshop Camera-Based Document Anal. Recognition, Barcelona, Spain, pp. 34–41, July, 2009.

31. B. Fu, M. Wu, R. Li, W. Li, Z. Xu, and C. Yang, "A model-based book dewarping method using text line detection," *in Proc. Int. Workshop Camera-Based Document Anal. Recognition,* Curitiba, Brazil, pp. 63–70, September, 2007.

32. Y. Zhang, C. Liu, X. Ding, and Y. Zou, "Arbitrary warped document image restoration based on segmentation and thin-plate splines," *in Proc. 19th Int. Conf. Pattern Recognition,* pp. 1–4, Florida, USA, December, 2008.

33. M. A. Schulze, "Biomedical image processing with morphology-based nonlinear filters," PhD *thesis,* The University of Texas at Austin, USA, 1994.

34. X. Guoxin, X. Pei, and L. Qiang, "A method to improve the retinex image enhancement algorithm based on wavelet theory," *International Symposium on Computational Intelligence and Design,* Changzhou University, China, pp 182-185, October, 2010.

35. W. Niblack, *An Introduction to Digital Image Processing*, Prentice-Hall, Englewood Cliffs, NJ: 1986.

36. J. Sauvola, T. Seppanen, S. Haapakoski, and M.Pietikainen, "Adaptive document binarization", *ICDAR '97*, pp. 147-152, August, 1997.

37. B. Gatos, I. Pratikakis, and S.J. Perantonis, "Adaptive degraded document image binarization", *Pattern Recognition,* 39(3): pp. 317-327, 2006.

38. S. Wang and Y. Wang, "Shadow detection and compensation in high resolution Satellite image based on retinex," *Fifth*

International Conference on Image and Graphics, Wuhan University, China, pp 209-212, September, 2009.

39. J. Serra, *Image Analysis and Mathematical Morphology, Vol. 2: Theoretical Advances*; Academic Press, 1988;

40. S. Samantaray "A DoG based approach for fingerprint image enhancement," MSc *thesis,* Department of Computer Science and Engineering National Institute of Technology Rourkela, Orissa, India, 2011.

41. M. Q. Hernandez, "Genetic programming applied to morphological image processing," *PhD thesis,* The University of Birmingham, UK, 2004.

42. E. Land, "An alternative technique for the computation of the designator in the retinex theory of color vision," *Proceedings of the National Academy of Science*, 83, pp. 3078–3080, May, 1986.

43. D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Trans. on Image Processing 6*, pp. 451–462, March 1997.

44. Z. Rahman, D. Jobson, and G. Woodell, "Retinex processing for automatic image enhancement," *Journal of Electronic Imaging*, pp. 100–110, January 2004.

45. S. Mallat, *A Wavelet Tour of Signal processing the Sparse Way*, 3rd edition, Elsevier, Burlington, USA, 2009.

46. L. Zhang, Y. Zhang, and C. L. Tan, "An improved physically-based method for geometric restoration of distorted document images," *IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 4,* pp. 728–734, April, 2008.

47. A. Yamashita, A. Kawarago, T. Kaneko, and K. T. Miura, "Shape reconstruction and image restoration for non-flat surfaces of document with a stereo vision system," *in Proc. 17th Int. Conf. Pattern Recognit.,* Cambridge, U.K., pp. 482–485*,* August, 2004.

48. M. S. Brown and W. B. Seales, "Image restoration of arbitrarily warped documents," IEEE *Trans. Pattern Anal. Mach. Intell., vol. 26, no. 10,* pp. 1295–1306, October. 2004*.*

49. H. Cao, X. Ding, and C. Liu, "Rectifying the bound document image captured by the camera: A model based approach," *in Proc. 7th Int. Conf. Document Anal. Recognition,* pp. 71–75*,* Edinburg, Scotland, UK, August, 2003.

50. J. Liang, D. DeMenthon, and D. Doermann, "Geometric rectification of camera-captured document images," *IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 4,* pp. 591–605, April. 2008*.*

51. C. L. Tan, L. Zhang, Z. Zhang, and T. Xia, "Restoring warped document images through 3-D shape modeling," *IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 2,* pp. 195–208, February, 2006*.*

52. L. Zhang, A. M.Yip, M. S. Brown, and C. L. Tan, "A unified framework for document restoration using in painting and shape-from-shading," *Pattern Recognition Journal,* vol. 42, no. 11, pp. 2961–2978, 2009.

53. O. Okun, M. Pietikainen, and J. Sauvola, "Document skew estimation without angle range restriction," *IJDAR 2,* pp. 132 - 144, 1999.

54. N. Tripathy and U. Pal, *"*Handwriting segmentation of unconstrained Oriya text," *in International Workshop on*

Frontiers in Handwriting Recognition, pp. 306–311, October, 2004.

55. C. Weliwitage, A. L. Harvey, and A. B. Jennings, "Handwritten document offline text line segmentation," *in Proceedings of Digital Imaging Computing: Techniques and Applications*, pp. 184-187, Cairns, Australia, December, 2005.

56. U. Pal, S. Datta, "Segmentation of Bangla unconstrained handwritten text," Proceedings *of Seventh International Conference on Document Analysis and Recognition,* pp 1128–1132, Edinburg, UK, August, 2003.

57. M. S. Brown and Y. C. Tsoi, "Geometric and shading correction for images of printed materials–A unified approach using boundary," *Conference on Computer Vision and Pattern Recognition,* pp. 240-246, Washington, DC, USA, June/July, 2004*.*

58. E. Kavallieratou, "Text line detection and segmentation: uneven skew angles and hill-and-dale writing," Journal *of Universal Computer Science, vol. 17, no. 1,* pp 16-29, 2010.

59. N. Stamatopoulos, B. Gatos, and I. Pratikakis, "A methodology for document image dewarping techniques performance evaluation," *in Proc. 10th Int. Conf. Document Anal. Recognition,* Barcelona, Spain, pp. 956–960, July, 2009.

60. R. Dharshana and K. Harada, "Connected component labeling algorithms for gray-scale images and evaluation of performance using digital mammograms," *IJCSNS International Journal of Computer Science and Network Security, vol.8 no.6,* June   pp 33-41, 2008.

61. N. Stamatopoulos, B. Gatos, and I. Pratikakis, "Goal-oriented rectification of camera-based document images," *IEEE*

transactions on image processing, vol. 20, no. 4, pp. 910-920, April 2011.

62. R. D. Yapa and K. Harada, "Connected component labeling algorithms for gray-scale images and evaluation of performance using digital mammograms," *International Journal of Computer Science and Network Security*, vol.8 no.6, June 2008.

63. U. V. Marti and H. Bunke, "Using a statistical language model to improve the performance of an HMM-based cursive handwriting recognition system*," Int. J. Pattern Recognit. Artif. Intell., vol. 15, no. 1,* pp. 65–90, 2001.

64. A. Vinciarelli, S. Bengio and H. Bunke, "Offline recognition of unconstrained handwritten texts using HMMs and statistical language models," *IEEE transactions on pattern analysis and machine intelligence, vol. 26,* pp 33-41, June 2004.

65. 1680. Nokia. http://www.nokia.com, 24th October 2012.

66. Fujifilm 4800Z. Fujifilm Global. http://www.fujifilm.com, 24th October 2012.

67. MATLAB version 7.14. Computer software. Math works, 2009.

68. ABBYY FineReader OCR, http://finereader.abbyy.com/, 24th October, 2012.

69. BookRestorer, [Online]. Available: http://www.i2s-bookscanner.com/, 24th October, 2012.

70. The OCROpus Open Source Document Analysis System, [Online]. Available: http://code.google.com/p/ocropus/, 24th October, 2012.

71. V. Petrovic and C. S. Xydeas, "Objective image fusion performance measure," *Proceedings of SPIE*, vol. 4051, pp. 88–99, April 2004.

72. P. Gemma, and H. Heink, "A new quality metric for image fusion," *CWI, Kruislaan 413*, 1098, SJ Amsterdam, The Netherlands.

73. Z. Wang, and A.C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters, vol. 9, no. 3*, pp. 81–84, March 2002.

APPENDIX

This appendix is divided into four parts. Part A1 contains a paper that was published during this research [7] and information about the conference. Part A2 presents the MATLAB code simulations that were used to generate the results presented in chapter 7 of this thesis. Part A3 contains three ground-truth images of heavily faded images. These images have been used as inputs in testing of the proposed algorithm. In part A4 an image blurring model is presented.

The IET Image Processing Conference 2012

3 - 4 July 2012 | University of Westminster, London, UK

A1 Published Paper

AN IMPROVED ENHANCEMENT OF DEGRADED BINARY TEXT DOCUMENT IMAGES USING MORPHOLOGICAL AND SINGLE SCALE RETINEX OPERATIONS

Henry Kiragu, Elijah Mwangi

Department of Electrical and Electronic Engineering

University of Nairobi, Kenya, P.O. BOX 30197-00100

Nairobi, Kenya.

Emails: hencharia@yahoo.com, elijah.mwangi@uonbi.ac.ke

Keywords: Enhancement, degradation, single scale retinex, binarisation, morphological dilation.

Abstract

This paper proposes a novel enhancement algorithm for degraded binary document images captured using a low end cell phone camera. The algorithm starts with colour to grey scale conversion followed by contrast stretching. Then single scale retinex enhancement is performed to increase the contrast, reduce global variance and improve the background uniformity of the image. Otsu's binarisation method is then used to dichotomize the image and finally, morphological dilation is performed to preserve stroke connectivity of the document characters by bridging any gaps resulting from the thresholding process. Computer simulation experiments have been used to demonstrate the effectiveness of the proposed algorithm. The results show a significant improvement over the Otsu's method in terms shadow removal, document legibility and optical character recognition. The results also compare well with the state of the art Sauvola's method results.

1 Introduction

Enhancement of binary document images is used as a preprocessing step in many document image processing operations such as image dewarping, Optical Character Recognition (OCR) and handwriting recognition. Accurate binarisation is critical to the success of these image processes. Techniques used in image binarisation can be classified as either global or local techniques. In global techniques, a single threshold value is calculated and used to dichotomize the image [5, 11], while in local methods, local image information is used to calculate the threshold value for the image pixels within an defined window [3, 4, 9, 10, 12, 14].

Degradations in document images caused by fading, poor resolution of the imaging device, noise, shadows, poor illumination, dirt and ink smear on the text paper may lead to erroneous determination of the threshold level. Various global and adaptive local binarisation methods have been proposed. The optimal global thresholding method proposed by Otsu gives a procedure for determining an optimum threshold level for binarising an image by maximizing the between-class variance of the image [11]. However the method does not address the issue of non-uniform illumination, fading, gaps in the text characters and the possibility of adjusting the global variance in the calculation of the threshold value. Adaptive thresholding techniques such as those reported in [2, 9, 10, 12] generally give good binarisation results but they mostly require adjustment of some parameters [4].

In this paper, we propose an enhancement technique which preserves most of the textual information in degraded binary document images. The method can be applied automatically without requiring parameter adjustments. We have also demonstrated that it is possible to obtain binarisation resultsthat are comparable to those obtained using the state of the art methods and better than those obtained using Otsu's method.

The rest of this paper is organized as follows: section 2 outlines the requisite theory that is utilized in developing the proposed algorithm. The theory includes: contrast stretching, homomorphic filtering, single scale retinex algorithm, Otsu's thresholding method and morphological dilation. Section 3 presents the proposed document image enhancement algorithm while section 4 presents some test results obtained using the proposed algorithm. Finally, section 5 gives a conclusion and suggestions for future work.

2 Background

2.1 Contrast stretching

Contrast stretching is a process that expands the range of intensity levels in an image so that it spans the full intensity range of the recording medium or display device [5]. For an 8-bit greyscale image $f(x,y)$, a linear contrast stretching operation can be represented by:

$$g(x,y) = \frac{255}{(f_{max} - f_{min})}(f(x,y) - f_{min}), \tag{1}$$

where f_{max} and f_{min} are the maximum and minimum intensity levels of the image $f(x,y)$ respectively and $g(x,y)$ is the contrast stretched image.

2.2 Homomorphic filtering

An image is characterized by the amount of light incident on the scene which is referred to as the illumination $i(x,y)$ and the amount of light reflected by the objects in the scene referred to as the reflectance $r(x,y)$. The two components combine as a product to form the image $f(x,y)$ as follows:

$$f(x,y) = i(x,y)r(x,y), \tag{2}$$

where $0 < i(x,y) < K$ and $0 < r(x,y) < 1$.

K is an upper bound for the illumination whose practical value is approximately equal to 9×10^4 Lumens/m^2 when the sun is used as the source [5].

The illumination component is characterized by low frequencies while the reflectance is characterized by higher frequencies. The homomorphic filtering algorithm reduces the contribution of the illumination component by means of a modified Gaussian high pass filter. In order to separate the two image components the logarithm of the image is taken as follows [5]:

$$\log f(x,y) = z(x,y) = \log i(x,y) + \log r(x,y). \tag{3}$$

Taking the two dimensional DFT yields:

$$Z(u,v) = DFT\{\log i(x,y)\} + DFT\{\log r(x,y)\}. \tag{4}$$

The transformed image $Z(u,v)$ is filtered using a Gaussian high pass filter having a transfer function $H(u,v)$ to give:

$$S(u,v) = H(u,v)F_i(u,v) + H(u,v)F_r(u,v). \tag{5}$$

The filtered image in spatial domain is obtained by the inverse Discrete Fourier transformation as:

$$s(x,y) = IDFT[H(u,v)F_i(u,v) + H(u,v)F_r(u,v)]. \tag{6}$$

Taking the anti-logarithm of $s(x,y)$ yields the enhanced image $g(x,y)$ as:

$$g(x,y) = \exp[s(x,y)]. \tag{7}$$

2.3 Single-scale retinex algorithm

The Single Scale Retinex (SSR) algorithm is an image enhancement method that improves the brightness perception, contrast and sharpness of a greyscale image. The algorithm achieves these improvements through a combination of spatial and spectral transformations that result in dynamic range compression and a reduction in global variance of the image [2, 6, 8, 13]. The algorithm belongs to the class of centre-surround functions. Each output value of these functions is determined by the corresponding input value called the centre and its neighbourhood or surround [8]. The mathematical form of single scale retinex is given by:

$$g(x,y) = \alpha(\log[f(x,y)] - \log[s(x,y) \otimes f(x,y)]) - \beta, \tag{8}$$

where: $g(x,y)$ is the output image, $f(x,y)$ is the input image, α is a scaling factor, β is an offset parameter, $\otimes$ is the convolution symbol and $s(x,y)$ is a Gaussian filter kernel defined as:

$$s(x,y) = a\exp[-(x^2 + y^2)/\sigma^2], \tag{9}$$

where σ is the standard deviation of the filter which typically ranges from 10 to 250 and a is a constant given by:

$$a = \frac{1}{\sum_x \sum_y s(x,y)}. \tag{10}$$

Due to the large sizes of filter kernels used, the convolution in spatial domain is normally performed as a multiplication in frequency domain to reduce the processing time. The result is then transformed back to spatial domain using Inverse Discrete Fourier Transformation (IDFT) as follows:

$$f(x,y) \otimes s(x,y) = IDFT\{F(u,v)S(u,v)\}, \tag{11}$$

where $S(u,v)$ and $F(u,v)$ are the discrete Fourier transforms of $s(x,y)$ and $f(x,y)$ respectively.

The single scale retinex algorithm reduces the global variance of an image as shown in the histograms in figure 1 thereby increasing further the separability measure proposed by Otsu (see equation 22).

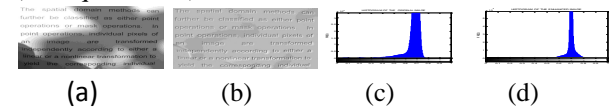(a)   (b)   (c)   (d)

Fig. 1: Dynamic range compression of SSR. (a) Input image (b) SSR enhanced image; (c) histogram of (a); (d) histogram of (b).

2.4 Otsu's thresholding method

Image thresholding is a statistical-decision theory problem whose objective is to minimize the average error in assigning pixels to two or more classes [5, 11]. Otsu's method is an optimal global thresholding technique. Let $\{0,1,2,...,L-1\}$

denote the L intensity levels of an image of size M x N pixels and let n_i denote the number of pixels having intensity level i. The total number of pixels in the image is given by:

$$MN = n_0 + n_1 + n_2 + ... + n_{L-1}. \tag{12}$$

The probability of occurrence of intensity level i is:

$$p_i = n_i \big/ MN, \tag{13}$$

where $\sum_{i=0}^{L-1} p_i = 1$ and $p_i \geq 0$.

Let a threshold level $T(k) = k$, $0 \leq k \leq L-1$ be used to dichotomize the image into two classes C_1 and C_2 where C_1 consists of all the pixels with intensity values in the range $[0, k]$ and C_2 consists of all the pixels with intensity values in the range $[k+1, L-1]$. The probabilities that a pixel is assigned to class C_1 and C_2 respectively are given by:

$$p_1(k) = \sum_{i=0}^{k} p_i = p, \tag{14}$$

and

$$p_2(k) = \sum_{i=k+1}^{L-1} p_i = 1 - p_1(k) = 1 - p_1. \tag{15}$$

The mean intensity values of the pixels assigned to classes C_1 and C_2 respectively are:

$$m_1 = \sum_{i=0}^{k} ip(i/C_1) = \frac{1}{p_1(k)} \sum_{i=0}^{k} ip_i, \tag{16}$$

and

$$m_2 = \sum_{i=k+1}^{L-1} ip(i/C_1) = \frac{1}{p_2(k)} \sum_{i=k+1}^{L-1} ip_i. \tag{17}$$

The cumulative mean up to intensity level k is given by:

$$m(k) = m = \sum_{i=0}^{k} ip_i. \tag{18}$$

The global mean of the image is given by:

$$m_G = \sum_{i=0}^{L-1} ip_i. \tag{19}$$

The global and between-class variances σ_G^2, and $\sigma_B^2(k)$ respectively are given by:

$$\sigma_G^2 = \sum_{i=0}^{L-1} (i - m_G)^2 p_i, \tag{20}$$

and

$$\sigma_B^2 = \frac{[m_G p_1(k) - m]}{p_1(k)[1 - p_1(k)]}. \tag{21}$$

Class separability metric, η is defined as follows

$$\eta = \frac{\sigma_B^2(k)}{\sigma_G^2} = \frac{[m_G p_1(k) - m(k)]^2}{p_1(k)[1 - p_1(k)]}. \tag{22}$$

The optimum threshold is the value, k_{opt} that maximizes the between-class variance obtained from

$$\sigma_B^2(k_{opt}) = \max \sigma_B^2(k), \tag{23}$$

for $0 \leq k_{opt} \leq 1$.

2.5 Local thresholding methods

In local thresholding methodologies, local information obtained from the neighbourhood of each pixel at the coordinates (x,y) is used to determine the threshold level T_{xy} for that pixel [5].

The thresholded image g(x,y) is obtained as:

$$g(x, y) = \begin{cases} 1 & \text{if} \quad f(x, y) > T_{xy} \\ 0 & \text{if} \quad f(x, y) \leq T_{xy} \end{cases} \tag{24}$$

where $f(x, y)$ is the input image.

In general, local binarisation methods perform better than global binarisation methods on degraded document images but are computationally slow, sensitive to the selection of parameter values [3]. The method proposed by Niblack [10] determines local threshold values using a rectangular neighbourhood. The threshold value for the pixel at the coordinates (x,y) is given by,

$$T_{xy} = k\sigma_{xy} + \mu_{xy}, \tag{25}$$

where k is a constant that determines the proportion of the total objet boundary considered to be a part of the object. K is usually set at -0.2.

The result of binarisation using this method retains background noise if the objects in the image are sparse [4].

Sauvola and Pietikainen [12] propose a methodology that reduces the background noise in the binarised image. The method adds a hypothesis on the intensity levels of the object and background pixels resulting in the following formula.

$$T_{xy} = \mu_{xy} \left[1 + k\left(\frac{\sigma_{xy}}{R} - 1 \right) \right] \tag{26}$$

where R is the dynamic range of the standard deviation fixed at 128 and k is a positive constant which is usually set at 0.5.

Evaluations of local binarisation methods have reported that Sauvola's method is better than the other types of local binarisation methods [3].

Gatos et al. [4] propose a technique for improving the Sauvola's method. The procedure starts with preprocessing using a Wiener filter. The foreground and background regions are then estimated. Next, adaptive thresholding is performed followed by up sampling. Finally, postprocessing using shrink and swell filters is performed.

2.6 Morphological dilation

Morphological dilation is an operation that grows objects in an image uniformly in spatial extent [5]. Dilation of an image A by a structuring element B is the set of all displacements z such that A and B overlap by at least one element which is written as:

$$A \oplus B = \left\{ z / [(B^{\bullet})_z \cap A] \subseteq A \right\}, \qquad (27)$$

where $B^{\bullet}$ and $(B)_z$ are the reflection about the origin and the translation by z of B respectively which are defined as follows:

$$B^{\bullet} = \{ w / w = -b, b \in B \}, \qquad (28)$$
$$(B)_z = \{ c / c = b + z, b \in B \}. \qquad (29)$$

The structuring element is a small set or subimage used to probe an image for properties of interest. The structuring elements used in morphological image processing are rectangular arrays which have defined origins. If the origin is not defined, it is assumed to be at the centre of symmetry of the structuring element. Dilation thickens the characters and also bridges gaps in the characters as reported in [6, 8, 15] and shown in figure 2.
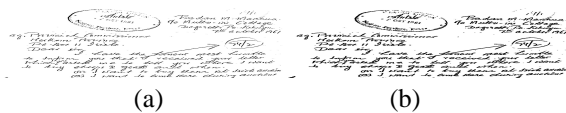

(a)   (b)

Fig.2 Bridging of gaps using morphological dilation: (a) Thresholded image; (b) dilation of (a).

3 Proposed method

An image of the document is captured using a camera and uploaded to a computer. Processing of the image is done using MATLAB version 7. The processing starts with conversion of the image from colour to greyscale image followed by linear contrast stretching. The image is then enhanced using the single scale retinex algorithm so as to remove shadows and compensate for poor illumination and noise. The output of the single scale retinex algorithm is then binarised using the Otsu's method which is based on maximization of the between-class variance. Any gaps in the text characters are then bridged using morphological dilation. The document image is finally stored in a memory device or printed. The algorithm is shown in form of a block diagram in fig.3.
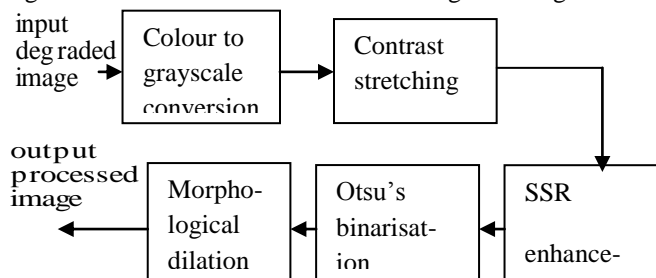

Fig.3: Block diagram of the proposed image enhancement procedure.

4 Experimental results

A series of experiments were conducted using MATLAB simulation. In all the experiments, the input is an 8-bit colour image having a spatial resolution of 960x1280 pixels and captured using Nokia 1680 cell phone. The Otsu's binarization function (greythresh) in MATLAB version 7 was used. Linear contrast stretching was used to expand the intensity range of the input image so that it spans over the entire range from 0 to 255. A 4x4 pixels square structuring element was used in morphological dilation. The parameters used in the single scale retinex algorithm are: $\alpha = 0.68, \beta = -0.72$ and $\sigma = 15$. The values of the parameters were selected based on the perceived quality of the output image after performing a series of experiments with different values.

4.1 Non uniformly illuminated images.

The results in figures 4 to 6 show the performance of the proposed algorithm on document images that are degraded by shadows in comparison with the Otsu's method. In each of the figures, the first image is the degraded input, the second shows the result of enhancing the input image using the proposed algorithm while the third shows the result of using Otsu's method on the same input. The results of the proposed algorithm show a significant improvement over the results of the Otsu's method in terms of shadow removal capability.

Figures 7 show the performance of the proposed algorithm on document images that are degraded by shadows in comparison with the Sauvola's method. The result of the proposed algorithm is subjectively comparable with that of the Sauvola's method.
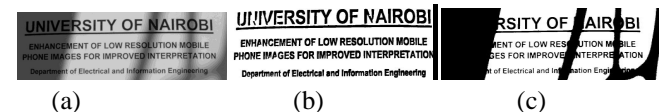

(a)   (b)   (c)

Fig. 4: Enhancement of an image with shadows (a) Input image; (b) Proposed algorithm result; (c) Otsu's method result.

(b)   (b)   (c)
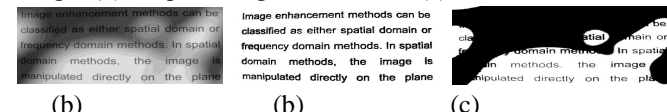
Fig. 5: Enhancement of an image with shadows (a) Input image; (b) Proposed algorithm result; (c) Otsu's method result.

(a)   (b)   (c)
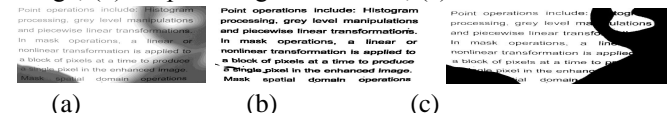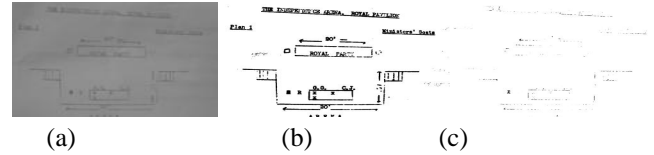
Fig. 6: Enhancement of an image with shadows (a) Input image. (b) Proposed algorithm result. (c) Otsu's method result.

(a)   (b)   (c)

Fig. 7: Enhancement of an image with shadows (a) Input image. (b) Proposed algorithm result. (c) Sauvola's method result.

4.2 Poorly illuminated images.

The results in figure 8 show the performance of the proposed algorithm on a document image that is degraded by low level of illumination in comparison with that of the Otsu's method. The first image is the degraded input, the second image shows the result of enhancement of the input using the proposed algorithm and the third image shows the result of using Otsu's method on the same input. The results of the proposed algorithm show a slight improvement over the results of the Otsu's method in terms of processed document legibility.
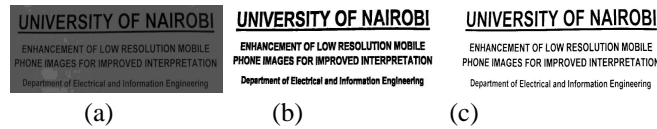

(a)   (b)   (c)

Fig. 8: Enhancement of poorly illuminated image (a) Input image. (b) Proposed method result. (c) Otsu's method result.

4.3 Faded document images.

The results in figures 9 to 12 show the performance of the proposed algorithm on document images that are degraded by fading in comparison with the Otsu's method. In each of the figures, the first image is the degraded input, the second image shows the result of enhancement of the input image using the proposed algorithm while the third image shows the result of using Otsu's method on the same input. The results of the proposed algorithm show a significant improvement over the results of the Otsu's method in terms of legibility of the processed image.
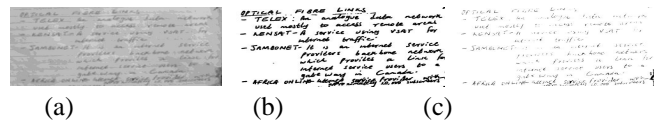

(a)   (b)   (c)

Fig 9: Enhancement of faded document image (a) Input image; (b) Proposed algorithm result; (c) Otsu's method result.

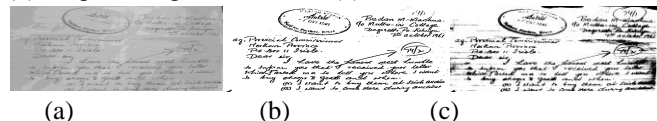(a)   (b)   (c)

Fig. 10: Enhancement of a historical handwritten document image (a) Input image. (b) Proposed algorithm result. (c) Otsu's method result.
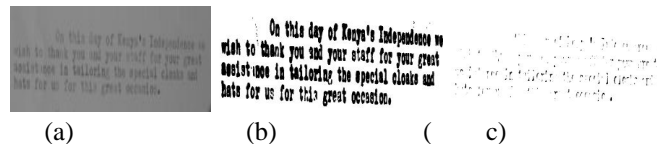

(a)   (b)   (c)

Fig. 11: Enhancement of a historical typed document images (a) Input image. (b) Proposed algorithm result. (c) Otsu's method result.

(a)   (b)   (c)

Fig. 12: Enhancement of a document image with diagrams (a) Input image (b) Proposed algorithm result. (c) Otsu's method result.

4.4 Effects of the single scale retinex process.

The results presented in this section demonstrate the effects of the single scale retinex process on images degraded by shadows as shown in figures 13 to 16 and table 1. In all the figures, the first image is the degraded input image while the second one is the single scale retinex enhancement result. Table 1 summarises the values of the global variances of the images in figures 13 to 16. The results show that single scale retinex enhancement improves the contrast and background uniformity of such images and also reduces the global variances of the images. These effects make the process of determining the threshold level more accurate.

(b)   (b)

Fig. 13: Shadow removal using SSR. (a) Input image   (b) SSR enhanced image.

(b)   (b)

Fig. 14: Shadow removal using SSR. (a) Input image   (b) SSR enhanced image.

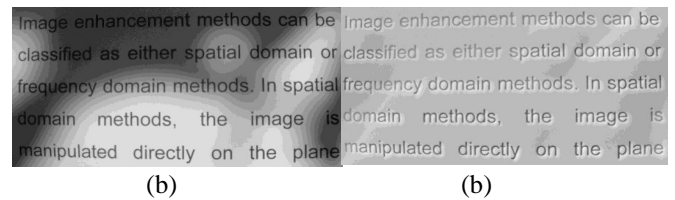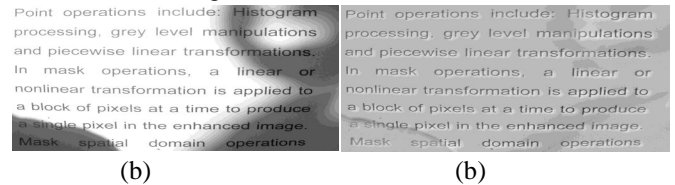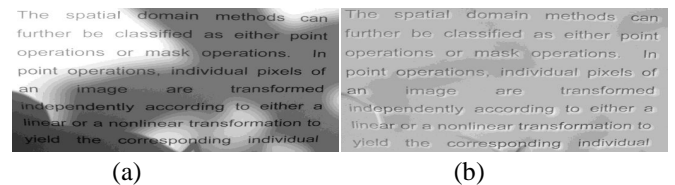(a)   (b)

Fig. 15: Shadow removal using SSR. (a) Input image   (b) SSR enhanced image.

(b)   (b)

Fig. 16: Shadow removal using SSR. (a) Input image (b) SSR enhanced image.

Image	Global variance	
	Input image	SSR enhanced image
Fig. 13	0.0490	0.0073
Fig. 14	0.0730	0.0055
Fig. 15	0.0654	0.0063
Fig. 16	0. 0464	0.0082

Table 1: Reduction of global variance by SSR enhancement.

4.5 Optical character recognition results

Table 2 shows the results of applying the ABBYY finereader optical character recognition software [1] on the results of the proposed method and those of the Otsu's method for images with shadows. From these results, a higher proportion of characters were recognised when the image is enhanced using the proposed method than when preprocessed using Otsu's method.

The results in the first row of the table show that, processing the input image using the proposed algorithm led to correct recognition of 132 out of the 133 characters while using Otsu's method, 78 characters were correctly recognised. For the second image, the proposed enhancement method resulted in correct recognition of all 147 characters while Otsu' method resulted in correct recognition of 57 characters.

Table 3 shows the results of applying the finereader optical character recognition software [1] on the results of the proposed method as well as those of the Sauvola's method on images with shadows.

In the first row, processing the input image using the proposed algorithm led to correct recognition of 232 out of the 262 characters while using Sauvola's method, 223 characters are correctly recognized. Results shown in the second row indicate that, the proposed algorithm resulted in correct recognition of all the 147 characters while Sauvola's method resulted in correct recognition of 145 characters.

Input image	Proposed method	Otsu's method
	UNIVERSITY OF NAIROBI ENHANCEMENT OF LOW RESOLUTION MOBIL PHONE IMAGES FOR IMPROVED INTERPRETATION Department of Electrical and Information	SITY OF AIR Bl ENT OF LOW RE$ TION M ILE ES FOR IMPROV A TERPR TION tof Electrical and In { ation Engi » =

		Engineering	
		Image enhancement methods can be classified as either spatial domain or frequency domain methods. In spatial domain methods, the image is manipulated directly on the plane	é ¢| M éin ori min - ; _ln spati A methods. the image ipulated directly on the pl

Table 2: Optical character recognition results.

Input image	Propose method	Sauvola's method
	Point operations include: Histogram processing, grey level manipulations and piecewise linear transformations. In mask operations, a linear or nonlinear transformation is applied to \ a block of pixels at a time to produce afilngleplxel in the enhanced image. Mask spatial d_omaln opemtivfls	Point operations include: H•stogrant processing, grey level manipulations and piecewise linear transformations. In mask operations, a linear of nonlinear transformation is applied. to: a block of pixels at a time to proglum a sin.gle_pixel in the enhancml tinPOR, Mask sp.atial domain ..opwatiOis:
	Image enhancement methods can be classified as either spatial domain or frequency domain methods. In spatial domain methods, the image is manipulated directly on the plane	Image enhancenrient methods can be classified as• either spatial domain or frequency domain methods. In spatial ,domain methods, the image is manipulated directly on the plane

Table 3: Optical character recognition results.

5 Conclusion

In this paper, a document image enhancement procedure has been proposed which addresses the effects of poor illumination, noise and fading simultaneously. We have also illustrated the effectiveness of the algorithm for each of these

degradations with some test results. The procedure improves the results of optical character recognition due to its improved binarisation results.

From the results, it is evident that the single scale retinex algorithm leads to better threshold level selection for images degraded by shadows than when the Otsu's method is applied on the original image. The improved performance is achieved mainly due the global variance reduction by the single scale retinex enhancement. The results of the proposed algorithm also compare well with those of the Sauvola's binarisation method in terms of optical character recognition. Future work shall focus on improving the bridging of the character gaps, more accurate removal of the shadow effects and text warping correction.

References

[1]   ABBYY, (www.finereader.com , (February 2012).

[2]   F. Bin, C. Jianwen, and X. Xinyan, "A novel adaptive vein image contrast enhancement method based on fuzzy and Retinex theory", *Proceedings of the IEEE International Conference on Information and Automation*, Harbin, China, pp 2447-2450, (2010).

[3]   S. S., Bukhari, F. Shafait, T. M. Breuel "Adaptive Binarization of Unconstrained Hand-Held Camera-Captured Document Images", Journal of Universal Computer Science, vol. 15, no. 18 (2009).

[4]   B. Gatos, I. Pratikakis, S.J. Perantonis, "Adaptive degraded document image binarization", *Pattern Recognition,* 39(3): 317-327, (2006).

[5]   C. Gonzalez and R.E. Woods, *Digital Image Processing*. 3rd edition, New Jersey, USA: Addison-Wesley, (2008).

[6]   X. Guoxin, X. Pei, and L. Qiang, "A method to improve the Retinex image enhancement algorithm based on wavelet theory", *International Symposium on Computational Intelligence and Design,* Changzhou University, China, pp 182-185, (2010).

[7]   M. Q. Hernandez, "Genetic Programming applied to Morphological Image Processing," PhD thesis*, The University of Birmingham,* UK, (2004).

[8]   D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex", *IEEE Trans. Image Process.*, **vol. 6, no. 3,** pp. 451–462, (1997).

[9]   I. K. Kim, D. W. Jung, R.H. Park, "Document image binarization based on topographic analysis using a water flow model", *Pattern Recognition* 35, pp. 265–277, (2002).

[10]   W. Niblack, *An Introduction to Digital Image Processing,* Prentice-Hall, Englewood Cliffs, NJ, 1986 pp. 115–116.

[11]   N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, **vol. SMC-9**, pp. 62–66, (1979).

[12]   J. Sauvola, M. Pietikainen, "Adaptive document image binarization", *Pattern Recognition*, **vol.33** pp. 225–236, (2000).

[13]   S. Wang and Y. Wang, "Shadow detection and compensation in high resolution Satellite image based on Retinex", Fifth *International Conference on Image and Graphics,* Wuhan University, China, pp 209-212, (2009).

[14]   H. Yang, A. C. Kot, and X. Jiang, "Binarization of Low-Quality Barcode Images Captured by Mobile Phones Using Local Window of Adaptive Location and Size", IEEE *transactions on image processing,* **vol. 21, no. 1,** pp 418-425, (2012).

[15]   A. Zadorozny and H. Zhang, "Contrast enhancement using morphological space scale", *Proceedings of the IEEE International Conference on Automation and Logistics,* Shenyang, China, (200

A2. MATLAB codes

```matlab
function p= enhan(A,alpha, beta, sigma, M, N)
% ENHAN converts a colour image to a grey scale image and then enhances it to
% improve on interpretability.
% P=enhan [A, ALPHA, BETA, SIGMA, M, N] is a function for enhancing  a degraded
% input image using bicubic interpolation, contrast stretching, Otsu's global thresholding
% and morphological dilation.
% The function reduces degradations caused by factors such as noise,  shadows, fading,
% low illumination levels and geometrical distortions
% A is the input degraded image,
% Alpha is the scaling factor used in the single scale retinex   enhancement.
% Beta is an offset parameter used in the single scale retinex  enhancement.
% Sigma is the standard deviation of the SSR Gaussian filtering kernel.
% M is the number of rows in the input degraded image.
% N is the number of columns in the input degraded image.
% Function was written by Kiragu Henry Macharia an MSc student in the department of
% Electrical and Electronic Engineering of the University of Nairobi on 15th October
% 2012.
% Function ENHAN was written for enhancing degraded images for an   MSc research
% thesis which was supervised by professor Elijah Mwangi of the department of Electrical
% and Electronic Engineering of    the University of Nairobi.

% Determination of the input arguments alpha, beta, sigma, M and N plus linear contrast
% stretching enhancement.
I2=rgb2gray(i);       % Coversion of the input image from colour to greyscale.
f1=im2double(I2);  % Conversion of uint_8 image data to double data.
[M, N]=size(f1);       % Determination of the input arguments M and N.
a=max(max(f1));      % Determination of the maximum image intensity.
b=min(min(f1));       % Determination of the minimum image intensity.
for m=1:M,
for n=1:N,
f2(m,n)=(f1(m,n)-b)/(a-b); % linear contrast stretching.
end
end
if nargin<=5            % Settings of the other input arguments not specified.
alpha=1.44;
beta=0.7;
sigma=10;
else
error ('too many input parameters')
end
if M<=240,
f=imresize(f2,4, 'bicubic'); % Bicubic interpolation.
else
f=f2;
% Reduction of degradations in the input image.
```

150

```matlab
for m=1:M,   % rows of the Gaussian lowpass filter kernel used in SSR  enhancement.
for n=1:N,    % Columns of a Gaussian lowpass filter kernel used in SSR  enhancement.
z (m, n)=exp(-(m^2+n^2)/sigma^2); % Calculation of the Filter coefficients.
end
end
K=sum (sum (z));      % Determination of the filter kernel scaling factor.
h1= z/K;               % Computation of the scaled filter kernel matrix.
H=fft2 (h1, M, N);        % Determination of the 2-D DFT of the filter kernel.
F=fft2 (f, M, N);          % 2-D DFT of the degraded greyscale image.
G=F.*H;                 % DFT of Gaussian filtered input image.
g1=abs (ifft2 (G));        % Determination the low pass image matrix.
I3=alpha*log (g1+1);  % Determines the log of the low pass image matrix.
I5=alpha*log (f+1);      % Determines the log of the input image matrix.
I6=(I5-I3)+beta;          % Computation of the degradation-reduced image.
% Otsu's global thresholding of the degradation-reduced image.
level = graythresh(I6);   % Determination of the optimum threshold level.
BW = im2bw(I6,level);   % Computation of the thresholded image.

% morphological dilation of the thresholded to bridge character gaps
SE=strel('square',5);  % Specification of the morphological structuring  element (SE).
D=(1-BW);               % Inverse transformatiom of the thresholded image.
Q=imdilate(D, SE);      % Dilation of the inverted image.
p=1-Q;                  % Computation of the output enhanced image.
imshow(f);              % Display of the degraded greyscale image.
figure, imshow(p);      % Display of the output enhanced image.
end
function q= debra(T, a, b, M, N)
% DEBRA converts a colour image to a greyscale image and then de-blurs.
% q= debra[T, a, b, M, N] computes the de-blurred image based on a linear relative
% motion model.
% 'a' and 'b' are constants whose values are are determined from the velocity
%  components in both x and y directions.
% T is the camera exposure time.
% M is the number of rows in the input image
% N is the number of columns in the input image

% process of determining the input argumentsalpha, beta, sigma,M and N.
I2=rgb2gray(i);      % coversion of the input image from colour to greyscale.
f1=im2double(I2);   % conversion of uint_8 image data to double data.
[M, N]=size(f1);      % determination of the input arguments M and N.
if nargin<=5          % setting of the input arguments that are not specified.
a=8*10^-4;
b=8*10^-4;
T=1;
else
error('too many input parameters')
```

```matlab
end
for m=1:M,
  for n=1:N,
      E(m,n)=T*(sin(pi*(a*m+b*n)*exp(-j*pi*(a*m+b*n))/(pi*(a*m+b*n));

  end

end

X=fft2(g);% Calculates the 2-D DFT of the logarithm of the image

Y= X./E;% Performs matrix element wise multiplication

y=ifft2(Y);% Calculates the 2-D inverse DFT

q=real(y); % Removes imaginary components in y

imshow(q); % Displys the final de-blurred image.
end
function r= retin(i, alpha, beta, sigma, M, N)
% ENHAN converts a colour image to a grey scale image and then enhances it to
% improve its quality.
% r=retin[i, alpha, beta,sigma M, N] computes the enhanced version of the input image
% using the SSR algorithm.
% i is the input degraded image,
% alpha is the scaling factor used in the SSR enhancement.
% beta is an offset parameter used in the SSR enhancement.
% sigma is the standard deviation of the SSR Gaussian filtering kernel.
% M is the number of rows in the input image
% N is the number of columns in the input image

% process of determining the input argumentsalpha, beta, sigma,M and N.
I2=rgb2gray(i);            % coversion of the input image from colour to greyscale.
f1=im2double(I2);         % conversion of uint_8 image data to double data.
[M, N]=size(f1);          % determination of the input arguments M and N.
if nargin<=5              % setting of the other input arguments if they are not specified.
alpha=1.44;
beta=0.7;
sigma=10;
else
error('too many input parameters')
end
 % SSR enhancement of the input image.
for m=1:M,           % setting of the number of rows of the Gaussian lowpass
                     % filter kernel used in SSR enhancement.
for n=1:N,           % setting of the number of Columns of a Gaussian lowpass
                     % filter kernel used in SSR enhancement.
z(m,n)=exp(-(m^2+n^2)/sigma^2); % Calculation of the filter coefficents matrix.
```

```matlab
    end
end
K=sum(sum(z));          % Determination of the filter kernel scaling factor.
h1= z/K;                % computation of the scaled filter kernel matrix.
H=fft2(h1,M,N);         % Determination of the 2-D DFT of the filter kernel
F=fft2(f,M,N);          % Determination of the 2-D DFT of the degraded image
G=F.*H;                 % Calculation of the 2-D DFT of the lowpass filtered input image.
g1=abs(ifft2(G));       % Determination  the low pass image matrix.
I3=alpha*log(g1+1);     % Determines the logarithm of the low pass image matrix
I5=alpha*log(f+1);      % Determines the logarithm of the input image matrix
r=(I5-I3)+beta*ones(M,N);  %Computation of the degradation-reduced image
imshow(f);              % Display of the degraded greyscale image
figure, imshow(r);      % Display of the SSR enhanced image.
end
function s= alrot(alpha, M, N)
% alrot converts a colour image to a grey scale image and then improves its contrast.
% s= alrot[alpha, M, N] computes the contrast enhanced  version of the input image
% alpha is a constant whose value is less than unity.
% M is the number of rows in the input image
% N is the number of columns in the input image

% process of determining the input arguments.
I2=rgb2gray(i);         % coversion of the input image from colour to greyscale.
f1=im2double(I2);       % conversion of uint_8 image data to double data.
[M, N]=size(f1);        % determination of the input arguments M and N.
if nargin<=3            % setting of the input arguments that are not specified.
alpha=0.02;
else
error('too many input parameters')
end
fd=im2double(I2);       % converts uint_8 image data to double

H=fft2(fd);             %Calculates the two dimensional DFT of the image

for m =1:1:M,           % move in steps of one from the first row to the last

   for n=1:1:N,         %  move in steps of one from the first column to the last

G(m,n)=abs(H(m,n))^alpha; % Modifies the magnitude of the 2-D DFT of the image.

end

end

K=G.*H;        % Modifies the magnitude of the 2-D DFT.

s=ifft2(K);    % Transforms the modified 2-D DFT into spatial domain.
```

```matlab
imshow(s);     % Displays the alpha root enhanced image.
end
```

function t= homom(i, gammal, gammah, c, d0, M, N)

```matlab
% HOMOM converts a colour image to a grey scale image and then enhances it using
% homomorphic filtering.
% t= homom(i,gammal,gammah,c,d0, M, N)computes the homomorphic filtered version of
% the input image.
% i is the input degraded image.
% gammal is a constant that is less than unity.
% gammah is a constant that is greater than unity.
% c is a constant less than unity that controls the filter sharpness.
% d0 is the cutoff frequency of the filter.

% process of determining the input arguments.
I2=rgb2gray(i);         % coversion of the input image from colour to greyscale.
f1=im2double(I2);       % conversion of uint_8 image data to double data.
[M, N]=size(f1);        % determination of the input arguments M and N.
if nargin<=7             % setting of the other input arguments if they are not specified.
gammal=0.6;
gammah=1.6;
c=0.8;
d0=100;
else
error('too many input parameters')
end

 %Homomorphic filtering of the input image.
for m =1:1:M,  % move in steps of one from the first row to the last
   for n=1:1:N, %  move in steps of one from the first column to the last
     D(m,n)=((m-M/2)^2+(n-N/2)^2)^0.5;   % calculates the 2D frequency matrix
H(m,n)=  (gammah-gammal)*exp(-c*D(m,n)^2/d0^2)-gammal; % Calculates the filter
%response using the 2-D DFT matrix.

     lf(m,n)=log(f1 (m,n)+0.1);%determines the image logarithm

   end

end

F=fft2(lf);         % Calculates the 2-D DFT of the logarithm of the image

G=F.*H;             % Performs matrix element wise multiplication

g=ifft2(G);         % Calculates the 2-D inverse DFT

t=exp(real(g));  % Determines the inverse logarithm and adds an offset
```

```
  imshow(t);       % Displays the final enhanced image

end




function p= sauv(i, k, R, M, N)
% SAUV binarises an image using the Sauvola's local thresholding method.
% p= sauv[i, R, k, M, N]computes the local thresholded version of the input image.
% i is the input degraded image,
% k is a positive constant less than one.
% R is the dynamic range of the standard deviation.
% M is the number of rows in the input image
% N is the number of columns in the input image
% process of determining the input arguments.
I2=rgb2gray(i);       % coversion of the input image from colour to greyscale.
f1=im2double(I2);    % conversion of uint_8 image data to double data.
[M, N]=size(f1);        % determination of the input arguments M and N.
if nargin<=5          % setting of the other input arguments if they are not specified.
K=0.5;
R=0.5;
else
error('too many input parameters')
end
h=(ones(3,3))/9;  % Mean filtering mask

A=imfilter(f,h);     % Local mean determination.

B=f.*f;                % Square of the subimage.

C=imfilter(B,h);    % Mean of the square of the subimage.

D=A.*A;              % Square of the mean of the subimage.

V=C-D;               % Variance of the subimage.

S=sqrt(V);           %  Standard deviation of the subimage.

T=k/R*A+0.4*S;     % Threshold level

F=f-T;

for m=1:1:M,   % Columns of the filter kernel

for n=1:1:N,     % Columns of the filter kernel

if F(m,n)<0, p(m,n)=0;

else
```

155

```
p(m,n)=1;

        end

    end

end

imtool(p);  %Displays the enhanced image.

end
```

```matlab
function r1= rtmsq(i1,i2,M, N)
% RTMSQ determines the RMSE of a degraded image in comparison with a reference
% ground-truth image.
% r1=rtmsq[i, alpha, beta, sigma m, n] computes the RMSE of an input image.
% i1 is the ground-truth image.
% i2 is the degraded image.
%M is the number of rows in the input image
%N is the number of columns in the input image

% process of determining the input arguments.
f1=im2double(i1);          % conversion of uint_8 image data to double data.
x=im2double(i2);           % conversion of uint_8 image data to double data.
[M, N]=size(f1);           % determination of the input arguments M and N.
a=0;

for m =1:1:M,              % move in steps of one from the first row to the last

    for n=1:1:N,           %  move in steps of one from the first column to the last

        a=a+f1(m,n);

    end

end

b=0;

for m =1:1:M,              % move in steps of one from the first row to the last

    for n=1:1:N,           %  move in steps of one from the first column to the last

        b=b+x(m,n);

    end

end

r=sqrt(abs((a-b)^2));
```

```matlab
q=0;

for m =1:1:M, % move in steps of one from the first row to the last

    for n=1:1:N,%  move in steps of one from the first column to the last

        q=q+f1(m,n)^2;

    end

end

s=sqrt(q);

r1=r/s;        % Root mean square error.

end


function sd= dev(I M, N)
% DEV calculates the standard deviation of an image..
% sd= dev(I M, N) computes the the standard deviation of a grayscale image.
% I is the input image.
% M is the number of rows in the input image
% N is the number of columns in the input image
I2=rgb2gray(I); % coverts image from RGB to grayscale.
f=im2double(I2); % converts uint_8 image data to double.
[M, N]=size(f1);    % determination of the input arguments M and N.
if nargin<3        % setting of the input arguments that are not specified.
error('too few input parameters')
end
 S=0;

for m =1:1:M, % move in steps of one from the first row to the last

 for n=1:1:N,%  move in steps of one from the first column to the last

S=S+f(m,n);

    end

end

k=M*N;

mn=S/k;

d=0;

for m =1:1:M, % move in steps of one from the first row to the last
```

```
 for n=1:1:N,%  move in steps of one from the first column to the last

d = d + ((f(m,n) - mn)^2).*f(m,n);

end

end

va= d/S;

sd=sqrt(va); % Standard deviation.

end
```

A3 Ground Truth Images

In this section of the appendix, ground-truth images of three historical documents which were used in obtaining some of the experimental results are presented. The documents were sourced from the Kenya National archives.
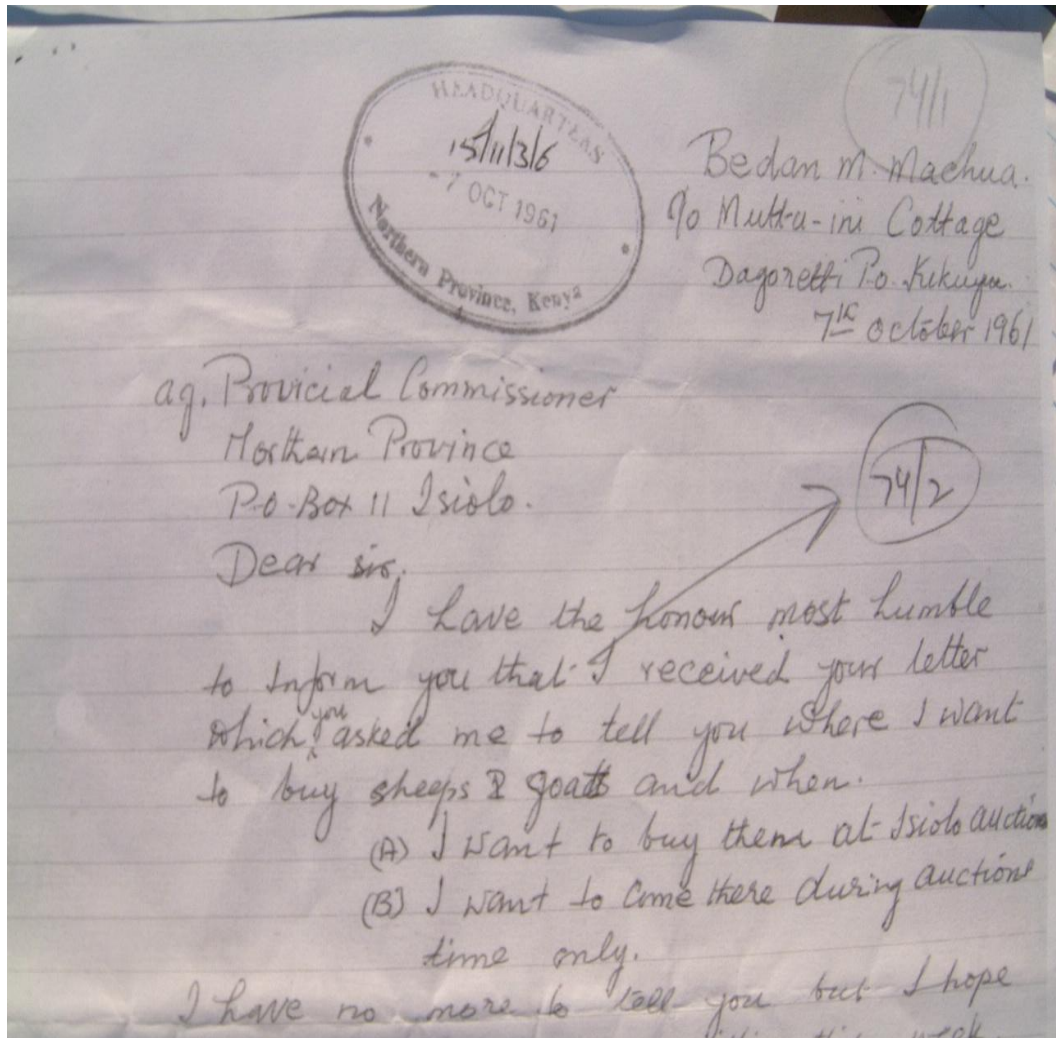

Figure A.1: Ground-Truth Image for the Input Image in Figures 7.7 and 7.23.
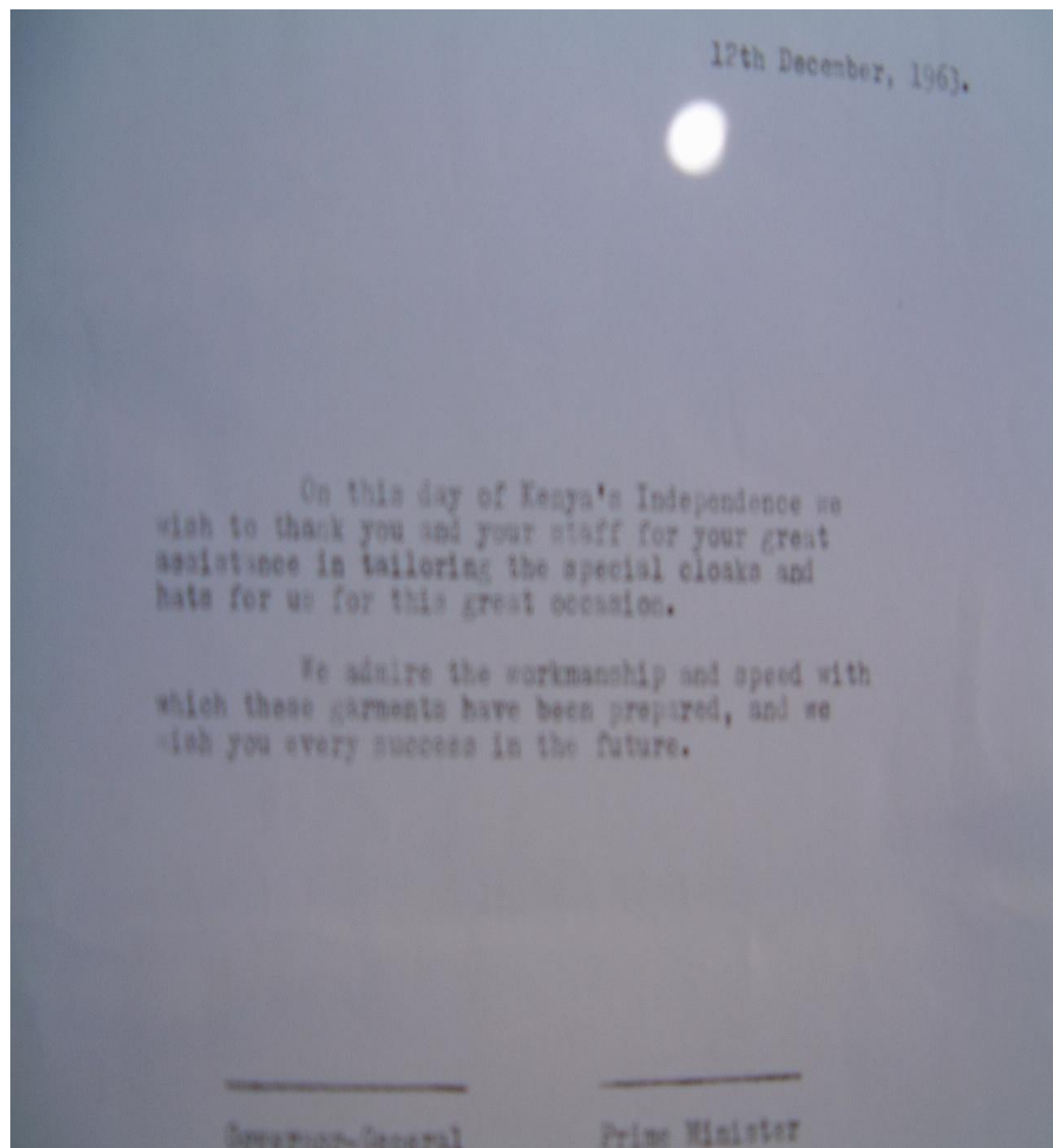
12th December, 1963.

On this day of Kenya's Independence we
wish to thank you and your staff for your great
assistance in tailoring the special cloaks and
hats for us for this great occasion.

We admire the workmanship and speed with
which these garments have been prepared, and we
wish you every success in the future.

Governor-General   Prime Minister

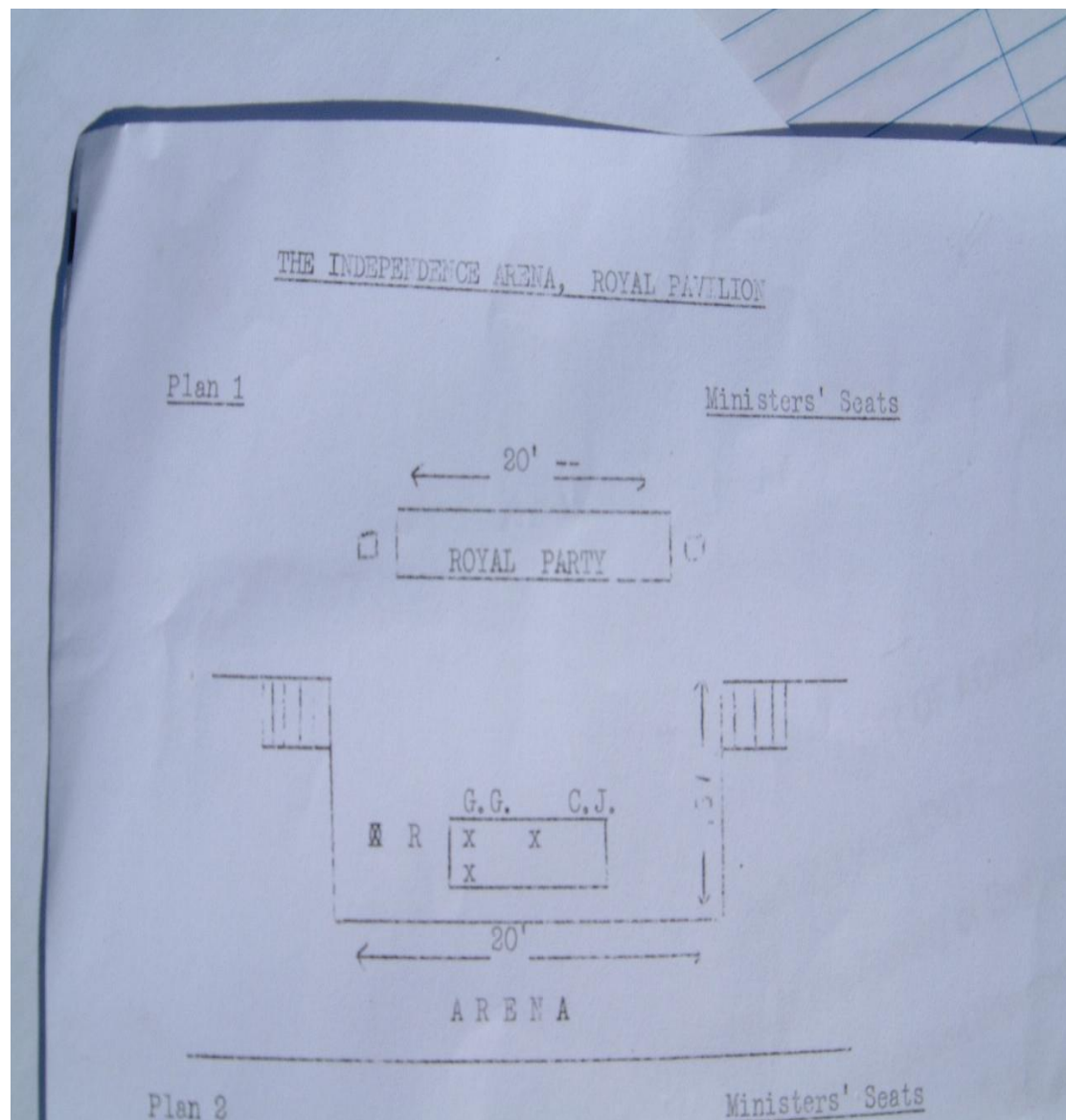Figure A.2: Ground-Truth Image for the Input Image in Figure 7.24.

Figure A.3: Ground-Truth Image for the Input Image in Figure 7.25.

A4 Image Blurring Model

Image blurring degradation is commonly caused by a relative motion between the scene objects and the sensor during image acquisition. The planer motion can be represented by time varying components $x_0(t)$ and $y_0(t)$ in the x and y directions respectively.

The degraded (blurred) image $g(x, y)$ can be represented as a continuous summation of many images for the duration during which the camera shutter is open. This summation can be modeled as integration as follows,

$$g(x, y) = \int_0^T f[x - x_0(t), y - y_0(t)]dt. \tag{A.1}$$

where T is the duration of exposure.

Taking the 2D-DFT of g(x, y) gives:

$$G(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)e^{-j2\pi(ux+vy)}dxdy \tag{A.2}$$

$$= \int_0^T \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f[x - x_0(t), y - y_0(t)]e^{-j2\pi(ux+vy)}dxdy \right]dt \tag{A.3}$$

$$= F(u, v)\int_0^T e^{-j2\pi(ux_0(t)+vy_0(t))}dt \tag{A.4}$$

The blurring function is therefore given by:

$$H(u, v) = \int_0^T e^{-j2\pi(ux_0(t)+vy_0(t))}dt \tag{A.5}$$

For linear motion, $x_0(t) = \dfrac{at}{T}, y_0(t) = \dfrac{bt}{T}$ where a and b are constants

The degradation function is therefore given by:

$$H(u, v) = \frac{T\sin[\pi(ua + vb)]e^{-j\pi(ua+vb)}}{\pi(ua + vb)} \tag{A.6}$$