



# Dewarping on Camera Document Images

<sup>1</sup>Sruthy S, <sup>2</sup>Dr. S. Suresh Babu

<sup>1</sup>Research Scholar, Computer science & Engineering, Noorul Islam Centre for Higher Education, Noorul Islam University, Kumaracoil, TamilNadu, India.

email: [ssruthyphdni@gmail.com](mailto:ssruthyphdni@gmail.com)

<sup>2</sup>Professor, Sree Buddha College of Engineering, Pattoor P.O, Nooranad, Alappuzha District Kerala, India

**Abstract**—Warping reduces the readability and accuracy of the camera document images. Hence it affects the OCR. Here Dewarping is proposed with two steps a coarsedewaring along with fine dewarping. To map curved document image to a 2D rectangular area transformation model is used. Then projection of the document is by fitting top and bottom curved lines and straight lines in sides. This process is coarse dewarping and then fine dewarping is performed for detection of words. Since words are normalized by coarse dewarping, it is found to be effective. Analysis of the results based on different images taken are done on multiple languages and proved to be efficient.

**Index Terms**—Coarse dewarping, Fine dewarping, Morphological Algorithms, Text line dewarping

## 1. INTRODUCTION

Digital imaging paved a new way for storing images electronically. It has numerous benefits like saves storage space, handles records easily, find documents rapidly and prevents lost records. The availability of high performance, low quality equipments have made prolific changes in digital imaging analysis. Document image processing has been extensively studied about past 40 years. In conventional times, document imaging has been done with huge flat bed scanning devices. The images from flat bed scanners give good start up

but when they are working on digital cameras they will be giving low clarity images, which cannot be readable by OCR systems. Recently, Portable devices like digital camcorders, digital cameras, PCcams, PDAs, and even cell phone cameras are most commonly used for image capture. They are small, light, portable, easily integrated with

various networks, and moreover they are more flexible for many document capturing tasks in

less constrained environments. These factors are heading to a natural extension of the document processing community where cameras are used for document image analysis.

Document imaging covers many different areas including pre-processing, graphics analysis, writer identification, digital libraries, office automation, and forensics etc. Digital cameras, camcorders, PDAs and phone cameras can supplement the scanners and makes document image processing more flexible. These technical developments leads to advanced researches in the field of document image processing which aims in the video files and also abstraction of images in the text files. The scanner based OCR applications are now being converted into new platforms which are using more flexible image capturing devices.

Document imaging analysis can be categorized into a number of ways: by the techniques used, by the devices equipped, by the intended application.

Extraction of document images can be of various types. Document image contains text files with scenes, video frames with captions, etc. The feature difference of the image causes many of the challenges in the extraction of the document. Camera captured images suffer problems like perspective distortion which causes warping along the spline of the book, geometric distortion, low resolution, uneven lighting, complex backgrounds. These types of distorted images cannot be readable by the current OCR systems. Non-linear warping is a major distortion that makes document imaging analysis disgusting. The strongly distorted text in the document image makes the processing more complicated.

Many recent approaches have addressed these problems and can be classified into two main categories based on three-dimensional (3D) document shape reconstruction and two dimensional (2D) document processing. Three-dimensional (3D) reconstruction requires specialised hardware like stereo cameras, laser scanners etc. So, it limits the flexibility of camera captured devices. (2D) document image processing uses single camera in an uncontrolled environment so these processing techniques are more ease to use [1-2].

### 1.1 OBJECTIVE

The main objective of this is to develop a dewarping algorithm for curled document images based on coarse to fine dewarping using enclosed box method. For the development of dewarping algorithm MATLAB was used. The page dewarping has set off lot of interest in the scientific fraternity over the last few decades. The goal of dewarping is to flatten curled document images and make it readable by the OCR systems. Document dewarping analysis can be classified into two dimensional document processing which requires only limited hardware for analysis and three dimensional document reconstruction which limits the flexibility of user. In this paper work, (2D)

document image processing method is used. This method provides coarse dewarping for whole image and encloses each character by a rectangular box to provide dewarping at word level. Fine dewarping is applied to improve the dewarping efficiency.

### 1.2 ORGANIZATION OF PAPER

The paper is organized as follows which includes seven sections. In section two, a review on the research works already carried out in dewarping is included. Section three would provide a background on document imaging analysis and analyse the challenges involved in image acquisition. Section four would outline the different dewarping approaches. Section five provides analysis on the performance of coarse to fine dewarping using enclosed box method.

## 2. LITERATURE REVIEW

This section represents a brief review of several studies and researches related to dewarping of document images. Several s were studied based on this paper and among them few s relevant to this work are reviewed here.

Masalovitch and Mestetskiy [3] proposed a method for approximation of whole image deformation as combination of single interlinear space deformations. Long continuous branches are used for defining interlinear spaces of the document. They are approximated by cubic Bezier curves so as to estimate the deformation of each interlinear space. After that a whole approximation of document image is built. Here, the initial image should be black and white with black text lines and white background and the initial image should contain one big text block. First the image is binarized and after that discrete binary image is represented as a set of continuous polygonal figures with lowest perimeter. Skeleton of polygonal figure can be represented as a planar graph, where nodes are points on a plane and bones are straight lines that connect the nodes. The main idea of the

algorithm is that the image is represented as continuous skeleton system, and then filtering of the skeleton is built such that the unwanted bones are removed. After that extracting long near horizontal branches and then each branch is approximated by cubic Bezier curves and Bezier patches are built based on the obtained curves. One of the steps of this algorithm is the pre-processing step, on which all small garbage branches and branches that can be obviously determined as non-interlinear from the skeleton are deleted. This method has a demerit that it will not give a satisfactory result in the case of vertical borders of image which isn't so accurate.

Zhang and Tan [4] divide the document image into shaded and non shaded region. Initially, the shaded region is identified and image is binarized using Niblack's method. They find the text line curves by the connected component analysis and move the components to restore straight horizontal baselines. Images must be always greyscale and have a shaded region. This restoration method uses connected component analysis and regression methods to dewarp the image. A top down scanning method is used to rectify the distortions in clean area and after that alignments are corrected using linear regression method. A bottom down approach is applied to shaded portion and polynomial regression method is used to rectify the distortions. After that, warped text alignments and linear text alignments in both areas are then paired up. The warped text lines are restored by correcting the quadratic curves accordingly based on the corresponding straight text lines. This approach can be applied only for gray scale images.

In Lu and Tan's method for the restoration of camera documents [5], the image is divided into three subsections. They are document partition, the target rectangle construction and document restoration. The document partition step includes two subdivisions. First one is that the distorted document image is divided into X line and base line. Second step converts the identified text line

into smaller patches. After that, a target rectangle correspondence is built for each image such that the distorted image is mapped to this rectangle. The target rectangle is constructed based on the number and the aspect ratios of enclosed characters. The character aspect ratios are determined based on character span, character ascender, descender, and character intersection numbers. For each partitioned image patch, a target rectangle correspondence must be constructed within the target image to rectify that partitioned image patch. This method classifies characters to six categories with six different aspect ratios. Characters are classified based on the features including character span, character ascender, descender, and character intersection number. Finally in the image restoration step, rectification homography is applied to dewarp the image. This approach cannot be used when the distortion angle is big.

Zhang and Tian [6] introduced a method for warped document restoration in digital libraries. This method particularly focussed on boundaries to reduce the warping effect and uses Gordon surface model for the text lines of 2d image. Natural cubic splines are used for representing text lines. The image clarity will be less using this method. Here, a document-boundary independent approach to correct arbitrarily warped document images taken using ordinary digital cameras is explained. It is based on the Gordon surface model constructed from a set of text lines extracted from the 2D image. The text lines are represented using Natural Cubic Splines interpolating a set of points extracted from connected component analysis. Most of the images do not have explicit boundary curves for boundary interpolation. However, a ruled surface model based on the text lines is constructed here. This Gordon surface model can be projected to a planar surface without distortion. This 3D Gordon surface model can be applied to the 2D projection image since straight lines are preserved under projection. The iso-parametric lines passing through must also pass through the corresponding 2D projection points. Therefore, the projection of

this Gordon surface model can also be parameterized using the projected text lines in the 2D image. This method cannot be used for more distorted image.

Koo, Kim and Cho [7] proposed an algorithm to compose a geometrically dewarped and visually enhanced image from two document images taken by a digital camera at different angles. From the corresponding points in these images, the surface of a book is reconstructed, and then stitches two rectified images for a visually better composite. Initially, a cost function is defined for the correction of geometric distortion, which is related with the geometric transformation of 3-D points. After that image stitching method is used to combine better patches from two images. Graph-Cut Optimization is used as the stitching method. Due to the misalignment of two rectified images and the asymmetry on the amount of information of each image, the simple average of two images is not a good solution to composing an enhanced image. So, better parts from each of the images are found, and then stitch and blend them into a single image. This method fails when distortion angle is big and not suitable for more curled images.

Tian and Narasimhan [8] in Rectification and 3D Reconstruction of Curved Document Images proposed a method that automatically reconstructs the 3D shape and rectifies a deformed text document from a single image. The regularity in the text pattern is used to constrain the 2D distortion grid. Here, the 2D distortion (warping) grid in an image is estimated by exploiting the line structure and stroke statistics in text documents. This estimation consists of two main steps: 1) text lines are automatically identified and densely traced 2) the text orientation is determined at every location in the image. In most documents 2D image grid can be regarded as a perspective projection of a 3D parallelogram mesh. Here, the process is done by tracing an initial set of text lines, called seed lines, across the document image from randomly selected initial points. These initial points are based on an image self-similarity measure.

Then these seed lines are re sampled and refined using dynamic programming. In this work, it is assumed that the camera projection is perspective and each cell of the 2D warping coordinate grid is a parallelogram in 3D space. The second assumption is reasonable because the surface can be assumed to be locally planar or rigid if grid cells are sufficiently small. For most undistorted planar documents, the text lines are parallel and so are local vertical text directions, thus forming a parallelogram grid. But, this method reduces the flexibility of the user because additional hardwares are required.

### 3. DOCUMENT IMAGES

Document imaging analysis can be categorized into a number of ways: by the techniques used, by the devices equipped and by the intended application. Extraction of document images can be of various types. Document image contains text files with scenes, video frames with captions, etc. The feature difference of the image causes many of the challenges in the extraction of the document. Three fourth of the work done in this area is on extraction of image and video text from broadcast video or still images. Different documents require different types of devices to convert them to digital format. The following section discusses the various types of imaging devices for this purpose.

#### 3.1 IMAGING DEVICES

Digital scanners are one of the most important documents imaging device used for past decades. Scanners vary from drum scanners to small desktop scanners. The speed of the scanners can be varied from several pages per second to one line per second. The resolution of consumer-grade flatbed scanners has recently passed 2400 dpi (dots per inch), and those for film scanning can be much higher, and at the same time the price of consumer-grade scanners has fallen well below \$100, making them very popular PC add-ons [2]. Fig 3.1 shows different kinds of industrial cameras.

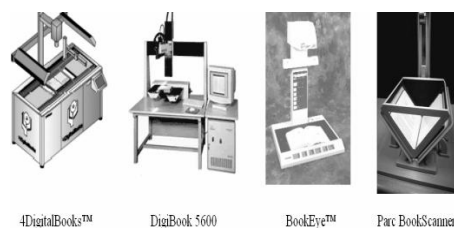


Fig 3.1: industrial cameras

In the case of analysis of documents like huge manuscripts, bounded volume books, brittle etc., scanners cannot be used. Cameras are most commonly used for document image analysis in such cases. These cameras are industrial grade, high quality and expensive as the systems use them. These cameras are called planetary cameras. To keep the data as flat as possible a particular environment is required and the environment should be well lit. The camera should be mounted on a high precision rack.

The advent of digital camera was a milestone in the entire document imaging analysis world. The most important advantage of them is their flexibility. They can be small as a business card and can be carried to anywhere. The border line between the imaging and video devices are disappearing by the invention of digital cameras and cam coders.

Current consumer-grade digital cameras are expanding to 8 megapixels and beyond, with resolutions of up to  $3500 \times 2200$ . In most of the ideal imaging conditions, this resolution is sufficient for capturing documents at a resolution (300 dpi) adequate for document image analysis. Digital video cameras that we use currently have much lower resolution ( $640 \times 480$ ) because they are designed primarily for low-bandwidth environment and are often highly compressed. The fact that they are not designed specifically for document image capture presents many interesting challenges. Ultimately, we hope to be able to perform various

document analysis tasks directly on the device. Now, companies are marketing compact flash cameras that can capture document images which are attached to tablets or PCs. Nokia and other telecom companies have recently released camera phones that capture at a resolution up to  $640 \times 480$  with over 1 megapixel [2].

### 3.2 ADVANTAGES OF CAMERA DOCUMENT ANALYSIS

Camera document analysis has many advantages over scanners. Cameras are small, they can be carried to any environment, and are more ease to use. In general, they are more flexible.

A study based on OCR systems conducted by Newman et al. [9] shows that desktop OCR using PC-cams is more flexible and productive than a scanner-based OCR for document image analysis. Fisher [10] analysed the possibility of substituting sheet-fed scanners used by soldiers in the battlefield, with digital cameras. He find out that sheet fed scanners cannot be used to capture images of thick bounded books, and are bulky and they limit the flexibility of users. All these conditions make them not suitable for battlefields. These results leads him to the conclusion that digital cameras are capable of capturing a whole A4 size document page at an equivalent 200 dpi resolution needed by OCR. Fig 3.2 shows the price ranges and resolution of different consumer-grade digital cameras.



Fig 3.2:price ranges and resolution of consumer-grade digital cameras

### 3.3 CHALLENGES

Major challenges in document image analysis are explained below:

**Perspective distortion:** This distortion occurs when the imaging plane is not parallel to the text plane. This cause the images appear too farther and cannot be easily readable to the OCR systems.

**Colour quantization and intensity:** In an ideal image acquisition device, each pixel in a photon sensor array should results the luminance of the inbound light and/or colour components corresponding to the frequency of the light. However, different hardware techniques have different spatial/intensity/colour quantization mechanisms. The first problem is the low-pass filter used in many digital cameras. Current CCD/CMOS- based camera sensors are in the Bayer format. This pattern has twice as many G sensors as R and B sensors. Each pixel can see only single colour. A low-pass filter is applied to spread the colour to nearest positions. Most scanners use separate CCD/CMOS sensors for RGB components so they do not have this low-pass filter and may produce clear images. The second issue is related to the size of the sensor. The larger photon sensor size results in better dynamic range. Current digital cameras can easily under-/overexpose due to their small photon sensor size on a crowded CCD/CMOS chip.

**Focussing and zooming:**Focus is an important factor in case of digital imaging devices. Character recognition and segmentation requires sharp edge response. At short distances and large apertures, even very minute perspective changes can cause uneven focus.

**Non planar surfaces:** pages of an opened book are flat and curled along the spline of the book. This warping causes many disturbances and cannot be readable by the optical character recognition systems.

**Low resolution:** Low resolution is another problem. The images captured by digitalcameras are of low resolution and cannot be readable by OCR systems.

**Complex backgrounds:** More complex background makes segmentation of the image more difficult. If the document image is of irregular shape, the segmentation becomes more difficult.

**Uneven lighting:** Uneven lighting is a major issue in case of image capturing. A camera has far less control of lighting conditions on an object than scanners. If on camera flash is used, the centre of the view is the brightest, and then lighting decays outward.

**Wide-angle-lens distortion:** When an object gets closer to the image plane, focus, lighting and layout problems occur on the periphery. Since many focus-free and digital cameras come with a cheap wide-angle lens, distortion can be a problem if they are used for document analysis.

**Sensor noise:** Dark noise and read-out noise are the two major sources of noise at the CCD/CMOS stage in digital cameras. Additional noise can be generated in amplifiers. The high shutter speed, the small aperture, the dark scene, and the high temperature make the noise level so high. Compared to digital cameras, scanners normally have less to worry in all these aspects. Fig 3.3 shows an image scanned at a particular resolution and the different types of distortions in that image.

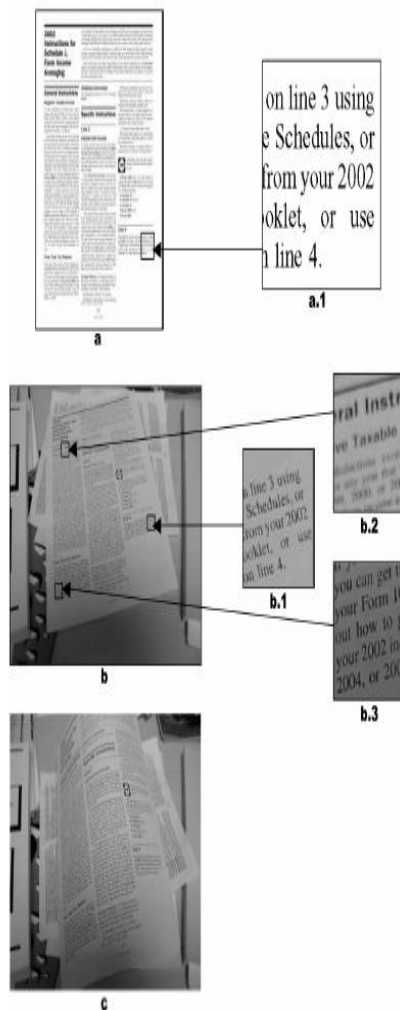


Fig 3.3: (a) document scanned at 300 dpi (b, c) same image captured by digital camera with perspective distortion, complex background and warping (a.1, b.1) small portion of camera captured and scanned image at full resolution (b.2, b.3) two small portions of image b at full resolution.

### 3.4 CAMERA BASED ACQUISITION OF DOCUMENT IMAGES

Main important property that differ a camera from a scanner is that we can capture images from a certain distance. The image can be zoomed and can capture the information but zooming at certain distance causes resolution problems. In order to increase the clarity of the image, the image should be sliced and capture the images of these partitioned texts. After that mosaicing techniques are used to combine all the texts of the document.

#### 3.4.1 AUTOFOCUS AND ZOOMING

In [11], Mirmehdi et al. propose an approach for general recognition problems by auto zooming. The variance in the window of observation can be used as an indicator for best zoom if the background variance is less. In [12], Zandifar et al. discussed auto focusing problems in designing a text reading system for the visually impaired. It is analysed that the best focus can be achieved when the image have sharper edges. Mirmehdi et al. [13] explains a system that can automatically locate text documents in a zoom-out view and control the camera to pan, tilt, and zoom in to get a closer look at the document. It is assumed that the documents are directly facing the camera so there will be no perspective distortion. The whole documents partitioned into several parts, and the camera captures each part after panning, tilting, and zooming. The divided parts are put together by mosaicing to obtain a complete document image, which is sent to an OCR package. The auto focusing and zooming problem is a very interesting one since it has direct application in robots.

#### 3.4.2 IMAGE MOSAICING

Jung et al. [14] use mosaicing technique to put together long text strings that appear in multiple video frames into a panorama image. In the CamWorks project [15], mosaicing is used to put together the images of the upper and lower part of a document page. In [16], a desktop OCR system using a PC-cam is described where the camera is placed on top of a desk pointing downwards but the

camera captures only a small part of an A4 document. The user moves the document while monitoring the computer screen until every part of the page appears in the sequence. During the capturing, frames are selected such that they are substantially different and yet successive ones overlap. This reduces the number of frames used in image registration and reduces blur that can result from the combination of too many images.

### 3.4.3 IMAGE COMPRESSION

Images and videos require large amount of space. In such cases compression is very important. Zunino and Rovetta [17] design a vector quantization (VQ) mechanism for license plate images. This method not only compresses images but also gives information as to the location of the plate in images.

### 3.5 PROCESSING CAPTURED IMAGES

While considering the processing of captured images, we want to consider the differences between processing image and video text and processing images of structured documents. This difference will have an impact on techniques needed to process them. Unlike images of structured documents, texts in images and videos are only a subset of a vast number of images or video frames, and detection of texts may be nontrivial. Overall, the procedures involved with processing document images or images of text will require text detection, localization, extraction, geometrical normalization, enhancement/binarization, and recognition.

### 3.6 MULTI FRAME PROCESSING

In the captured image processing section, issues related to the processing of a single image known to contain text are explained. Often, however, when processing a sequence of images, there are both new challenges and advantages. The most common case is the well-known video text analysis, but the user may also simply take two or more pictures of the same document, each picture containing either the whole document or part of it. The motivation

may be to make sure a clear copy is obtained or a high enough resolution obtained. Many of the same problems are shared by all these cases including frame selection, text tracking, and multi frame enhancement.

### 3.7 CAMERA BASED APPLICATIONS

Over the past 30 years, there have been numerous applications on camera-based text recognition, such as reading license plates, book sorting [18], visual classification of magazines and books, reading freight train IDs, road sign recognition, detection of danger labels, and reading signs in warehouses. Fig 3.4 shows some examples of camera based applications. In addition to these types of applications, the ability to process signs using mobile, low-cost hardware enables numerous other applications.

Mobile text recognizer and speech generator for the visually impaired: camera-based OCR techniques can be used in a head-mounted smart video camera system to help the visually impaired. It helps to detect and recognize text in the environment and then convert text to speech. The problems they confront on the vision side include the detection of text and the adjustment of cameras (such as zooming) so clear focus can be achieved.

Text acquisition: Text acquisition can be implemented to small levels. For example, while barcodes are widely used, they have the disadvantage of not being readable to humans and require expensive, specialized laser readers. A recent trend is to enhance barcode readers depending on PDAs and cameras [19]. The ability to capture and recognize text would be a further useful complement to barcode readers. Similarly, in the package delivery industry, it would be helpful to recognize addresses and automatically route them to an appropriate destination.

Document archiving: High quality digital cameras have been used for digitizing large historical manuscripts. As for consumer-grade equipment,

due to their flexibility and independence of bulky computers, it will not be surprising to find digital cameras and camcorders being used as document digitizing and archiving devices in the future. A user can carry such a device conveniently anywhere and record interesting document pages instantly.

Cargo container and warehouse merchandise code reader: Lee and Kankanhalli [20] present a system used in ports to automatically read cargo container codes. A single greyscale image captured by a camera is provided for reading container codes. The uneven surface may make text look warped. Their text detection is based on vertical edges found in the image and a verification stage uses domain knowledge that container codes are standardized in a fixed format, four letters followed by seven digits in one or two lines.

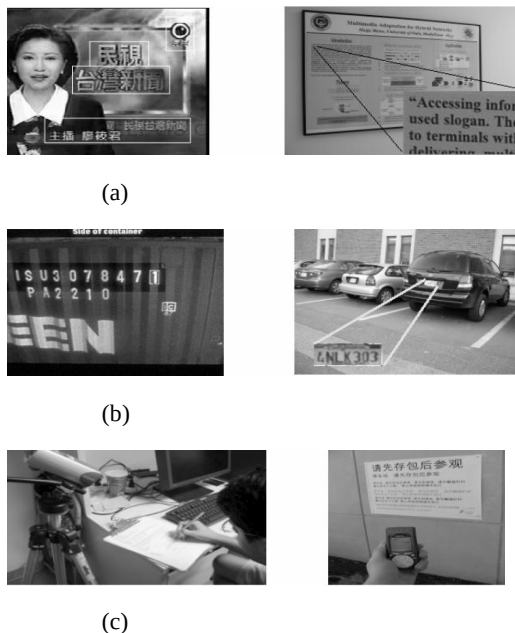


Fig 3.4: camera based applications: (a)Video caption text recognition (b) Cargo container code (c) Camera-based handwriting recognition (d) Poster capturing. (e) License plate reading (f)Sign translation.

#### 4. DEWARPING OF DOCUMENT IMAGES

Investigations on document analysis and recognition have conventionally been focused on analysing scanned documents. Digitizing analogue media is an important process in the field of media preservation. Digital media proffers numerous benefits compared to its physical counterpart, such as less physical storage space, and increased accessibility and functionality. By using less expensive commodity equipment and software, the KB has reduced the investments necessary for digitization. To further decrease the cost of digitization, the KB aims to lessen the manual labour needed during the process. At the time of writing, the printed media at the KB is digitized such that there is a distinct margin between the object and the borders of the digital image. The redundant area captured leads to higher digital storage needs, and may also result in difficulties in further processing steps, such as optical character recognition (OCR).

In many cases, the images captured will not be on flat surfaces, they may be on curved surfaces. In the case of images captured using cameras, they can take the form of any arbitrary shapes. The straight lines will be appeared as curled lines and the rectangular and square shapes will be deformed. These deformations are strictly non-linear and cannot be explained as linear transformations as in the case of perspective distortions.

Many novel approaches have been introduced over the years for performing page segmentation and optical character recognition (OCR) on scanned documents. With the emergence of digital cameras, the traditional way of capturing images is changing

from flat-bed or sheet bed scans to capture by hand-held digital cameras. These hand held cameras are more flexible than traditional scanners. Recognition and segmentation of documents captured with hand-held cameras have many technical challenges like perspective distortion, non-planar surfaces, low resolution, uneven lighting, zooming and focussing, complex backgrounds and wide-angle lens distortions. The fatal distortion that mostly happens in camera-captured document analysis is to deal with the page curl and perspective distortions. Current document imaging analysis and optical character recognition (OCR) systems do not expect these types of degradations, and show very poor performance when applied directly to camera-captured document images. The aim of page dewarping is to flatten a curled camera captured image such that it becomes readable by current OCR systems.

OCR systems provide a full alphanumeric representation of hand written or printed documents at electronic speed by scanning them. The document images are scanned by the scanner in the OCR systems and then the data is analysed and interpreted and converts into machine readable formats. Thus OCR helps the user to quickly automate the content and eliminates key strokes and maintains high level of accuracy. Intelligent character recognition systems are modules of optical character recognition systems which also converts data into machine readable formats. Fig 4.1 shows examples of warped images.

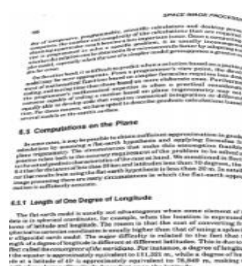
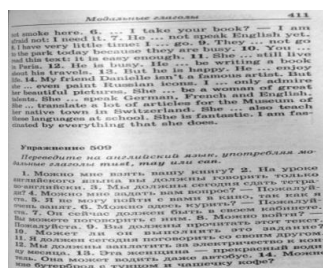


Fig 4.1: warped images

Over the past years, many different approaches have been proposed for document image dewarping. These approaches can be broadly categorized into two according to the acquisition of images.

- 1) 3-D shape reconstruction of the page using specialized hardware like stereo-cameras, structured light sources, or laser scanners.
- 2) 2-D reconstruction of the page using a single camera in an uncontrolled environment.[21]

The first approach for page dewarping was those based on 3-D shape reconstruction where processing is done with the aid of hardware. This is one of the major drawbacks of these approaches such that they require specialized hardware. These hardware limit the flexibility of capturing documents with cameras. Therefore, the approaches based on a single camera in an uncontrolled environment have caught more attention recently.

## 5. COARSE TO FINE DEWARPING OF DOCUMENT IMAGES USING ENCLOSED BOX METHOD

Image warping is a common problem in the case of scanning or capturing document images from thick volume books or from huge historical manuscripts. Warping causes shade along the spline of the book and also causes curliness on the text lines. This reduces the OCR accuracy and also impairs the readability of the user. Dewarping methods are used for flattening the curled document images and to rectify the distortions in the document image. In this section, Coarse to fine dewarping with enclosed box method is explained for dewarping document images. Image binarization (threshold selection) is the first step in this process. It converts a gray scale image into a binary image. Second step concerns with the detection of noisy black border and removal. Third step includes the corner detection of the curled document image. Coarse dewarping which is the fourth step deals with the transformation of the curled document image with a rectangular model. After that each of the letters in the text document is enclosed within rectangular boxes. This is the fifth step. Then fine dewarping is done for better dewarping results. Finally, image enhancement is done with morphological operators to enhance the document image. Fig 5.1 explains different steps in the dewarping process.

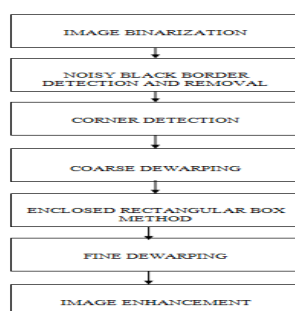


Fig 5.1: Flow chart of coarse to fine dewarping using enclosed box method

### 5.1 IMAGE BINARIZATION

Document image binarization refers to the conversion of a gray-scale image into a binary image. It is the initial step in most of the document imaging analysis. Usually, it differentiates text areas from background areas, so it is used as a text locating technique [22]. Binarization plays a prominent role in document analysis since its performance affects the character segmentation and recognition results. When processing distorted document images, binarization is a hilarious task. Distortions appear quite often and may occur due to several reasons which range from the accession source type to environmental conditions. Examples of degradation influence may include the appearance of variable background intensity, shadows, smudge and low contrast. The binarization scheme consists of five basic steps. The first step includes a dedicated denoising procedure using a low-pass Wiener filter. Here uses an adaptive Wiener method based on local statistics. In the second step, rough estimation of foreground regions is done. Next, as a third step, the background surface of the image is calculated by interpolating neighbouring background intensities into the foreground areas that result from the previous step. In the fourth step, final binarization is done by combining information from the calculated background surface and the original image. Text areas are located if the distance of the original image from the calculated background overshoot a threshold. This threshold adapts to the gray-scale value of the background surface in order to preserve textual information even in very dark background areas. In the last step, we proceed to a post-processing that eliminates noise, improves the quality of text regions and preserves stroke connectivity [22]. Fig 5.2 shows the block diagram of image binarization.

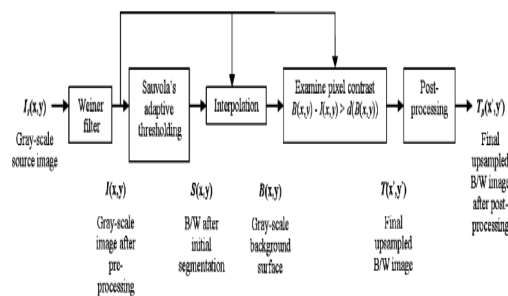


Figure 5.2: Block diagram of image binarization

### 5.1.1 PRE- PROCESSING

Pre- processing stage is required for poor quality degraded image. Pre –processing procedure improves the quality of the image by smoothing the image and by removing noise from the image. Weiner filter is used in this stage for image restoration. This adaptive wiener filter works by calculating the neighbouring pixel.

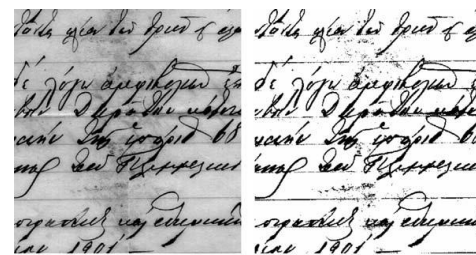
The gray scale image  $I$  is converted to filtered image  $I$  as follows:

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

Where  $\mu$  is the local mean,  $\sigma^2$  is the variance and  $v^2$  is the average of all variance calculated from the neighbourhood of each pixel.

### 5.1.2 CALCULATION OF FOREGROUND REGIONS

In this step, foreground and background regions are segmented. Sauvola's approach is used for segmenting the foreground and background regions. Here, from the image  $I(x, y)$  calculating the foreground regions in which the foreground regions have the value one and after that calculating the binarized image  $S(x, y)$ . Fig 5.3 shows adaptive thresholding using Sauvola's approach.



(a)

Fig 5.3: Adaptive thresholding using Sauvola's approach: (a) original image; (b) Rough estimation of foreground regions.

### 5.1.3 BACKGROUND SURFACE ESTIMATION

Background surface estimation should be equal to the difference between the original image to the foreground estimated image. Background surface estimation is done by neighbourhood pixel interpolation.

### 5.1.4 FINAL THRESHOLDING

Final thresholding is done by combining the background regions with the processed image. Text areas are calculated if the distance from the processed image with the background area exceeds a particular threshold  $d$ . Threshold value for dark region is selected small because the threshold value  $d$  changes according to the gray scale value of the background region in order to preserve textual information in dark background regions. The final binary image  $T(x, y)$  is given by :

$$T(x, y) = \begin{cases} 1 & \text{if } B(x, y) - I(x, y) > d(B(x, y)) \\ 0, & \text{otherwise} \end{cases}$$

### 5.1.5 UP SAMPLING

Upsampling is done for image enhancement. It is done by bicubic interpolation. It estimates the value of a pixel by calculating neighbouring 16 pixel

values surrounding that pixel. Theupsampled image is calculated as follows:

$$T(x',y') = 1 \text{ if } B(x,y) - I(x',y') > d(B(x,y))$$

0, otherwise

Fig 5.4 shows an input document image and fig 5.5 shows binarized document image which is the first step in the dewarping process.

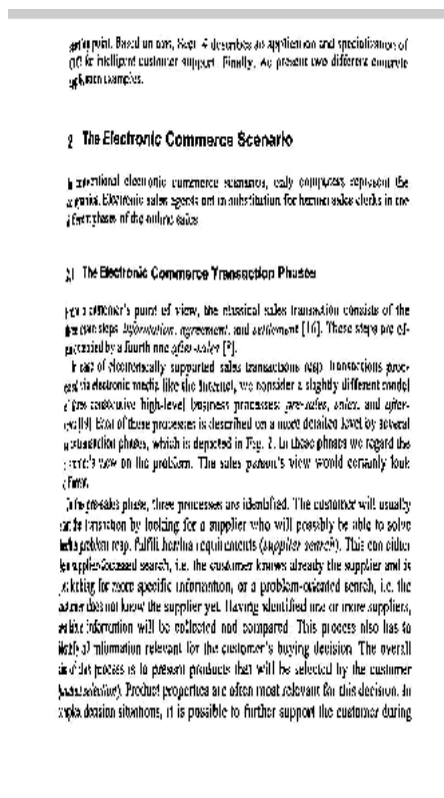


Fig 5.4: input image

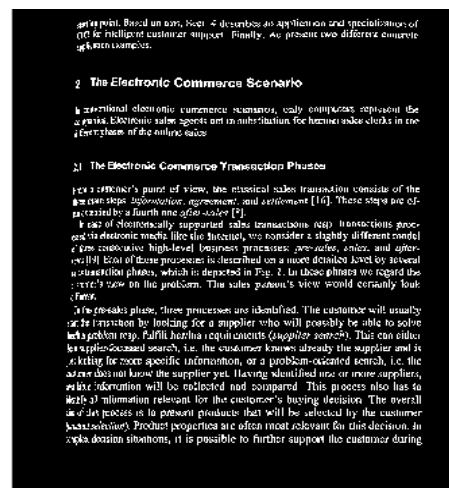


Fig 5.5: binarized image

## 5.2 AUTOMATIC BLACK BORDER DETECTION AND REMOVAL

Document images may have a noisy black border or contains noisy text regions from neighbouring pages when captured by a digital camera. Approaches proposed for document segmentation and character recognition usually consider ideal images without noise. However, there are many factors that may generate imperfect document images. When a page of a book is captured by a camera, text from an adjacent page may also be captured into the current page image. These unwanted regions are called "noisy text regions". Additionally, there will usually be black borders in the image. These unwanted regions are called "noisy black borders". All these problems influence the performance of segmentation and recognition processes. In this stage, noisy black borders (vertical and horizontal) of the image are detected and removed. This method is mainly based on horizontal and vertical profiles. In order to calculate the borders, first proceed to an image smoothing, then calculate the starting and ending offsets of borders and text regions and then

calculate the borders limits. The final clean image without the noisy black borders is calculated by using the connected components of the image [23]. Figure 5.6 represents the flow chart for black border detection and removal.

### 5.2.1 NOISY BLACK BORDER DETECTION AND REMOVAL

RLSA: Horizontal and vertical smoothing is done with the use of the Run Length Smoothing Algorithm (RLSA). This algorithm examines the white runs in the horizontal and vertical direction. For each direction, white runs with length less than a threshold are eliminated.

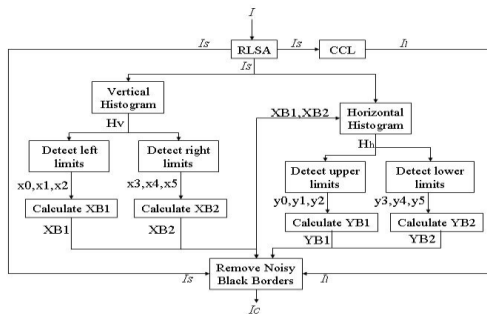


Fig 5.6: Flowchart for noisy black border detection and removal

The empirical value of horizontal and vertical length threshold is four pixels. The resulting image is  $I_s(x, y)$ .

CCL (Connected Component Labelling): Calculate the connected components of the image  $I_x$ .

The image consists of CS connected components  $c_i$ .

$$I(x, y) = \begin{cases} iif(x, y) \in c_i, 0 < i < CS \\ 0, \text{ otherwise} \end{cases}$$

Vertical Histogram: Calculate vertical histogram  $H_v$ , which is the sum of black pixels in each column.

$$H_v(x) = \sum_{y=0}^{I_y-1} I_s(x, y) \text{ where } 0 < x < I_x$$

Detect left limits: Detect vertical noisy black borders in the left side of the image. Initially searching for the start and the end ( $x_0, x_1$ ) of the left vertical black border. Calculate  $X_0$  as follows:

$$X_0 = \min(x) : (H_v(x) \geq L_1) \text{ or } (H_v(x) \leq L_2) \text{ where } 0 < x < I_x/5$$

The first condition  $H_v(x) \geq L_1$  is satisfied when the black border starts from the left side of the image, which is the most usual case while the second condition,  $H_v(x) \leq L_2$  is satisfied when white region exists before the black border. If we don't find any  $x_0$  that satisfies the above conditions we set  $x_0 = -1$ ,  $x_1 = 1$  and  $x_2 = -1$  and stop this process.

Calculate  $XB_1$ : Calculate left limit ( $XB_1$ ) of text regions.

$$\begin{cases} XB_1 = 0 & \text{if } X_0 = -1 \\ X_0 + (x_1 - x_0)/2 & \text{if } x_2 = -1 \\ X_1 + (x_2 - x_1)/2 & \text{if } x_2 \neq -1 \end{cases}$$

A similar process is applied in order to detect the vertical noisy black border of the right side of the image as well as the right limit  $XB_2$  of text regions.

Horizontal Histogram: Calculate horizontal histogram  $H_h$ , which is the sum of black pixels in each row at  $XB_1$  and  $XB_2$  limits.

$$H_h(y) = \sum_{x=XB_1}^{XB_2} I_s(x, y) \text{ where } 0 \leq y \leq I_y$$

A similar process as for the vertical noisy black borders are applied in order to detect the horizontal

noisy black borders as well as the upper (YB1) and bottom (YB2) limits of text regions.

**Remove Noisy Black Borders:** All black pixels that belong to the connected component which includes at least one pixel that is outside the limits are transformed into white.

$I_c(x,y) = 0$  if  $I_l(x,y) = i$  and  $(x_1 < XB_1$  or  $X_1 > XB_2$  or  $y_1 < yb_1$  or  $y_1 > yb_2$   
 $I(x,y)$  otherwise.

### 5.2.2 NOISY TEXT REGION DETECTION AND REMOVAL

Noisy text regions of the image are detected and removed in this stage. Here initially an image smoothing is done. The detailed flowchart representing the steps are shown in fig 5.7

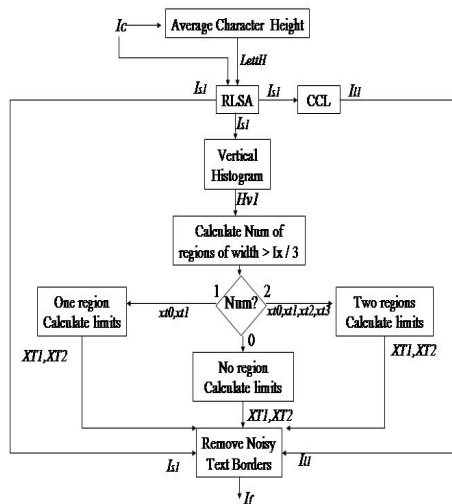


Fig 5.7: Noisy text region detection and removal

**RLSA:** Vertical and horizontal smoothing is done here with the aid of dynamic parameters which depends on the average character height.

**Vertical histogram:** Vertical histogram is calculated as follows:

$$Hv1(x) = \sum Is1(x, y)$$

Then calculate the number of regions with width  $> Ix'/3$ ; Check the total number of consecutive X which satisfies this condition:

$$Hv1(x) > L4; W = 1/3 * Ix$$

Two region calculate limits: Here signal correlation of two regions are calculated.

$$Sc(a,y) = 1 - 2/m \sum_{k=0}^M (Is1(k,y) \text{ xor } Is1(k,y+a))$$

Where 'M' is the region width and 'a' is the distance between two lines.

Then calculate limits XT1 and XT2:

If  $(XT1 = xt0 \text{ \& } XT2 = xt3)$  else if  $(SC0 < SC1)$

then  $(XT1 = Xt0 \text{ \& } XT2 = xt1)$  else  $(XT1 = xt2 \text{ \& } XT2 = xt3)$

One region calculate limits: Here dividing the region into eight segments and calculating the signal correlation of each regions  $(SC1...SC8)$ .

- 1) If  $SC1 < .5$  and  $SC8 < .5$ , there should be no text regions, so  $XT1 = xt0$  &  $XT2 = xt1$ .
- 2) If  $SC1 > .5$ , searching for last consecutive region i where  $SCi > .5$  and finding an x' where HV1 is minimum.
- 3) If  $(Xt1 - xt') > W$  then  $xt1 = x'$  &  $XT2 = xt1$  else  $XT1 = xt0$  &  $XT2 = xt1$
- 4) If  $(x' - xt0) > W$  then  $XT1 = xt0$  and  $XT2 = x'$  else  $XT1 = xt0$  &  $XT2 = xt1$
- 5) If  $SC8 > .5$ , searching for last consecutive region of i where  $SCi > .5$ , then finding x' where HV1 is minimum.
- 6) If  $(x' - xt0) > W$ , then  $XT1 = xt0$  &  $XT2 = x'$  else  $XT1 = xt0$  &  $XT2 = xt1$

No region calculate limits: Here, the text region consists of two or more columns. This step

calculates the noise in this region. In this stage, checking whether the condition  $(HV1(x) > L4) > W/4$  is satisfied. If two or less regions satisfy this condition, then  $XT1 = XB1$  &  $XT2 = XB2$ . If three or more region satisfy this condition, then correlation is taken. Remove noisy text region: All black pixels which is not in the limit  $XT1$  and  $XT2$  is converted into white pixels.

After that the final image is calculated as follows:

$$If(x,y) = I_c(x,y) \text{ If } I_i(x,y) = i \in (x1, y1): (x1 > XT1) < XT2 \text{ \& } I_i(x1, y1)$$

0, otherwise

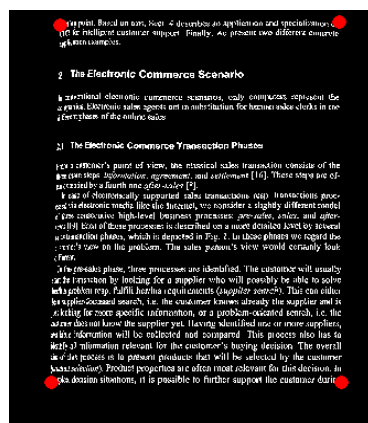


Fig 5.8: detected corners of the image

### 5.3 COARSE DEWARPING

In this step, a transformation model which maps the projection of the curved surface to the 2D rectangular area is applied. The extraction of the curved surface is achieved from the left, right boundaries and top and bottom curled lines. At first the borders are detected [24]. Let  $NL$  denotes the number of lines in the curled image and  $CH$  denotes the height of each character. Fig 5.8 shows an example of extracted curved surface.

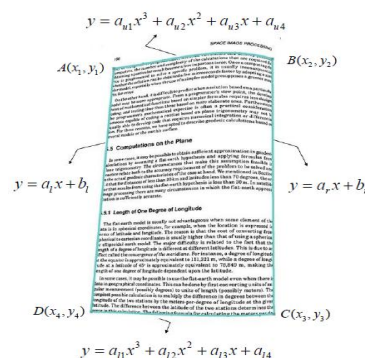


Fig5.9: Extraction of curved surface projection

#### 5.3.1 ESTIMATION OF CURVED SURFACE

After identifying the text lines, all the corner points should be detected.  $A(X1,Y1)$ ,  $B(X2,Y2)$ ,  $C(X3,Y3)$ ,  $D(X4, Y4)$  [24].

##### Left/Right line Estimation

Here, all the leftmost points in each text line are detected:  $(Xli, yli), 0 < i < NL$ . After that we calculate the average value of  $Xl$  and every point of it is eliminated if it does not satisfy this condition:

$$|Xli - Xl| > 2 \cdot CH$$

This condition is used to eliminate the subtitles, titles etc of the document. Least square estimation method is used to get straight line AD. After this process, the straight line AD is defined as:

$$Y = alx + bl$$

Consequently, the straight line BC, is defined as

$$Y = arx + br$$

##### Top/bottom curved line estimation

Assume that  $Dli$  be the distance between the leftmost point of text line  $i$  and the straight line AD, and let  $Dri$  be the distance between the

rightmost point of text line  $i$  and the straight line  $BC$ . After that, applying the condition which satisfies that the text line selected is not too small, not a title or not a subtitle etc. Then detect all the upper most points  $(x_{ui}, y_{ui})$  of the text line by the previous step explained earlier. After this process, the curved line  $AB$  is defined as:

$$Y = au_1x^3 + au_2x^2 + au_3x + au_4$$

Consequently, the curved line  $DC$  is defined as:

$$Y = al_1x^3 + al_2x^2 + al_3x + al_4$$

Fig 5.9 shows the detected borders of the document image using the method explained earlier.

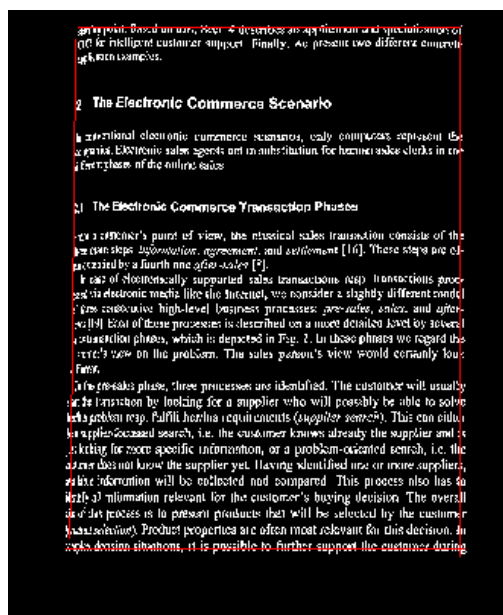


Fig 5.10: Detected borders of the image

### 5.3.2 RECTANGULAR TRANSFORMATION MODEL

In this step, the projection of the curved surface is transformed with the 2D rectangular area. Let  $A'$  ( $X1'$ ,  $Y1'$ ),  $B'$  ( $X2'$ ,  $Y2'$ ),  $C'$  ( $X3'$ ,  $Y3'$ ),  $D'$  ( $X4'$ ,  $Y4'$ ) represents the points on the rectangular area. Let  $\widehat{AB}$  be the arc length between points  $A$  and  $B$  and  $|AB|$  represent the Euclidian distance between points  $A$  and  $B$ .

Here, we want to calculate the width  $W$  of the rectangle. It is calculated as follows:

$$W = \widehat{AB}, \widehat{DC}$$

Height of the rectangular area is calculated as:

$$H = \min(|AD|, |BC|)$$

After that we want to calculate the corner points of the rectangle as follows:

$$\begin{aligned} x1' &= x1 & y1' &= y1 \\ x2' &= x1' + W & y2' &= y1' \\ x3' &= x2 & y3' &= y2' + H \\ x4' &= x1' & y4' &= y3' \end{aligned}$$

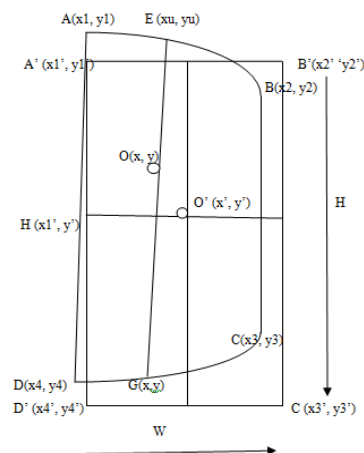


Fig 5.11: Transformation model

Our aim is to represent the points in the curved surface to the points in the rectangular area. Fig

5.10 shows the transformation model which maps the curved surface area. Consider the point  $O(x, y)$ . Let us transform this point to the rectangular area by calculating new position  $O'(x, y)$  for  $O(x, y)$  as follows:

$$X' = x1' + |A'Z|$$

$$Y' = y1' + |A'H|$$

Where  $|P'Z|$ ,  $|P'H|$  are calculated as follows:

$$\begin{aligned} \frac{|AQ|}{|AE|} &= \frac{W}{|A'Z|} \\ &= |A'Z| = \frac{W}{|AQ|} * |AE| \end{aligned}$$

$$\begin{aligned} \frac{|EG|}{|EO|} &= \frac{H}{|AH|} \\ &= |A'H| = \frac{H}{|EG|} * |EO| \end{aligned}$$

Repeating this procedure for all the points in the curved surface area. Finally each and every point on the curved surface area is transformed to new points in the rectangular area. Figure 5.11 shows output image after coarse dewarping.

part. Based on this, Sect. 4 describes an application to a specialization of the present customer support. Finally, we present two different concrete application examples.

### 3 The Electronic Commerce Scenario

In traditional electronic commerce scenarios, only computers represent the partners. Electronic sales agents act in substitution for human sales clerks in the different phases of the online sales.

#### 3.1 The Electronic Commerce Transaction Phases

From a customer's point of view, the classical sales transaction consists of the following steps: information, agreement, and settlement [26]. These steps are often extended by a fourth one, after-sales [8].

In case of electronically supported sales transactions (resp. transactions processed by electronic media like the Internet), we consider a slightly different model of four consecutive high-level business processes: pre-sales, sales, and after-sales [19]. Each of these processes is described on a more detailed level by several transaction phases, which is depicted in Fig. 1. In these phases we regard the customer's view on the problem. The sales person's view would certainly look different.

In the pre-sale phase, three processes are identified. The customer will usually start a transaction by looking for a supplier who will possibly be able to solve his problem (e.g. fulfill her/his requirements (supplier search)). This can either be a supplier-focused search, i.e. the customer knows already the supplier and is looking for more specific information, or a problem-oriented search, i.e. the customer does not know the supplier yet. Having identified one or more suppliers, sales information will be collected and compared. This process also has to take all information relevant for the customer's buying decision. The overall goal of this process is to present products that will be selected by the customer (product selection). Product properties are often most relevant for this decision, in order to make a decision situation, it is possible to further support the customer during

Fig 5.12: image after coarse dewarping

## 5.4 ENCLOSED RECTANGULAR METHOD

After the coarse dewarping procedure, all the lines of the curved document image are not straightened. So, applying an enclosed rectangular box method to flatten all the characters in the document image. All the characters in the document image are enclosed within rectangular boxes. To construct the enclosed rectangular box, the slope of the image and the distance between the base lines should be known. The upper and lower baselines are used for considering the top and bottom lines of the rectangular box. The vertical lines in the rectangular box are estimated by the neighbouring enclosed boxes. Fig 5.12 is an example of enclosed rectangular boxes.



Fig 5.13: Enclosed rectangular boxes

The width of each cell is different for different characters. The correct width should not be similar to the input image in most of the cases because of different types of distortion in the image. In ideal cases, the width should be equal to the width of the input image. The width of each of the cell should be equal to the Euclidean distance between left and right corner points.

The line spacing of each of the characters is also an important factor while considering enclosed boxes. The top point and bottom point of each of the characters should be known. Many of the characters have ascenders and descenders. The letter P is a descender and d is an ascender.

The depth value of each of the characters is also an important factor. Considering that the objects are of constant orientation and line spacing, size of each of the object in the image depends on the distance from focus of the camera.

Considering all these factors, enclosed rectangular boxes can be built for the curled characters in the document image.

### 5.5 FINE DEWARPING

In fine dewarping, word level dewarping is done. Here, first detect all the text lines and words. For this, remove all the non text components using the condition:

$$Height > 3 * CH \text{ or } Height < \frac{CH}{4} \text{ or } Width < \frac{CH}{4}$$

After that, upper baseline of the word is defined as:

$$y = a_{ij} + b_{ij}$$

Then all the words are rotated and translated as follows:

$$y^{rs} = y^r + d_{ij}$$

$$x^{rs} = x = x^r$$

Where

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

$$d_{ij} = \begin{cases} y_{i0}^{ru} - y_{ij}^{ru}, & \text{if } |\phi_{ij}^u| - |\phi_{ij-1}| < |\phi_{ij}^l| - |\phi_{ij-1}| \\ y_{i0}^{rl} - y_{ij}^{rl}, & \text{otherwise} \end{cases}$$

$$y_{ij}^{ru} = (a_{ij} x_{ij} + b_{ij}) * \cos \phi_{ij}$$

$$y_{ij}^{rl} = (a_{ij} x_{ij} + b_{ij}) * \sin \phi_{ij}$$

Where  $\phi_{ij}$  is the slope of the word and  $x_{min}^{ij}$  is the left side of the enclosed rectangular box. At last all the components that we have been removed are added.

### 5.6 IMAGE ENHANCEMENT

Morphological operators are used for image enhancement in curled document images. They affect the structure, layout or shape of an image. The two significant morphological operations are dilation and erosion. Object expansion can be done by dilation. It potentially fills in small holes and connects disjoint objects. Erosion shrinks objects by etching away (eroding) their boundaries. By the proper selection of structuring element these operations can be customized for an application, which determines exactly how the objects will be dilated or eroded [25].

### 5.6.1 DILATION

The dilation process is performed by placing the structuring element B on the image A and sliding it over the image in a manner similar to convolution. The difference is in the operation performed. No change occurs, if the origin of the structuring element coincides with a 'white' pixel in the image; Then move to the next pixel. If the origin of the structuring element concurs with a 'black' in the image, make black all pixels from the image covered by the structuring element.

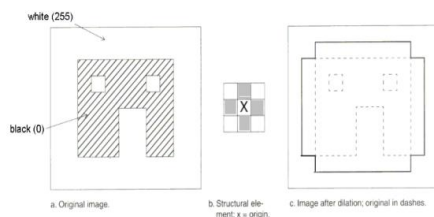


Fig. 5.14: Illustration of the dilatation process

An example is shown in Fig. 5.13. With a dilation operation, all the 'black' pixels in the original image will be retained, any boundaries will be expanded, and small holes will be filled.

### 5.6.2 EROSION

The erosion process is similar to dilation. Here returning pixels to 'white', not 'black'. The processes involved are: 1) If the origin of the structuring element falls over a 'white' pixel in the image, there is no change; move to the next pixel 2) If the origin of the structuring element coincides with a 'black' pixel in the image, and at least one of the 'black' pixels in the structuring element falls over a white pixel in the image, then change the 'black' pixel in the image (corresponding to the position on which the centre of the structuring element falls) from 'black' to a 'white'. Fig 5.14 shows the illustration of the erosion process

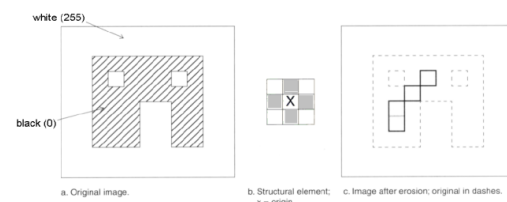


Fig 5.15: Illustration of the erosion process

### 5.6.3 OPENING AND CLOSING

When dilation and erosion are combined, complex sequences can be formed. The most useful of these for morphological filtering are called opening and closing. Opening consists of an erosion followed by a dilation and can be used to eliminate all pixels in regions that are too tiny to contain the structuring element. In this case, the structuring element is often called a probe, because it is probing the image looking for small objects to filter out of the image. Fig 5.15 shows the illustration of the opening process.

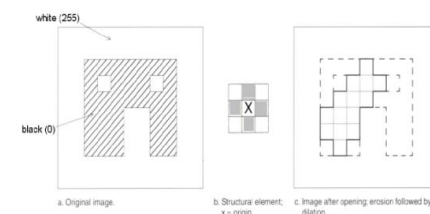


Fig. 5.16: Illustration of the opening process  
Closing consists of a dilation followed by erosion and can be used to fill in holes and small gaps which is shown in fig 5.16. Closing operation has the effect of filling in holes and closing gaps. The order of operation is important. Closing and opening will generate different results even though both consist of erosion and dilation[25].

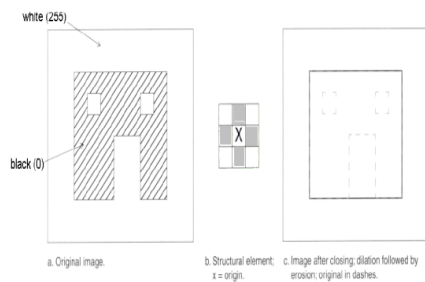


Fig. 5.17: Illustration of the closing process.

## 5.6.4 MORPHOLOGICAL ALGORITHMS

### 5.6.4.1 BOUNDARY EXTRACTION

The boundary of a set  $A$ , denoted by  $\beta(A)$  can be obtained by first eroding  $A$  by  $B$  and then calculating the set differences between  $A$  and its erosion.

$$\beta(A) = A - (A \odot B)$$

Where  $B$  is a suitable structuring element and ‘ $-$ ’ is the difference operation on sets which is shown in fig 5.17.

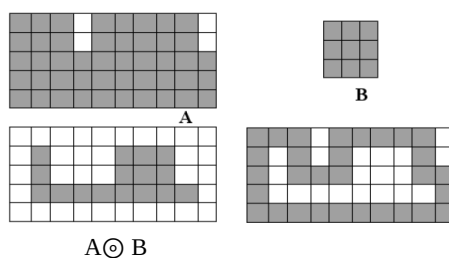


Fig. 5.18: Illustration of the boundary extraction algorithm

### 5.6.4.2. REGION FILLING

This is a simple algorithm for region filling based on set dilations, complementation, and intersections. Beginning with a point  $p$  inside the boundary, the objective is to fill the entire region with ‘black’. If we adopt the convention that all non-boundary (background) points are labelled ‘white’, then we assign a value of ‘black’ to  $p$  to begin. The following procedure then fills the region with ‘black’:

$$X_k = (X_{k-1} \oplus B) \cap A^c \quad k = 1, 2, 3...$$

Where  $X_0 = p$ ,  $B$  is the symmetric structuring element;  $\cap$  is the intersection operator

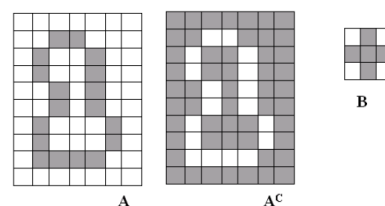


Fig. 5.19: Illustration of the region filling algorithm

algorithm. Based on our, Sect. 4 describes an application and specifications of (C) for intelligent customer support. Finally, we present two different concrete system examples.

#### 2) The Electronic Commerce Scenario

In traditional electronic commerce scenarios, only online sales represent the customer. Electronic sales agents act as substitution for human sales clerks in one-to-one interactions of the online sales.

#### 2.1) The Electronic Commerce Transaction Phases

From a customer's point of view, the classical sales transaction consists of the phases: product information, agreement, and settlement [16]. These steps are often realized by a fourth step: delivery [17].

In case of electronically supported sales transactions (e-s) transactions proceed via electronic media like the Internet, we consider a slightly different model of the customer's high-level business processes: pre-sale, sales, and after-sale [18]. Most of these processes is described in a more detailed level by several transaction phases, which is depicted in Fig. 2. In these phases we regard the customer's view on his problem. The sales person's view would certainly look different.

In the pre-sale phase, three processes are identified. The customer will usually start his search by looking for a supplier who will possibly be able to solve his problem (e.g. fulfill his/her requirements (supplier search)). This can either be a problem-oriented search, i.e. the customer knows already the supplier and is seeking for more specific information, or a problem-oriented search, i.e. the customer does not know the supplier yet. Having identified one or more suppliers, further information will be collected and compared. This process also has to do with information relevant for the customer's buying decision. The overall search process is to present products that will be selected by the customer (purchase decision). Product properties are often most relevant for this decision. In the decision phase, it is possible to further support the customer during

Fig 5.20: Curled input image

using point. Based on this, Sect. 4 describes an application and specialization of CEM for intelligent customer support. Finally, we present two different concrete application examples.

## 2 The Electronic Commerce Scenario

In conventional electronic commerce scenarios, only computers represent the companies. Electronic sales agents act in substitution for human sales clerks in the different phases of the online sales.

### 2.1 The Electronic Commerce Transaction Phases

From a customer's point of view, the classical sales transaction consists of the five main steps: *information, agreement, and settlement* [16]. Those steps are often extended by a fourth one *after-sales* [8].

In case of electronically supported sales transactions resp. transactions proceeding via electronic media like the Internet, we consider a slightly different model of three consecutive high-level business processes: *pre-sales, sales, and after-sales* [19]. Each of these processes is described on a more detailed level by several sub transaction phases, which is depicted in Fig. 1. In those phases we regard the customer's view on the problem. The sales person's view would certainly look different.

In the *pre-sales* phase, three processes are identified. The customer will usually start the transaction by looking for a supplier who will possibly be able to solve *customer problems* resp. fulfill her/his requirements (*supplier search*). This can either be a supplier-focused search, i.e. the customer knows already the supplier and is at looking for more specific information, or a problem-oriented search, i.e. the customer does not know the supplier yet. Having identified one or more suppliers, suitable information will be collected and compared. This process also has to verify all information relevant for the customer's buying decision. The overall task of this process is to present products that will be selected by the customer (*product selection*). Product properties are often most relevant for this decision. In complex decision situations, it is possible to further support the customer during

Fig 5.21: Dewarped image

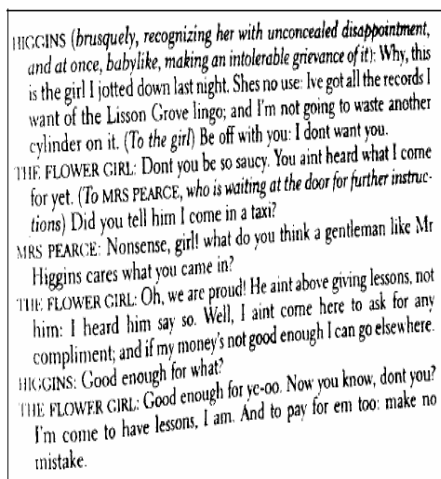


Fig 5.22: Portion of a curled document image

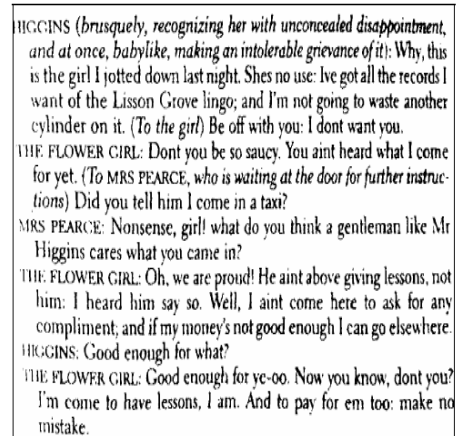


Fig 5.23: Dewarped output image

Packages 201  
Quasimodo 17  
Chapter 12

As your models get bigger, you will find that many use cases tend to cluster together in groups that are conceptually and semantically related. In the UML, you can use packages to model these clusters of classes.

## Hints and Tips

When you model use cases in the UML, every use case should represent some distinct and identifiable behavior of the system or part of the system. A well-structured use case

- Names a single, identifiable, and reasonably atomic behavior of the system or part of the system.
- Factors common behavior by pulling such behavior from other use cases that it includes.
- Factors variants by pushing such behavior into other use cases that extend it.
- Describes the flow of events clearly enough for an outsider to easily understand it.
- Is described by a minimal set of scenarios that specify the normal and variant semantics of the use case.

When you draw a use case in the UML,

- Show only those use cases that are important to understand the behavior of the system or the part of the system in its context.
- Show only those actors that relate to those use cases.

Fig 5.24: Warped image from a document

*Packages are discussed in Chapter 12.* As your models get bigger, you will find that many use cases tend to cluster together in groups that are conceptually and semantically related. In the UML, you can use packages to model these clusters of classes.

#### Hints and Tips

When you model use cases in the UML, every use case should represent some distinct and identifiable behavior of the system or part of the system. A well-structured use case:

- Names a single, identifiable, and reasonably atomic behavior of the system or part of the system.
- Factors common behavior by pulling such behavior from other use cases that it includes.
- Factors variance by pushing such behavior into other use cases that extend it.
- Describes the flow of events clearly enough for an outsider to easily understand it.
- Is described by a minimal set of scenarios that specify the normal and variant *sentences* of the use case.

When you draw a use case in the UML,

- Show only those use cases that are important to understand the behavior of the system or the part of the system in its context.
- Show only those actors that relate to these use cases.

Fig 5.25: Dewarped image of the document

Different types of curled document images with their output dewarped images are shown from fig 5.20 to fig 5.25.

## 5. RESULTS AND DISCUSSIONS

This paper work is implemented in MATLAB. In order to verify the validity of this work, hundred images of different resolution are taken. After applying a coarse dewarping which is done by transforming the curled document image with a rectangular model, not a better dewarped result is obtained. So, in order to obtain a better dewarped result, each of the characters in the document image is enclosed within rectangular boxes and a fine dewarping is done on word level.

Table 6.1 Angle measurements of dewarped images

| Image                           | Image 1 | Image 2 | Image 3 |
|---------------------------------|---------|---------|---------|
| Angle of original image         | 0.0800  | -1.9800 | -1.4300 |
| Rotated angle of dewarped image | 2.24300 | 3.1700  | 3.4300  |

Images with different warping amount are taken to check the angle variation of document images. The original angle of the warped image and the measured angle of the dewarped images are measured. This paper work can correctly rectify the distortions upto angle of -2.

OCR evaluation can be done for checking the accuracy of the image. Hundred images of different resolution before and after dewarping are taken for checking the accuracy of the image. OCR accuracy is defined as the ratio of number of correct characters (number of characters in document – number of errors identified) to the total number of characters in the document.

Hundred document images before and after the dewarping processes are scanned to check the OCR accuracy. The curled document images before dewarping produced poor result when read by an OCR engine. This paper work performed the OCR test using ABBYY finereader. After applying coarse and fine dewarping alone, they did not produce a better satisfactory result. The rectified image after applying coarse to fine dewarping with enclosed box methodology, the OCR accuracy is improved by 25%.

Table 6.2: OCR accuracy

|                                                         |        |
|---------------------------------------------------------|--------|
| Without dewarping                                       | 55.07% |
| With coarse to fine dewarping                           | 85.56% |
| With coarse to fine dewarping using enclosed box method | 93.98% |

This paperwork used data set containing different font sizes and different fonts. The methodology can dewarp document images irrespective of font size and font diversities. This work requires approximately 9sec to process one page.

## 6. CONCLUSION

Document imaging analysis has interest over past few years. Many types of distortion affects document images. The prominent one is the warping affect. Warping effect reduces the OCR accuracy and also OCR systems cannot read the curled document images. Many dewarping methods are introduced to straighten the curled document images. In this paper work, coarse to fine dewarping using enclosed box method is introduced. The curled surface area is projected by a rectangular transformation model to achieve coarse dewarping. In order to enhance dewarping at word level each of the characters in the document image are enclosed within rectangular boxes. Each of the letters is translated and rotated using fine dewarping to improve the dewarping result. Results show that this method can dewarp document images efficiently and improves the OCR accuracy.

## REFERENCES

- [1] F. Shafait, T. M. Breuel, "Document Image Dewarping Contest", In 2nd Int. Workshop on Camera Based Document Analysis and Recognition, Curitiba, Brazil, 2007, pp. 181-188.
- [2] J. Liang, D. Doermann, H. Li. "Camera-based analysis of text and documents: survey", International Journal on Document Analysis and Recognition, 7(2-3), 2005, pp. 84-104.
- [3] Masalovitch A, 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, 2007, pp. 45-53.
- [4] Z. Zhang , C. L. Tan, "Correcting document image warping based on regression of curved text lines", International Conference on Document Analysis and Recognition, Edinburgh, Scotland, 2003, pp. 589-593.
- [5] Shijian Lu, Chew Lim Tan, "The Restoration of Camera Documents through Image Segmentation", Workshop on Document Analysis Systems VIII, Nelson, New Zealand, 2006, pp. 484-495.
- [6] Z. Zhang & C. L. Tan, "Warped image restoration with applications to digital libraries", International Conference on Document Analysis and Recognition, Seoul, Korea, 2005, pp. 192-196.
- [7] HyungIl Koo, Jinho Kim, and Nam Ik Cho, "Composition of a Dewarped and Enhanced Document Image From Two View Images", IEEE Transactions On Image Processing, Vol. 18, No. 7, July 2009.
- [8] Yuandong Tian, Srinivasa G. Narasimhan "Rectification and 3D Reconstruction of Curved Document Images", in ICDAR, 2005
- [9] Newman W, Dance C, Taylor A, Taylor S, Taylor M, Aldhous T, "Cam Works: a video-based tool for efficient capture from source documents", In: Proc. international conference on multimedia computing and systems, pp 647-653, 1999.
- [10] Fisher F, Digital camera for document acquisition", In: Proc. symposium on document image understanding technology, pp 75-83, 2001.
- [11] Cai M, Song J-Q, Lyu MR "A new approach for video text detection", In: Proc. ICIP, pp 117-120, 2002.

- [12] Mirmehdi M, Palmer PL, Kittler J  
“Towards optimal zoom for automatic target recognition”, In: Proc.10th Scandinavian conference on image analysis, 1:447– 453, 1997
- [13] Zandifar A, Duraiswami R, Chahine A, Davis L, “A video based interface to textual information for the visually impaired”, In: Proc. IEEE 4th international conference on multimodal interfaces, pp 325–330, 2002.
- [14] Mirmehdi M, Clark P, Lam J, “Extracting low resolution text with an active camera for OCR”, In: Proc.IX Spanish symposium on pattern recognition and image processing, pp 43–48, 2001.
- [15] Jung K, Kim KI, Han J-H, “Text extraction in real scene images on planar planes”, In: Proc. ICPR, pp 469–472, 2002.
- [16] Newman W, Dance C, Taylor A, Taylor S, Taylor M, Aldhous T, “Cam Works: a video-based tool for efficient capture from source documents”, In: Proc international conference on multimedia computing and systems, pp 647–653, 1999.
- [17] Zappala A, Gee A, Taylor M, “Document mosaicing. Image Vis Comput: 17(8):585–595, 1999.
- [18] Zunino R, Rovetta S, “Vector quantization for license-plate location and image coding. IEEE Trans IndustElectr 47(1):159–167, 2000.
- [19] Gotoh T, Toriu T, Sasaki S, Yoshida M, “A flexible vision-based algorithm for a book sorting system, IEEE Trans Pattern Anal Mach Intell 10(3):393–399, 1988.
- [20] Moravec KLCA “grayscale reader for camera images of XEROX dataglyphs.”In: Proc. 13th BMVC, pp 698–707,2002.
- [21] N. Stamatopoulos, B. Gatos, I. Pratikakis, and S. J. Perantonis, “A two-step dewarping of camera document images,” in Proc. 8th Int.Workshop Document Analysis Syst., Nara, Japan, 2008, pp. 209–216.
- [22] B. Gatos, I. Pratikakis & S.J. Perantonis, “Adaptive Degraded Document Image Binarization”, Pattern Recognition, 39, 2006, pp. 317-327.
- [23] N. Stamatopoulos, B. Gatos, and A. Kesidis, “Automatic Borders Detection of Camera Document Images” In 2nd Int.Workshop on Camera-Based Document Analysis and Recognition, Curitiba, Brazil, 2007, pp. 71-78.
- [24] N.Stamatopoulos, B. Gatos, and I. Pratikakis, “Goal-oriented rectification of camera-based document images”, IEEE Trans On Image Processing, Vol. 20, No. 4, April 2011.

