Chapter 2

Theoretical Background

Introduction

2.1 Image

2.2 Digital image

2.3 RGB image

2.4 Gray Image

2.5 Image Acquisition

2.6 Image Preprocessing

2.7 Image Enhancement

2.7.1 Enhancement by point processing

2.7.2 Spatial Filtering:

2.7.3 Enhancement in the frequency domain:

2.7.4 Pseudo color image processing

2.8 Geometric Distortion

2.8.1 Translation

2.8.2 Rotation

2.9 Restoration

Conclusion

Introduction

This thesis Optical Character Recognition of handwritten Bangla Character focuses on quality improvement and efficient classification technique. The concept of quality improvement falls under the category of images enhancement and image restoration. These images are registered before enhancement following which the restoration carried out. The text documents considered in this thesis are handwriting documents. This chapter provides the theoretical background about the images, digital devices used to capture the digital images, followed by image enhancement and some of its techniques and image restoration in brief. Geometrical distortions are considered in OCR. It is also briefly discussed in this section.

2.1 Image

An image [9] is a binary representation of visual information such as drawings, pictures, graphs, logos, or individual video frames. It is a two-dimensional signal that is perceived by human visual system. Mathematically an image is a two dimensional function $f(x, y)$, where x and y are the spatial coordinates, and the value or amplitude of the function at coordinate (x, y) is the intensity of the image at that point. The intensity of an image perceived by human is infinite level and always positive.

2.2 Digital Image

Digitization is the process of transforming images, text, or sound from analog media (generally formats or objects that we can see or hear) into electronic data that we can save, organize, retrieve, and restore through electronic devices into perceptible surrogates of the original works. Of the vast number of digital assets that are being created, still images, texts, motion pictures, and sound recordings predominate. A digital image, then, is one that has been created through the process of digitization.

A digital image is a numeric representation of (normally binary) a two dimensional image. Depending on whether the image resolution is fixed, it may be of vector or raster type. By itself, the term "digital image" usually refers to raster images or bitmapped images. A digital image consists of discrete picture elements. An analog image is continuous with respect to coordinates and in amplitude. To convert an analog image in digital image, it is needed to digitize both coordinates and amplitude. Digitization of coordinates is referred to as sampling and digitization of the amplitude referred to as quantization. In digital image the 2D continuous image is divided into N rows and M columns. The intersection of row and column are known as pixel. The value assigned to integer coordinates $[m,n]$ with $m = \{0,1,2 \ldots \ldots \ldots \ldots .., M-1\}$ and $n = \{0,1,2 \ldots \ldots \ldots \ldots .., N-1\}$ is $f[m,n]$

Suppose, that a continuous image $f(x,y)$ is approximated by equally spaced and arranged in the form of an M x N array as shown follows:

$$
\begin{array}{l}
f(0,0), f(0,1) \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots f(0, N-1) \\
f(x,y) = f(1,0), f(0,1) \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots f(1, N-1) \\
\vdots \\
\vdots \\
F(M-1,0), F(M-1,1) \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots F(M-1, N-1)
\end{array}
$$

Here $f(x,y)$ represents the digital image and each array elements is referred to as an image element, picture element or a pixel. In this thesis, we will use the term pixel to refer to the samples of the digital image.

2.3 RGB Image

An RGB (red, green, blue) image is a three-dimensional byte array that explicitly stores a color value for each pixel. RGB image arrays are made up of width, height, and three channels of color information. Scanned photographs are commonly stored as RGB images. The color information is stored in three sections of a third dimension of the image. These sections are known as color channels, color bands, or color layers. One channel represents the amount of red in the image (the red channel), one channel represents the amount of green in the image (the green channel), and one channel represents the amount of blue in the image (the blue channel). An RGB image is stack of three matrices; representing the red, green and blue values for each pixel. These three images forming the RGB color image are referred to as red, green and blue components images. If each of these Components has a range 0-255, this gives a total of 2563 different colors.

2.4 Gray Image

A grayscale (or gray level) image is simply one in which the only colors are shades of gray. The reason for differentiating such images from any other sort of color image is that less information needs to be provided for each pixel. In fact a `gray' color is one in which the red, green and blue components all have equal intensity in RGB space, and so it is only necessary to specify a single intensity value for each pixel, as opposed to the three intensities needed to specify each pixel in a full color image.

Often, the grayscale intensity is stored as an 8-bit integer giving 256 possible different shades of gray from black to white. If the levels are evenly spaced then the difference between successive gray levels is significantly better than the gray level resolving power of the human eye.

Grayscale images are very common, in part because much of today's display and image capture hardware can only support 8-bit images. In addition, grayscale images are entirely sufficient for many tasks and so there is no need to use more complicated and harder-to-process color images.

Table 2.1 Ranges of gray scale

Gray Scale	No. of levels	Ranges
2^1	2	0-1
2^2	4	0-3
2^4	16	0-15
2^8	256	0-255

The lowest value of the range represents black pixel intensity and highest value represents the white pixel intensity.

2.5 Image Acquisition

Before any video or image processing can commence an image must be captured by a camera and converted into a manageable entity. This is the process known as image acquisition. An imaging sensor is required to acquire a digital image. These sensors are sensitive to certain bands in the electromagnetic energy spectrum such as X-ray, ultraviolet, infrared or visible bands, and produce an electrical signal output proportional to the energy. The imaging sensor can be digital or analog. Analog sensor generates a digitized signal as an output that can be digitized using an Analog to

Digital (A/D) converter. In this thesis a digital camera is used to scan the document and obtain the image in its digitized form, i.e., a signal matrix.

2.6 Image Preprocessing

The term image processing denotes the manipulation and analysis of image. This can be divided in the following types:

a. Digitization and compression: This technique is used for converting an image to a discrete signal for computer processing and compressing it to economize on storage capacity or communication bandwidth for transmitting purpose.
b. Enhancement, restoration and reconstruction: The images signals are sometimes degraded by noise, low contrast or blurring. To obtain the original image or to improve it for analysis purposes, image enhancement, restoration or reconstruction techniques are used depending on the objective.
c. Matching and recognition: When images taken by different sensors at different times are to be compared, we use the matching or registration techniques. This analysis includes segmentation of an image, measuring the properties of different parts an obtaining a relationship between the parts and comparing. The resulting descriptions are examined using certain models. The ultimate goal of the above techniques is to help an observer translate the contents of an image into useful information.

2.7 Image Enhancement

After a digital image is obtained, it needs to be processed to improve its quality and make it suitable for a specific application. This technique is called image enhancement. The principal objective of image enhancement is to process a given image so that the result is more suitable than the original image for a specific application. It accentuates or sharpens image features such as edges, boundaries, or contrast to make a graphic display more helpful for display and analysis. The enhancement doesn't increase the inherent information content of the data, but it increases the dynamic range of the chosen features so that they can be detected easily.

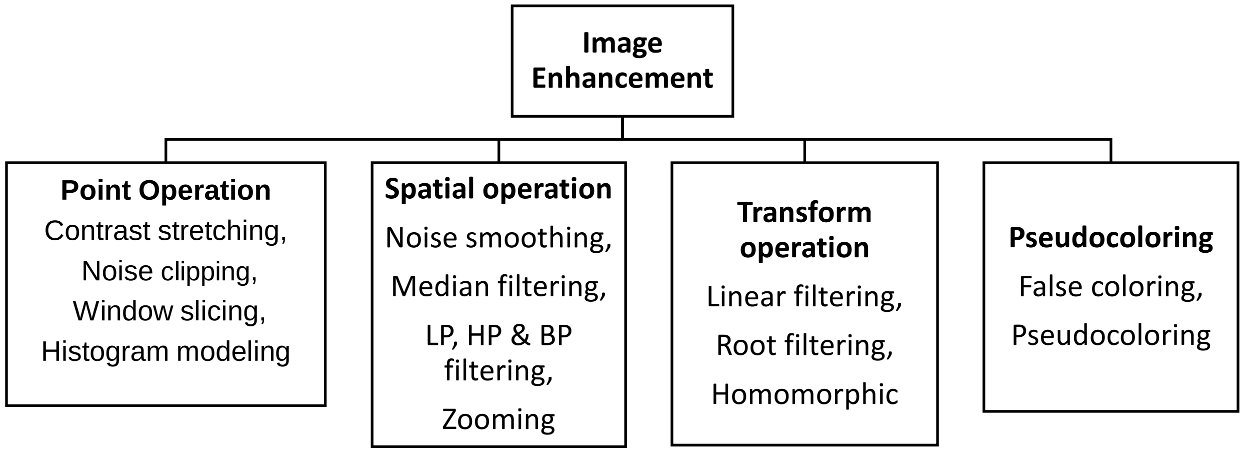

Fig. 2.1: Image Enhancement Operation.

The greatest difficulty in image enhancement is quantifying the criterion for enhancement and, therefore, a large number of image enhancement techniques are empirical and require interactive procedures to obtain satisfactory results. The image enhancement operations generally fall into two categories:

- Spatial Domain Techniques
- Frequency Domain Techniques

Spatial domain techniques are performed to the image plane itself and they are based on direct manipulation of pixels in an image. The operation can be formulated as $g(x, y) = \boldsymbol{T}[f(x, y)]$, where g is the output, f is the input image and $\boldsymbol{T}$ is an operation on f defined over some neighborhood. According to the operations on the image pixels, it can be further divided into 2 categories:

- Point operations and
- spatial operations(including linear and non-linear operations)

The 2D convolution is performed in frequency domain with DFT.

Spatial domain: $g(x, y) = f(x, y) * h(x, y)$

Frequency domain: $G(w1, w2) = F(w1, w2)H(w1, w2)$

2.7.1 Enhancement by point processing

• These processing methods are based only on the intensity of single pixels.

- ➢ **Simple intensity transformation:**
- ✓ **Image negatives:**

- Negatives of digital images are useful in numerous applications, such as displaying medical images and photographing a screen with monochrome positive film with the idea of using the resulting negatives as normal slides.
- Transform function $T: g(x,y) = L - f(x,y)$, where L is the max. intensity.

✓ **Contrast stretching**

- Low-contrast images can result from poor illumination, lack of dynamic range in the image sensor, or even wrong setting of a lens aperture during image acquisition.
- The idea behind contrast stretching is to increase the dynamic range of the gray levels in the image being processed.

✓ **Compression of dynamic range**

- Sometimes the dynamic range of a processed image far exceeds the capability of the display device, in which case only the brightest parts of the images are visible on the display screen.
- An effective way to compress the dynamic range of pixel values is to perform the following intensity transformation function:

$$s = c\log(1 + |\mathrm{r}|) \tag{2.1}$$

Where, c is a scaling constant, and the logarithm function performs the desired compression

✓ **Gray-level slicing**

- Highlighting a specific range of gray levels in an image often is desired. Applications include enhancing features such as masses of water in satellite imagery and enhancing flaws in x-ray images

➢ **Histogram processing:**

- The histogram of a digital image with gray levels in the range [0,L-1] is a discrete function p(r_k)= (n_k)/n, where r_k= is the kth gray level, n_k is the number of pixels in the image with that gray level, n is the total number of pixels in the image, and k=0,1..L-1.

- $P(r_k)$ gives an estimate of the probability of occurrence of gray level r_k.
- The shape of the histogram of an image gives us useful information about the possibility for contrast enhancement.
- A histogram of a narrow shape indicates little dynamic range and thus corresponds to an image having low contrast

- ✓ **Histogram equalization**
- The objective is to map an input image to an output image such that its histogram is uniform after the mapping.
- Let r represent the gray levels in the image to be enhanced and s is the enhanced output.
- Transformation of the form $s = T(r)$
- ✓ **Histogram specification**
- Histogram equalization only generates an approximation to a uniform histogram.
- Sometimes the ability to specify particular histogram shapes capable of highlighting certain gray-level ranges in an image is desirable
- ✓ **Local enhancement**
- It is often necessary to enhance details over small areas.
- The number of pixels in these areas may have negligible influence on the computation of a global transformation, so the use of global histogram specification does not necessarily guarantee the desired local enhancement

2.7.2 Spatial Filtering:

- The use of spatial masks for image processing is called spatial filtering.
- The masks used are called spatial filters.
- The basic approach is to sum products between the mask coefficients and the intensities of the pixels under the mask at a specific location in the image. (2D convolution)

$$R(x,y) = \sum_{i=-d}^{d} \sum_{-d}^{d} w(i,j)\, f(x-i, y-j) \tag{2.2}$$

Where, $(2d+1)X(2d+1)$ is the mask size, $w(i,j)$'s are weights of the mask, $f(x,y)$is input pixel at coordinates (x,y), $R(x,y)$is the output value at (x,y).

- ➢ If the center of the mask is at location (x,y) in the image, the gray level of the pixel located at (x,y) is replaced by R, the mask is then moved to the next location in the image and the process is repeated. This continues until all pixel locations have been covered.
- ➢ **Smoothing filter:**
- Smoothing filters are used for blurring and for noise reduction.
- Blurring is used in preprocessing steps, such as removal of small details from an image prior to object extraction, and bridging of small gaps in lines or curves.

- Noise reduction can be accomplishing by blurring with a linear filter and also by nonlinear filtering.

✓ **Low pass filtering**

- The key requirement is that all coefficients are positive.
- Neighborhood averaging is a special case of LPF where all coefficients are equal.
- It blurs edges and other sharp details in the image

Example: $\frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

✓ **Median filtering**

- If the objective is to achieve noise reduction instead of blurring, this method should be used.
- This method is particularly effective when the noise pattern consists of strong, spike-like components and the characteristic to be preserved is edge sharpness.
- It is a nonlinear operation.
- For each input pixel $f(x, y)$, we sort the values of the pixel and its neighbors to determine their median and assign its value to output pixel $g(x, y)$

➢ **Sharpening Filters**

- To highlight fine detail in an image or to enhance detail that has been blurred, either in error or as a natural effect of a particular method of image acquisition.
- Uses of image sharpening vary and include applications ranging from electronic printing and medical imaging to industrial inspection and autonomous target detection in smart weapons.

✓ **Basic high-pass spatial filter**

- The shape of the impulse response needed to implement a high-pass spatial filter indicates that the filter should have positive coefficients near its center, and negative coefficients in the outer periphery.

  Example: filter mask of a 3x3 sharpening filter

$$\frac{1}{9}\begin{bmatrix}-1-1-1\\-1 \quad 8-1\\-1-1-1\end{bmatrix}$$

- The filtering output pixels might be of a gray level exceeding $[0, L-1]$.
- The results of high-pass filtering involve some form of scaling and/or clipping to make sure that the gray levels of the final results are within $[0, L-1]$

✓ **Derivative filters.**

- Differentiation can be expected to have the opposite effect of averaging, which tends to blur detail in an image, and thus sharpen an image and be able to detect edges.
- The most common method of differentiation in image processing applications is the gradient.
- For a function $f(x, y)$, the gradient of fat coordinates (x', y') is defined as the vector

$$\nabla \mathrm{f}(\mathrm{x}', \mathrm{y}') = \begin{bmatrix}\frac{\partial f}{\partial x}\\\frac{\partial f}{\partial y}\end{bmatrix}$$

- Its magnitude can be approximated in a number of ways, which result in a number of operators such as Roberts, Prewitt and Sobel operators for computing its value

2.7.3 Enhancement in the frequency domain:

We simply compute the Fourier transform of the image to be enhanced, multiply the result by a filter transfer function, and take the inverse transform to produce the enhanced image.

Spatial domain: $g(x, y) = f(x, y) * h(x, y)$ (2.3)

Frequency domain: $G(w1, w2 = F(w1, w2)H(w1, w2)$ (2.4)

➢ **Low-pass filtering**

Edges and sharp transitions in the gray levels contribute to the high frequency content of its Fourier transform, so a Low-pass filter smooths an image. Formula of ideal LPF function

$$H(u, v) = \begin{cases}1 & if\ D(u, v) \le D_0\\0 & else\end{cases} \quad (2.5)$$

➢ **High-pass filtering**

A high-pass filter attenuates the low frequency components without disturbing the high frequency information in the Fourier transform domain can sharpen edges. Formula of ideal HPF function

$$H(u,v) = \begin{cases} 0 & if\ D(u,v) \le D_0 \\ 1 & else \end{cases} \quad (2.6)$$

2.7.4 Pseudo color image processing

- In automated image analysis, color is a powerful descriptor that often simplifies object identification and extraction from a scene.
- Human eye performs much better in discerning shades of color than gray scale.
- A monochrome image can be enhanced by using colors to represent different gray levels or frequencies

2.8 Geometric Distortion

When a document image is obtained from scanner they have some distortion and misalignment between the recto and verso images. There are mainly three types of distortions- translation, rotation and scaling. Here all transformations are described in 2-dimentional Cartesian co-ordinates. Let the initial point (x_1, y_1) in image f1 be transformed in the image f2 at point (x_2, y_2). The geometric transformation between two images can be described between two images by two equations $x_2 = h_1(x_1, y_1)$ and $y_2 = h_2(x_1, y_1)$ where $h1$ and $h2$ are transformation functions.

2.8.1 Translation

The translate operator performs a geometric transformation which maps the position of each picture element in an input image into a new position in an output image, where the dimensionality of the two images often is but need not necessarily be, the same [10]. Under translation, an image element located at (x_1, y_1) in the original is shifted to a new position (x_2, y_2) in the corresponding output image by displacing it through a user-specified translation $\left(B_x, B_y\right)$. The treatment of elements near image edges varies with implementation. Translation is used to improve visualization of an image, but also has a role as a preprocessor in applications where registration of two or more images is required. Translation is a special case of one transformation. The translation operator performs a transformation of the form:

$$X_2 = x_1 + B_x \quad (2.7)$$

$$Y_2 = y_1 + B_y \quad (2.8)$$

2.8.2 Rotation

The rotation [f3] operator performs a geometric transform which maps the position $(x, y)$of a picture element in an input image onto a position (x_2, y_2) in an output image by rotating it through a user-specified angle about an origin. In most implementations, output location 0 about an origin 0. In most implementations, output locations (x_2, y_2)which are outside the boundary of the image are ignored. Rotation is most commonly used to improve the visual appearance of an image, although it can be useful as a preprocessor in applications where directional operators are involved. Rotation is a special case of affine transformation. The rotation operator performs a transformation of the form:

$$x_2 = \cos\theta * (x_1 - x_0) - \sin\theta * (y_1 - y_0) + x_0 \tag{2.9}$$

$$y_2 = \sin\theta * (x_1 - x_0) - \cos\theta * (y_1 - y_0) + y_0 \tag{2.10}$$

2.9 Restoration

The operation of taking a corrupt/noisy image and estimating the clean, original image is known as restoration. Corruption may come in many forms such as motion, blur, noise and camera miss-focus. Image restoration is performed by reversing the process that blurred the image and such is performed by imaging a point source and use the point source image, which is called the Point Spread Function (PSF) to restore the image information lost to the blurring process. Image restoration is different from image enhancement in that the latter is designed to emphasize features of the image that make the image more pleasing to the observer, but not necessarily to produce realistic data from a scientific point of view. Image enhancement techniques (like contrast stretching or de-blurring by a nearest neighbor procedure) provided by imaging packages use no a priori model of the process that created the image. With image enhancement noise can effectively be removed by sacrificing some resolution, but this is not acceptable in many applications. In a fluorescence microscope, resolution in the z-direction is bad as it is more advanced image processing techniques must be applied to recover the object. The objective of image restoration techniques is to reduce noise and recover resolution loss. Image processing techniques are performed either in the image domain or the frequency domain. The most straightforward technique for image restoration is Deconvolution, which is performed in the frequency domain and after computing the Fourier Transformation of both the image and the PSF and undo the resolution loss caused by the blurring factors. Deconvolution technique assumes absence of noise and that the blurring process is shift-invariant and hence more sophisticated techniques have been developed to deal with the different types of noises and blurring functions.

Conclusion

The success of any system totally depends on the preprocessing phase and it includes Image processing. To enhance the performance of the system the issues related to image processing are taken care of. Actually, a stable OCR system cannot be imagined without the proper uses of these operations.