1. Fundamentals of Digital Image

Digital image is a representation of a 2 dimensional intensity function $f(x, y)$, where (x, y) refers to the spatial location, and the value of intensity or brightness is represented using a finite set of digital values, known as gray values. The digital image is a discrete in both- the spatial dimensions (x and y coordinates) and the brightness (gray values). One can think of a digital image as a 2-dimensional matrix whose rows and columns indices refer to a location of the point, and the value of the corresponding matrix element is equal to the gray value at that point.

The elements of this digital matrix are known as picture elements, and commonly abbreviated as *"pixels"*. The total number of digital values used to represent the gray values indicate the range of image. This number is usually chosen to be a power of 2 for efficient storage. Thus, if an image has L gray values, equally spaced between $[0, L-1]$, such that $L = 2^m$, then the pixel value at each location can be represented using m bits.

For a given pixel $f(x, y)$, there exist a set of 4 horizontal and vertical neighboring pixels. These are identified as:

$$f(x+1, y), f(x-1, y), f(x, y+1), f(x, y-1).$$

Similarly one can find out a set of 4 diagonal neighbors for a pixel $f(x, y)$ as:

$$f(x+1, y+1), f(x+1, y-1), f(x-1, y+1), f(x-1, y-1).$$

The set of horizontal and vertical neighbors for a pixel p, is denoted as $N_4(p)$, and the set of its diagonal neighbors is denoted as $N_D(p)$. Both these sets together constitute the 8-neighbors of pixel p, denoted as $N_8(p)$. When the pixels are on the image boarder, some of the neighbors of the pixels lie outside the image.

2. Image Scaling, Translation and Rotation

Image scaling, translation, and rotation are some of the primary operations in the manipulation of an image.

2.1. Image Scaling

The scaling is resizing of an image with minimal error in smoothness or sharpness of the image. The scaling of an image can result either in reduction or in enlargement in the size. If one can think of an image as a 2D array of pixels, then first procedure which decreases the size of the image includes downsampling of the pixels along X and Y directions by the given factor. The later procedure of enlargement is commonly known as "zooming", and it includes upsampling of the image pixels in both the directions. During the size reduction, one performs downsampling of the image, which means removal of some of the information contents of the image. The zooming operation, however, involves upsampling followed by interpolation. Since it is not possible to obtain any extra information from the given image, the quality of the enlarged image is usually lower than that of the original image. Several techniques for efficient interpolations are used to minimize the loss of image quality.

Consider an image of size $M \times N$, to be enlarged by the factor of 2. The new image will have $2M \times 2N$ pixels, that is, four pixels will appear in the place of every single pixel of original image. The simplest technique for assigning the values of the "new" pixels in the enlarged image is replication, where the the 4 new pixels are assigned the value of the original single pixel. Although this method is very simple to implement, the resultant images have blocky-ness. The edges and boundaries of the image also get adversely affected. A bilinear interpolation is sometimes used to produce better quality enlarged images. To estimate the value of a given pixel, the bilinear interpolation uses 4 pixels in the neighborhood of that particular pixel. The value of the pixel is calculated as the linear combination of the values of these neighboring pixels (and sometimes a constant). Since this technique uses only 4 pixels and linear operations, it is faster in implementation and also better at preserving smoothness in the image. Another technique of interpolation uses polynomials of the degree 3, and hence known as bicubic interpolation. The polynomial functions such as cubic splines or Lagrange polynomials are usually used. The resultant images have a better quality and smoothness as compared to the results of the first two algorithms, however the bicubic interpolation demands more computation, and hence they are slower.

2.2. Image Translation

The translation of an image is shifting (or translating) each pixel in the image by a specific displacement (x_0, y_0). If we represent the coordinates of each pixel in the image as a 2-dimensional vector $\mathbf{p} = [x, y]$, and the displacement as a vector $\mathbf{v} = [x_0, y_0]$, then the new location for the pixel, $\mathbf{p}^*$ can be written as:

$$\mathbf{p}^* = \mathbf{p} + \mathbf{v}$$

2.3. Image Rotation

The rotation of an image refers to the rotation of each pixel in the image by a desired degree in the xy-plane. If an image is to be rotated by an angle θ, then the corresponding transformation matrix $\mathbf{R}$ can be written as:

$$\mathbf{R}_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

This rotation matrix is an orthonormal matrix, with determinant equal to 1. The rotation operation for every pixel at location $\mathbf{p}$ is then carried out as:

$$\mathbf{p}^* = \mathbf{R}_\theta \mathbf{p}$$

In the case of the rotation of a digital image, sometimes the location of the pixels may not exactly coinside with any of the locations at the given resolution. Therefore, some interpolation algorithm is applied to generate a final rotated image. The rotation of image slightly increases the dimensions of the rectangular boarder that contains the image. It is a common practice to assign 0 values to the pixels which are not a part of the image.

Sometimes, the translation and rotation operations are combined together which is known as affine transformation. This combined operation can be represented as:

$$\mathbf{p}^* = \mathbf{R}_\theta \mathbf{p} + \mathbf{v}$$

—

3. Study of statistical properties – Mean, Standard Deviation, Variance, Profile, and Histogram plotting

In this experiment, we will study some of the commonly useful properties of a digital image. Consider an image f of size $M \times N$. The image contains MN pixels, which can be represented as $f(x, y)$, where x and y indicate the number of column and the number of row, respectively.

Mean = The mean of an image is the average value of the gray values of all the pixels in the image. As the gray value of a pixel represents the amount of intensity at the given location, one can think of the mean value as an average intensity of the image. The mean, μ_{f} of the image is calculated as:

$$\mu_{\mathrm{f}} = \frac{1}{M\,N} \sum_{x=1}^{M} \sum_{y=1}^{N} f(x, y).$$

Thus, the mean is calculated by adding the gray values of all the pixels, and then dividing the sum by image dimensions. A low value of mean indicates less amount of intensity, and therefore the images appear darker. The images with a higher value of mean appear bright.

Variance = The mean of an image gives an indication of average image intensity, but does not give any idea regarding the spread of the gray values around the mean. The variance of an image is a measure of spread of the gray values around its mean. The variance is calculated as follows: For every pixel, the difference between the gray value of the pixel and the image mean is calculated. This difference value is then squared to give a squared difference. Due to this step, the positive and negative differences are treated in the same manner. Finally the sum of the squared differences for all the pixels is divided by the image dimensions to yield the variance σ^2 of the image. We can write the mathematical expression as:

$$\sigma^2 = \frac{1}{M\,N} \sum_{x=1}^{M} \sum_{y=1}^{N} (f(x, y) - \mu_{\mathrm{f}})^2.$$

Sometimes, the positive square root of the variance is used as a statistical measure. It is known as standard deviation (σ). The variance or standard deviation gives some indication of the image contrast. Images with lower contrast have large number of pixels close to the mean value. Such images have small value of the variance. For images having higher contrast, the spread of the pixel gray values is large, and hence a high variance.

Profile = For each of the N rows of the image, calculate the sum of all the pixels in the given row. As there are N rows, there will be N such values, each containing a summation of particular row of M pixels. Let's call this array of N elements as h_p. The ith element of h_p represents the sum of the pixels in the ith row of the image. Since the array h_p indicates the sum of the gray values of the image along the horizontal direction, this array is called as the horizontal profile of the image. We can write the expression for h_p as:

$$h_{\mathrm{p}}(i) = \sum_{x=1}^{M} f(x,i) \qquad\qquad i = 1, 2, \cdots, N.$$

Using an exactly similar analogy, we can calculate an array of M elements to represent the sum of the gray values of the image along its vertical direction. Here, each element of this array is equal to the sum of the pixels along the corresponding column. This array is called as the vertical profile of the image v_p, and it can mathematically written as:

$$v_{\mathrm{p}}(i) = \sum_{y=1}^{N} f(i,y) \qquad\qquad i = 1, 2, \cdots, M.$$

The profiles of image are particularly useful in the processing of documents, or processing of handwritten scripts for the extraction of individual lines and characters in the text. Sometimes, plots of the horizontal and vertical profiles are used to get quick idea of the distribution of gray values in the corresponding direction.