Aula-2: Visão da disciplina (Gonzalez-cap.1 and 2) – v16 (5/agosto)

 # Digital Image Processing, Gonzalez / Woods, 3. edição, 2008, caps. 1 e 2.
Consultar 4. edição, 2018.

CAPÍTULO-1 Introduction (Gonzalez)

1. What Is Digital Image Processing ?

An image may be defined as a two-dimensional function, f(x, y) , where x and y are spatial (plane) coordinates, and the amplitude of f at any pair of coordinates (x, y) is called the intensity or gray level of the image at that point.

When x, y, and the intensity values of f are all finite, discrete quantities, we call the image a digital image.

The field of digital image processing refers to processing digital images by means of a digital computer.

This way, a digital image is composed of a finite number of elements, each of which has a particular location and value. These elements are called picture elements, image elements, pels, and pixels. Pixel is the term used most widely to denote the elements of a digital image.

Related areas: (lembrar Horn, aula 1, Figuras 1.2, 1.3 e 1.10)

There is no general agreement among authors regarding where **image processing** stops and other related areas, such as **image analysis** and **computer vision**, start.

Sometimes a distinction is made by defining **image processing** as a discipline in which both the input and output of a process are images. We believe this to be a limiting and somewhat artificial boundary.

For example, under this definition, even the trivial task of computing the average intensity of an image (which yields a single number) would not be considered an image processing operation.

On the other hand, there are fields such as **computer vision** whose ultimate goal is to use computers to emulate human vision, including learning and being able to make inferences and take actions based on visual inputs.

The area of **image analysis** (also called image understanding) is in between image processing and computer vision.

There are no clear-cut boundaries in the continuum from image processing at one end to computer vision at the other.

However, one useful paradigm is to consider three types of computerized processes in this continuum: low-, mid-, and high-level processes.

Low-level processes involve primitive operations such as image preprocessing to reduce noise, contrast enhancement, and image sharpening. A low-level process is characterized by the fact that both its inputs and outputs are images.

Mid-level processing on images involves tasks such as segmentation (partitioning an image into regions or objects), description of those objects to reduce them to a form suitable for computer processing, and classification (recognition) of individual objects. A mid-level process is characterized by the fact that its inputs generally are images, but its outputs are attributes extracted from those images (e.g., edges, contours, and the identity of individual objects).

Finally, higher-level processing involves "making sense" of an ensemble of recognized objects, as in image analysis, and, at the far end of the continuum, performing the cognitive functions normally associated with vision.

2. The Origins of Digital Image Processing

The first computers powerful enough to carry out meaningful image processing tasks appeared in the early 1960s.

Work on using computer techniques for improving images from a space probe began at the Jet Propulsion Laboratoryin 1964 when pictures of the moon transmitted by Ranger 7 were processed by a computer to correct various types of image distortion inherent in the on-board television camera.

From the 1960s until the present, the field of image processing has grown vigorously. In addition to applications in medicine and the space program, digital image processing techniques now are used in a broad range of applications. Computer procedures are used to

- enhance the contrast or code the intensity levels into color for easier interpretation of X-rays and other images used in industry, medicine, and the biological sciences;
- Geographers use the same or similar techniques to study pollution patterns from aerial and satellite imagery;
- Image enhancement and restoration procedures are used to process degraded images of unrecoverable objects or experimental results too expensive to duplicate;
- In archeology, image processing methods have successfully restored blurred pictures that were the only available records of rare artifacts lost or damaged after being photographed;
- In physics and related fields, computer techniques routinely enhance images of experiments in areas such as high-energy plasmas and electron microscopy;
- Similarly successful applications of image processing concepts can be found in astronomy, biology, nuclear medicine, law enforcement, defense, and industry.

(These examples illustrate processing results intended for human interpretation.)

The second major area of application of digital image processing techniques mentioned at the beginning of this chapter is in solving problems dealing with machine perception. In this case, interest is on procedures for extracting from an image information in a form suitable for computer processing.

Often, this information bears little resemblance to visual features that humans use in interpreting the content of an image.

Examples of the type of information used in machine perception are statistical moments, Fourier transform coefficients, and multidimensional distance measures. Typical problems in machine perception that routinely utilize image processing techniques are automatic character recognition, industrial machine vision for product assembly and inspection, military recognizance, automatic processing of fingerprints, screening of X-rays and blood samples, and machine processing of aerial and satellite imagery for weather prediction and environmental assessment.

The continuing decline in the ratio of computer price to performance and the expansion of networking and communication bandwidth via the World Wide Web and the Internet have created unprecedented opportunities for continued growth of digital image processing. Some of these application areas are illustrated in the following section.

CAPÍTULO-2 Digital Image Fundamentals (Gonzalez)

2.1. Elements of Visual Perception

Although the field of digital image processing is built on a foundation of mathematical and probabilistic formulations, human intuition and analysis play a central role in the choice of one technique versus another, and this choice often is made based on subjective, visual judgments.

In particular, our interest is in the mechanics and parameters related to how images are formed and perceived by humans. We are interested in learning the physical limitations of human vision in terms of factors that also are used in our work with digital images. Thus, factors such as how human and electronic imaging devices compare in terms of resolution and ability to adapt to changes in illumination are not only interesting, they also are important from a practical point of view.

2.1.1 Structure of the Human Eye

Figure-2.1 (cross section of human eye) + Fig. 2.2

At its anterior extreme, the choroid is divided into the ciliary body and the iris. The latter contracts or expands to control the amount of light that enters the eye. The central opening of the iris (the pupil) varies in diameter from approximately 2 to 8 mm. The front of the iris contains the visible pigment of the eye, whereas the back contains a black pigment.

The lens is made up of concentric layers of fibrous cells and is suspended by fibers that attach to the ciliary body. It contains 60 to 70% water, about 6% fat, and more protein than any other tissue in the eye. The lens is colored by a slightly yellow pigmentation that increases with age. In extreme cases, excessive clouding of the lens, caused by the affliction commonly referred to as cataracts, can lead to poor color discrimination and loss of clear vision.

The lens absorbs approximately 8% of the visible light spectrum, with relatively higher absorption at shorter wavelengths. Both infrared and ultraviolet light are absorbed appreciably by proteins within the lens structure and, in excessive amounts, can damage the eye.

The innermost membrane of the eye is the retina, which lines the inside of the wall's entire posterior portion. When the eye is properly focused, light from an object outside the eye is imaged on the retina. Pattern vision is afforded by the distribution of discrete light receptors over the surface of the retina.

There are two classes of receptors: cones and rods.

The **cones** (cones) in each eye number between 6 and 7 million. They are located primarily in the central portion of the retina, called the fovea, and are highly sensitive to color. Humans can resolve fine details with these cones largely because each one is connected to its own nerve end. Muscles controlling the eye rotate the eyeball until the image of an object of interest falls on the fovea. Cone vision is called photopic or bright-light vision.

The number of **rods** (bastonestes) is much larger: Some 75 to 150 million are distributed over the retinal surface. The larger area of distribution and the fact that several rods are connected to a single nerve end reduce the amount of detail discernible by these receptors. Rods serve to give a general, overall picture of the field of view. They are not involved in color vision and are sensitive to low levels of illumination. For example, objects that appear brightly colored in day-light when seen by moonlight appear as colorless forms because only the rods are stimulated. This phenomenon is known as scotopic or dim-light vision.

Figure 2.2 shows the density of rods and cones for a cross section of the right eye passing through the region of emergence of the optic nerve from the eye. The absence of receptors in this area results in the so-called blind spot (see Fig. 2.1). Except for this region, the distribution of receptors is radially symmetric about the fovea.

The fovea itself is a circular indentation in the retina of about 1.5 mm in diameter. However, in terms of future discussions, talking about square or rectangular arrays of sensing elements is more useful. Thus, by taking some liberty in interpretation, we can view the fovea as a square sensor array of size 1.5 mm * 1.5 mm. **The density of cones in that area of the retina is approximately 150,000 elements per mm^2.**

Based on these approximations, the number of cones in the region of highest acuity in the eye is about 337,000 elements. Just in terms of raw resolving power, a charge-coupled device (CCD) imaging chip of medium resolution can have this number of elements in a receptor array no larger than 5 mm * 5 mm. While the ability of humans to integrate intelligence and experience with vision makes these types of number comparisons somewhat superficial, keep in mind for future discussions that the basic ability of the eye to resolve detail certainly is comparable to current electronic imaging sensors.

2.1.2 Image Formation in the Eye

- for photographic camera, the lens has a fixed focal length, and focusing at various distances is achieved by varying the distance between the lens and the imaging plane, where the film (or imaging chip in the case of a digital camera) is located.

- for human eye, the converse is true; the distance between the lens and the imaging region (the retina) is fixed, and the focal length needed to achieve proper focus is obtained by varying the shape of the lens. The fibers in the ciliary body accomplish this, flattening or thickening the lens for distant or near objects, respectively. The distance between the center of the lens and the retina along the visual axis is approximately 17 mm. The range of focal lengths is approximately 14 mm to 17 mm, the latter taking place when the eye is relaxed and focused at distances greater than about 3 m.

The geometry in Fig. 2.3 illustrates how to obtain the dimensions of an image formed on the retina. For example, suppose that a person is looking at a tree 15 m high at a distance of 100 m. Letting h denote the height of that object in the retinal image, the geometry of Fig. 2.3 yields 15/100 = h/17 or h = 2.55 mm. As indicated in Section 2.1.1, the retinal image is focused primarily on the region of the fovea. Perception then takes place by the relative excitation of light receptors, which transform radiant energy into electrical impulses that ultimately are decoded by the brain.

2.3 Image Sensing and Acquisition

2.3.4 A Simple Image Formation Model

As introduced in Section 1.1, we denote images by two-dimensional functions of the form f(x, y).

The value or amplitude of "f" at spatial coordinates (x, y) is a positive scalar quantity whose physical meaning is determined by the source of the image. When an image is generated from a physical process, its intensity values are proportional to energy radiated by a physical source (e.g., electromagnetic waves). As a consequence, f(x, y) must be nonzero and finite; that is,

$$0 < f(x, y) < \infty$$

The function f(x, y) may be characterized by two components:

(1) the amount of source illumination incident on the scene being viewed, and
(2) the amount of illumination reflected by the objects in the scene.

Appropriately, these are called the illumination and reflectance components and are denoted by i(x, y) and r (x, y), respectively.The two functions combine as a product to form f(x, y) :

$$f(x, y) = i(x, y)\, r(x, y), \quad \text{(Eq. 2.3-2)}$$

$$0 < i(x, y) < \infty \ \text{(Eq. 2.3-3)} \quad \text{and} \quad 0 < r(x, y) < 1 \ \text{(Eq. 2.3-4)}$$

Equation (2.3-4) indicates that reflectance is bounded by 0 (total absorption) and 1 (total reflectance). The nature of i (x, y) is determined by the illumination source, and r (x, y) is determined by the characteristics of the imaged objects.

2.4 Image Sampling and Quantization

There are numerous ways to acquire images, but our objective in all is the same:

to generate digital images from sensed data.

The output of most sensors is a continuous voltage waveform whose amplitude and spatial behavior are related to the physical phenomenon being sensed. To create a digital image, we need to convert the continuous sensed data into digital form.

This involves two processes: **sampling** and **quantization**.

2.4.1 Basic Concepts in Sampling and Quantization

The basic idea behind sampling and quantization is illustrated in Fig. 2.16.

Figure 2.16(a) shows a continuous image f that we want to convert to digital form. An image may be continuous with respect to the x- and y-coordinates, and also in amplitude. To convert it to digital form, we have to sample the function in both coordinates and in amplitude.

Digitizing the **coordinate values** is called **sampling**. Digitizing the **amplitude values** is called **quantization**.

The one-dimensional function in Fig. 2.16(b) is a plot of amplitude (intensity level) values of the continuous image along the line segment AB in Fig. 2.16(a). The random variations are due to image noise.

To sample this function, we take equally spaced samples along line AB, as shown in Fig. 2.16(c). The spatial location of each sample is indicated by a vertical tick mark in the bottom part of the figure. The samples are shown as small white squares superimposed on the function.The set of these discrete locations gives the sampled function.

However, the values of the samples still span (vertically) a continuous range of intensity val-
ues. In order to form a digital function, the intensity values also must be converted (quantized) into discrete quantities. The right side of Fig. 2.16(c) shows the intensity scale divided into eight discrete intervals, ranging from black to white. The vertical tick marks indicate the specific value assigned to each of the eight intensity intervals. The continuous intensity levels are quantized by assigning one of the eight values to each sample.The assignment is made depending on the vertical proximity of a sample to a vertical tick mark.

The digital samples resulting from both sampling and quantization are shown in Fig. 2.16(d). Start-
ing at the top of the image and carrying out this procedure line by line produces a two-dimensional digital image. It is implied in Fig. 2.16 that, in addition to the number of discrete levels used, the accuracy achieved in quantization is highly dependent on the noise content of the sampled signal.

Others possible sampling forms: sensing strip; single sensing element combined with mechanical motion; sensing array; etc.

2.4.2 Representing Digital Images

Let f(s, t) represent a continuous image function of two continuous variables, s and t. We convert this function into a digital image by sampling and quantization, as explained in the previous section.

Suppose that we sample the continuous image into a 2-D array, f(x, y) , containing M rows and N columns, where (x, y) are discrete coordinates.

> For notational clarity and convenience, we use integer values for these discrete coordinates: $x = 0, 1, 2, \dots, M - 1$ and $y = 0, 1, 2, \dots, N - 1$.

Thus, for example, the value of the digital image at the origin is f (0, 0) , and the next coordinate value along the first row is f (0, 1) . Here, the notation (0, 1) is used to signify the second sample along the first row.

The section of the real plane spanned by the coordinates of an image is called the spatial domain, with x and y being referred to as spatial variables or spatial coordinates.

Representation forms:

> Figure 2.18(a) is a plot of the function, with two axes determining spatial location and the third axis being the values of f (intensities) as a function of the two spatial variables x and y. This representation is useful when working with gray-scale sets whose elements are expressed as triplets of the form (x, y, z) , where x and y are spatial coordinates and z is the value of f at coordinates (x, y) .

The representation in Fig. 2.18(b) is much more common. It shows f(x, y) as it would appear on a monitor or photograph. Here, the intensity of each point is proportional to the value of f at that point. In this figure, there are only three equally spaced intensity values. If the intensity is normalized to the interval [0, 1], then each point in the image has the value 0, 0.5, or 1. A monitor or printer simply converts these three values to black, gray, or white, respectively, as Fig. 2.18(b) shows.

The third representation is simply to display the numerical values of f(x, y) as an array (matrix). In this example, f is of size 600 * 600 elements, or 360,000 numbers. Clearly, printing the complete array would be cumbersome and convey little information. When developing algorithms, however, this representation is quite useful when only parts of the image are printed and analyzed as numerical values. Figure 2.18(c) conveys this concept graphically.

We conclude from the previous paragraph that the representations in Figs. 2.18(b) and (c) are the most useful. Image displays allow us to view results at a glance. Numerical arrays are used for processing and algorithm development. In equation form, we write the representation of an M * N numerical array as

$$f(x,y) = \begin{vmatrix} f(0,0) & f(0,1) & f(0,N-1) \\ & & \\ f(M-1,0) & & f(M-1,N-1) \end{vmatrix} \qquad (2.4\text{-}1)$$

Both sides of this equation are equivalent ways of expressing a digital image quantitatively. The right side is a matrix of real numbers. Each element of this matrix is called an image element, picture element, **pixel**, or pel. The terms image and pixel are used throughout the book to denote a digital image and its elements.

Outra notação seria $a_{i,j}$ equivalentemente.

Returning briefly to Fig. 2.18, note that the **origin of a digital image is at the top left**, with the positive x-axis extending downward and the positive y-axis extending to the right. This is a conventional representation based on the fact that many image displays (e.g., TV monitors) sweep an image starting at the top left and moving to the right one row at a time.

More important is the fact that the first element of a matrix is by convention at the top left of the array, so choosing the origin of f(x, y) at that point makes sense mathematically. Keep in mind that this representation is the standard right-handed Cartesian coordinate system with which you are familiar. † We simply show the axes pointing downward and to the right, instead of to the right and up.

2.4.4 Image Interpolation

Interpolation is a basic tool used extensively in tasks such as zooming, shrinking, rotating, and geometric corrections. **Our principal objective in this section is to introduce interpolation and apply it to image resizing (shrinking and zooming), which are basically image resampling**

methods. Uses of interpolation in applications such as rotation and geometric corrections are discussed in Section 2.6.5. We also return to this topic in Chapter 4,where we discuss image resampling in more detail.

Fundamentally, interpolation is the process of using known data to estimate values at unknown locations.

Suppose that an image of size (500x500) pixels has to be enlarged 1.5 times to (750x750) pixels. A simple way to visualize zooming is to create an imaginary grid with the same pixel spacing as the original, and then shrink it so that it fits exactly over the original image. Obviously,the pixel spacing in the shrunken (750x750) grid will be less than the pixel spacing in the original image.

To perform intensity-level assignment for any point in the overlay,we look for its closest pixel in the original image and assign the intensity of that pixel to the new pixel in the (750x750) grid. When we are finished assigning intensities to all the points in the overlay grid, we expand it to the original specified size to obtain the zoomed image.

The method just discussed is called **nearest neighbor interpolation** because it assigns to each new location the intensity of its nearest neighbor in the original image (pixel neighborhoods are discussed formally in Section 2.5). This approach is simple but, as we show later in this section, it has the tendency to produce undesirable artifacts, such as severe distortion of straight edges. For this reason, it is used infrequently in practice.

A more suitable approach is **bilinear interpolation**, in which we use the **four nearest neighbors** to estimate the intensity at a given location. Let denote the coordinates of the location to which we want to assign an intensity value (think of it as a point of the grid described previously), and let denote that intensity value. For bilinear interpolation,the assigned value is obtained using the equation

$$v(x,y) = a.x + b.y + c.x.y + d \qquad (2.4\text{-}6)$$

where the four coefficients are determined from the four equations in four unknowns that can be written using the four nearest neighbors of point(x,y).

The next level of complexity is **bicubic interpolation**, which involves the sixteen nearest neighbors of a point. The intensity value assigned to point (x,y) is obtained using the equation

$$v(x,y) = \sum_{i=0,3} \sum_{j=0,3} a_{ij}\, x^i\, y^j \qquad (2.4\text{-}7)$$

where the sixteen coefficients are determined from the sixteen equations in sixteen unknowns that can be written using the sixteen nearest neighbors of point . Observe that Eq.(2.4-7) reduces in form to Eq.(2.4-6) if the limits of both summations in the former equation are 0 to 1.Generally, bicubic interpolation does a better job of preserving fine detail than its bilinear counterpart. Bicubic interpolation is the standard used in commercial image editing programs,such as Adobe Photoshop and Corel Photopaint.

2.5 Some Basic Relationships between Pixels

In this section,we consider several important relationships between pixels in a digital image. As mentioned before,an image is denoted by f(x,y).When referring in this section to a particular pixel,we use lowercase letters,such as p and q.

2.5.1 Neighbors of a Pixel: $N_4(p)$, $N_D(p)$ and $N_8(p)$

4-neighbors of p: **a** pixel p at coordinates (x,y) has four horizontal and vertical neighbors whose coordinates are given by: (x+1,y); (x-1,y); (x,y+1); (x,y-1)

This set of pixels is denoted by $N_4(p)$. Each pixel is a unit distance from (x,y) ,and some of the neighbor locations of p lie outside the digital image if is on the border of the image.We deal with this issue in Chapter 3.

Four diagonal neighbors of p have coordinates (x+1,y+1); (x+1,y-1); (x-1,y+1); (x-1,y-1) and are denoted by $N_D(p)$. These points, together with the 4-neighbors,are called the **8-neighbors** of p, denoted by $N_8(p)$. As before, some of the neighbor locations in $N_D(p)$ and $N_8(p)$ fall outside the image if (x,y) is on the border of the image.

2.6 An Introduction to the Mathematical Tools Used in Digital Image Processing

Objectives:
(1) to introduce you to the various mathematical tools we use throughout the book; and
(2) to help you begin developing a "feel" for how these tools are used.

2.6.1 Array versus Matrix Operations

Consider two 2 x 2 images (matrixes 2x2 a_{ij} and b_{ij}).

[a11 a12] and [b11 b12]
[a21 a22] [b21 b22]
Array product is: [a11.b11 a12b12]
[a21b21 a22b22]

Matrix product is: [a11.b11+a12.b21 a11.b12+a12b22]
[a21.b11+a22.b21 a21.b12+a22.b22]

2.6.5 Spatial Operations

Spatial operations are performed directly on the pixels of a given image. We classify spatial operations into three broad categories:

(1) single-pixel operations

The simplest operation we perform on a digital image is to alter the values of its individual pixels **based on their intensity**.This type of process may be expressed as a transformation function, T, of the form:

$s = T(z)$ (2.6-20), where z is the intensity of a pixel in the original image and s is the (mapped) intensity of the corresponding pixel in the processed image.

For example, Fig. 2.34 shows the transformation used to obtain the negative of an 8-bit image, such as the image in Fig.2.32(b), which we obtained using set operations. We discuss in Chapter 3 a number of techniques for specifying intensity transformation functions.

(2) neighborhood operations

Let Sxy denote the set of coordinates of a neighborhood centered on an arbitrary point (x, y) in an image,f.

Neighborhood processing generates a corresponding pixel at the same coordinates in an output (processed) image, g, such that the value of that pixel is determined by a specified operation involving the pixels in the input image with coordinates in Sxy.
For example, suppose that the specified operation is to compute the average value of the pixels in a rectangular neighborhood of size m x n centered on (x,y).The locations of pixels in this region constitute the set Sxy. Figures 2.35(a) and (b) illustrate the process. We can express this operation in equation form as

$$g(x,y) = 1/(m.n) \sum f(r,c) \qquad (2.6\text{-}21)$$

The net effect is to perform local blurring in the original image.This type of process is used, for example, to eliminate small details and thus render "blobs"corresponding to the largest regions of an image.

(3) geometric spatial transformations.

Geometric transformations modify the spatial relationship between pixels in an image. These transformations often are called rubber-sheet transformations because they may be viewed as analogous to "printing" an image on a sheet of rubber and then stretching the sheet according to a predefined set of rules. In terms of digital image processing, a geometric transformation consists of two basic operations:

(a) a spatial transformation of coordinates, and
(b) intensity interpolation that assigns intensity values to the spatially transformed pixels.

The transformation of coordinates may be expressed as

$$(x,y) = \{(v, w)\}$$

where (v,w) are pixel coordinates in the original image and (x, y) are the corresponding pixel coordinates in the transformed image.
For example, the transformation

$$(x,y) = T\{(v, w)\} = ((v/2,w/2)$$

shrinks the original image to half its size in both spatial directions.
One of the most commonly used spatial coordinate transformations is the affine transform (Wolberg [1990]), which has the general form

$$[x \;\; y \;\; 1] = [v \;\; w \;\; 1\} \; T = [v \;\; w \;\; 1] \left| \quad \right| \qquad (2.6\text{-}23)$$

This transformation can scale, rotate, translate, or sheer a set of coordinate points,depending on the value chosen for the elements of matrix T.

Image registration is an important application of digital image processing **used to align two or more images of the same scene**.

In the preceding discussion, the form of the transformation function required to achieve a desired geometric transformation was known. In image registration, we have available the input and output images, but the specific transformation that produced the output image from the input generally is unknown. **The problem, then, is to estimate the transformation function and then use it to register the two images.**

To clarify terminology, the input image is the image that we wish to transform, and what we call the reference image is the image against which we want to register the input.
Comments on image alignment

For example, it may be of interest to align (register) two or more images taken at approximately the same time, but using

different imaging systems, such as an MRI (magnetic resonance imaging) scanner and a PET (positron emission tomography) scanner.

Or, perhaps the images were taken at different times using the same instrument, such as satellite images of a given location taken several days, months, or even years apart.

In either case, combining the images or performing quantitative analysis and comparisons between them requires compensating for geometric distortions caused by differences in viewing angle, distance, and orientation; sensor resolution; shift in object positions; and other factors.

One of the principal approaches for solving the problem just discussed is to use **tie points** (also called control points), which are corresponding points whose **locations are known precisely in the input and reference images**. There are numerous ways to select tie points,

ranging from interactively selecting them to applying algorithms that attempt to detect these points automatically.

In some applications, imaging systems have physical artifacts (such as small metallic objects) embedded in the imaging sensors. These produce a set of known points (called reseau marks) directly on all images captured by the system, which can be used as guides for establishing tie points.