Image Processing
Summary

Giacomo Deodato

Fall, 2017

Sampling & Quantization

Image sampling is the digitization of *(x, y)* coordinates.

Grey-level quantization is the digitization of pixel amplitude.

Aliasing is an effect that causes different signals to become indistinguishable (or aliases of one another) when sampled. It also refers to the distortion or artifact that results when the signal reconstructed from samples is different from the original continuous signal.
Aliasing can occur in spatially sampled signals, for instance moiré patterns in digital images. Aliasing in spatially sampled signals is called spatial aliasing. Aliasing is generally avoided by applying low pass anti-aliasing filters to the analog signal before sampling.

Interpolation can be zero-order (replication) or first-order (linear):

$$f(n+a) = (1-a) \cdot f(n) + a \cdot f(n+1), \qquad 0 < a < 1$$

Shannon's theorem : $f_s > 2f_N$ establishes a sufficient condition for a sample rate that permits a discrete sequence of samples to capture all the information from a continuous signal of finite bandwidth.

Scalar quantization uses *decision level* (d_i, input) to get *reconstruction level* (r_i, output):

$$if\ d_i \leq f(x,y) < d_{i+1}\ then\ f_q(x,y) = r_i$$

Histogram

Image histogram shows the probability density associated to the grey-level k: $p(k) = n(k)/N$

Histogram equalization is used to enhance images, it follows the *Khoros routine*: build the histogram and the cumulative histogram (CH); for each grey level k do: $k \leftarrow INT(CH(k) \cdot (K-1)/N)$, where $N$is the number of pixels.
The main limitation of this algorithm is it can't take into consideration the position of the pixels.

Histogram thresholding is used to binarize an image (black and white), it corresponds to a 2-levels quantization.

Fourier transform

$$F(u) = \int_{-\infty}^{\infty} f(x) e^{-2\pi jux} dx, \qquad F^{-1}(x) = \int_{-\infty}^{\infty} F(u) e^{2\pi jux} du$$

DC value

$$\bar{f}(x,y) = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y), \qquad F(0,0) = \frac{1}{N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y)$$

$$\bar{f}(x,y) = \frac{1}{N} F(0,0)$$

Separability A 2D Fourier transform can be separated into two 1D Fourier transforms.

Display It is better to display $log(1+|F(u,v)|)$ than the normal 2D Fourier transform $|F(u,v)|$.

Spatial and frequency domains When we translate into the spatial domain, nothing changes in the frequency domain. When we rotate, then also the frequencies rotate. If we zoom, the frequencies scale down but they maintain the same shape.

Filtering

Low and high frequencies represent respectively uniform areas, and edges and noise.

Image filtering can be performed either via *Fast Fourier Transform* (FFT), by multiplying the Fourier transform of the image and the filter and then reversing the result; or using *convolution*.
The relevant parameters are the size fo the kernel used and the value of its coefficients.

Low pass spatial filters

Average filter is a spatial low pass filter that performs image smoothing: it reduces the noise but smooths the edges. The sum of the kernel coefficients must be 1 and they have to be all the same.

Median filter is non linear, it replaces the pixel with the median of the neighbourhood defined by the kernel. It is useful to reduce noise while preserving the edges. Moreover it does not create new values.

Wiener filter tailors itself to the local image variance: when the variance is large, it performs little smoothing; when the variance is small, it performs more smoothing. This approach often produces better results than linear filtering like averaging, and it is more selective, it preserves edges and other high-frequency parts of an image. Moreover, it performs better than the median filter because it takes into consideration both the mean and the standard deviation of the neighbourhood of each pixel instead of just the median.

High pass spatial filters

Gradient filter uses the following kernels to get the gradient on the x and y axis and then it sums then.

$$G_x = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix}, \qquad G_y = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix}$$

It is used to detect edges, if the value of the gradient is greater than a threshold value then the point is an edge.
Other kernel similar to the gradient are:

$$Prewitt = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}, \qquad Sobel = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$$

Laplacian filter is also used for edge detection, it can use one the following kernels:

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \qquad \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

To detect the edges it is necessary to perform the zero-crossings, an operation that could easily be not precise depending on the slope of the curve.

Canny edge detector has the best results, the algorithm can be broken down to 5 different steps:
1. Apply Gaussian filter to smooth the image in order to remove the noise
2. Find the intensity gradients of the image
3. Apply non-maximum suppression to get rid of spurious response to edge detection
4. Apply double threshold to determine potential edges
5. Track edge by hysteresis: Finalize the detection of edges by suppressing all the other edges that are weak and not connected to strong edges.

Edge detection steps

Pre-processing noise reduction: low pass

Processing gradient and thresholding

Post-processing edge thinning and outliers removal

Representation and description hough/radon transform

Hough Transform

It was originally designed for line detection but can also be used for other analytical curves.
A point in the Hough space (H) corresponds to a line in the image space (I).
A point in I corresponds to a sinusoid in H.
Points on the same line in I give curves passing through a common point in H. Points on the same curve in H give lines passing from the same point in I.
The Hough transform can work in different ways (we assume to work on a binarized image):

- From m to 1. Where m is the number of pixels from the image. In our case all the couples of points ($m = 2$) in the image are associated to a line and the parameters (ρ, θ) of the line are used to index and increment the value of a point in H. If the final value is greater than a given threshold, then the coordinates of that point are the parameters of and edge line.

- From 1 to m. Where m is the number of curves corresponding to a point. In our case it corresponds to the different slopes of the lines passing from a point. For each non-zero pixel, increment the points in H whose coordinates correspond to the parameters of the lines passing to the point.

- From m to n. Where m is number of points from the image and n is the number of possible curves passing from them.

Radon transform is strictly related to Hough transform when doing line detection, we could say that the former is a simplified version of the latter. The main difference between the two transformations is in the mathematical formulation, while the Hough transform is a discrete algorithm, the Radon transform is defined as an integral. The Radon transform is also easier to understand from a conceptual point of view: it projects the whole image over an axis rotating from 0 to 179 degrees. The Hough transform instead takes one or more features (pixels) from the image and associates it to one or more parametric curves in the Hough space.

The sum over any column of the Radon transform is always the same because it corresponds to the amount of information of the image, it is constant because each column contains the entire projection. For example, if we make the Radon transform of an horizontal line we'll notice that the vector corresponding to 0 degrees has uniform distributed values while the column corresponding to 90 degrees has a unique peak whose value is equal to the sum of all the values of the horizontal projection.

Morphological operators

Set operators Union, intersection, complement, inclusion and difference.

Erosion Given an object A and a structuring element B_p:
$er(A, B_p) = \{p | B_p \subset A\}$.

Dilatation Given an object A and a structuring element B_p:
$dil(A, B_p) = \{p | B_p \cap A \neq \varnothing\}$.

Opening $dil(er(A, B_p), B_p)$, it cuts narrow isthmus, remove small islands and narrow capes.

Closing $er(dil(A, B_p), B_p)$, it fills narrow canals, removes small lakes and narrow gulfs.

Properties They are translation invariant. Erosion and dilatation are not the inverse of each other but the dual: $dil(A, B_p) = (er(A^c, B_p))^c$.

Local Binary Pattern

LBP is a gradient based descriptor (differences are more relevant than the value of the pixel itself). It is widely used to compare face images by comparing their LBP histograms. To build the histogram we calculate the LBP codes from the original image by analysing the differences between a pixel and its neighbourhood. Then we do quantization by cutting the LSBs of the code and we represent these on the histogram.

Quad tree: split & merge segmentation

The process to build a quad tree goes trough two main steps: split and merge. During the split we divide the picture into smaller and smaller areas until reaching a given uniformity criterion. Each region is divided into smaller region if a its test value is greater than a given threshold. The test value is build by dividing for the cardinality of the region the sum of the squared differences between the value of each pixel and the average value of the region. During the merge phase we merge neighbouring areas according to a similarity criterion. The similarity is calculated as the sum of the squared differences between each point inside the regions to be merged and the mean value of all the points of the two regions.

Optical flow

The optical flow, or apparent motion field, is the relative motion between an observer and the scene. It is calculated in a two step process.

Local estimation

Calculate the displacement $d_i = (d_x, d_y)$ for each pixel $p_i = (x_i, y_i)$, where $d_x = x'_i - x_i$ for each pixel in the image.
To start the procedure we need to suppose that the luminance of a pixel is

constant over time. Assuming that δx and δy are small, we can write:

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t)$$

We can develop the second term with Taylor series:

$$I(x + \delta x, y + \delta y, t + \delta t) = I(x, y, t) + \frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t + H.O.T.$$

$$\frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t = 0, \qquad \frac{\partial I}{\partial x}V_x + \frac{\partial I}{\partial y}V_y = -\frac{\partial I}{\partial t}$$

The final equation is the relation between spatial and temporal gradients, but it is an equation in two variables, therefore we need to compute V_x and V_y by means of an iterative process that will lead to:

$d^{(i+1)} = d^{(i)}(p, t) - \epsilon \cdot sign(DFD(p, t, d^{(i)})) \cdot sign(\nabla I(x - d_x^{(i)}, y - d_y^{(i)}, t - 1))$

Global interpretation

Finally we'll be able to interpolate the linear equation for the displacement of all the pixels. From this equation the slope is the *zoom factor*, if it is close to 1, then there is no zoom, else, if it is very smaller than 1 there is a backward zoom, otherwise there is a forward zoom.
Furthermore, the two components of the known term of the equation will give us the vertical and horizontal pan of the image, that are the mean displacement of all the pixels.

Depth cameras (RGB-D)

Stereo setup
Passive system.
Two calibrated cameras.
Find points match between two images and triangulate to estimate the distance.
Not very precise and not reliable in presence of textured surfaces.

Time of flight
Measures light pulses (IR) round trip from emitter to object to sensor, pulses generation is hard to achieve but there are different modulation techniques available.
Resolution depends on time measure precision (high accuracy required and integration time disadvantages).
Indoor and outdoor (errors due to light scattering).

Structured light
Measures distortion in a pattern of IR projected over the scene.
Very fast but short depth range.
Only indoor.

Light field
Passive system.
Find the direction where the light is from.
Sensible to light variations.
Indoor and outdoor.

Stereo

Internal parameters are the focal length f and the projection of the optical centre in the image plane, *c(0, 0, f)*.

External parameters are translation and rotation, the camera pose and orientation.

Disparity is the difference of projected positions into the left and right images.
From the internal parameters we can link a point in the scene to a point in the image, from disparity we can derive depth under the assumption that the two cameras are parallel.

$$\begin{cases} x = f\dfrac{X}{Z} \\ y = f\dfrac{Y}{Z} \end{cases} \qquad \begin{cases} x_r - x_l = f\dfrac{B}{Z}, \\ y_r = y_l \end{cases} \qquad B: \ baseline$$

Epipolar geometry defines the segment where, given a point on the left image, is possible to find it on the right one. The segment goes from E_l, the intersection between the line $C_l - C_r$ and the left image plane, to p_l, the position of the point P in the left image.

The human eye

The retina is a membrane on the inner wall of the eyeball, it receives the image and converts it to nerve impulses.

The fovea at the center of the retina, is the region of highest visual activity and cone density.

Rods and cones are the photosensitive cells in the retina. Rods are sensitive to brightness while cones are sensitive to colours. The cones predominate in the fovea and the rods are at the periphery of the vision. Color sensitivity is given by three type of cones. Depending on the frequency they detect, they are called red (L) 64%, green (M) 32% and blue (S) 4%.

Colorimetry

Color perception depends on:
1. Illumination source
2. How the object absorbs light
3. Characteristics of human eyes
4. Color interpretation of the human brain

Light spectrum visible to the human eye goes from 780 nm to 380 nm of wavelength. The wavelength determines the colour and the amplitude determines the brightness.

Additive color mixing uses RGB as primary colors and YCM as secondary ones. Starting from black it adds lights to create colors, used in television. It exploits Grassman's laws.

Subtractive color mixing uses YCM as primary colors and RGB as secondary ones. It starts from white, used in printers with an auxiliary black.

White balance is a technique used with color correction to get better colors from a digital image.

Bayer interpolation is an arrangement of color filters where there are more G. R and B values are interpolated from the nearest neighbours.

Colormaking attributes are: the hue (color), lightness or brightness or value (perceptual response to luminance), chroma or saturation (purity of hue).

Color models

Munsell uses a visual approach, the central column represents the color value, each arm radiating from the central column represent a hue and along each arm, from the center to the outer limit there is the chroma.

RGB model uses a physical approach. The color space can be represented as a cube which is then cut by the plane R + G + B = 1 (forming the Maxwell triangle). Colors are expressed as a mixture of red, green and blue but, in some cases, the red value has to be negative. Moreover this model is not very intuitive, it is not easy to determine RGB values from a color, and it is perceptually non linear.

HSV / HLS are color models that transform the RGB model in order to describe colors in terms more natural to an artist (a cone). Value and lightness are not the same thing (also, value = sqrt(luminance)).

XYZ model is a transformation of RGB defined to describe the full space of perceptible colors without the negative values problem.
Colors are indexed using x and y, while Y is used fro the brightness. To pass from xyY to color space the following transformation is used:

$$\begin{cases} x = \dfrac{X}{X + Y * Z} \\ y = \dfrac{Y}{X + Y * Z} \\ z = \dfrac{Z}{X + Y * Z} \end{cases}$$

Inside the *gamut* (section of the colors prism) we change the saturation while moving on the border we change the hue. It is also possible to define dominant and complementary wavelengths and to do additive color mixing.
Inside the gamut are defined the Mac Adam's ellipses which define regions where color differences are not perceivable because fo the scales of the chromaticity diagram are not uniform.
In order to solve this problem we can transform the coordinates according to the Uniform Chromaticity Scale (UCS).

Lab model is a perceptually uniform space where L is luminosity, a is the red/green axis and b is the yellow/blue axis

YIQ is a color model used for TV. Y encodes luminance while I and Q encode the color. More bits are used to encode Y since people are more sensitive to the luminance variations.

YCrCb is a standard color model where Y is a linear combination of RGB, Cr = R - Y and Cb = B - Y.