# Image processing

https://github.com/Grufoony/Physics_Unibo

# Contents

# Chapter 1

# Introduction to images

The first question that someone should ask themselves is, of course, what is an image? We know from everyday life that we can't have an image without light. In more general terms, in order to have light we need to have some sort of radiation.
So we can say that an image is a distribution of matter, which is called an *object*, that is visible when illuminated. Alternatively it can be defined as a measure of the intensity of the reflected radiation.

Any imaging technique is characterized by the way that the object and the radiation interact. An imaging system collects radiation emitted by objects.

We define two quantities: The energy intensity, $E$, and the energy flux, $Q$.
Through this two quantities we can define *irradiance*, *radial intensity* and *radiance*:

$$irradiance = \frac{dQ}{dA} \left[ \frac{W}{m^2} \right]$$

$$radialintensity = \frac{dQ}{d\omega} \left[ \frac{W}{sterad} \right]$$

$$radiance = \frac{dQ}{d\omega dA} \left[ \frac{W}{sterac \cdot m^2} \right]$$

Usually sources of radiation are polychromatic, so there is a spectrum. It is thus essential to know how the radiation interacts with objects.
Different wavelengths interact differently, thus producing different images. This can be used in what is called *multispectral imaging*.

# Chapter 2

# Digital images

When talking about digital images, we must define the concepts of geometry and radiometry.

Geometry is the relationship between location and size of the objects in the 3D world and their representation in the image plane.

Radiometry is the relationship between the amount of light radiating from a point and the amount of light impinging on the correspondend image point.

The most basic image formation device (and the first one historically) is the pin-hole camera. The pin-hole camera consists of a closed room with a small hole (of the order of the millimiter) in one of the walls. When the light coming from the object reaches the hole, the image is formed upside down in the opposite wall of the chamber. The size of the image in the image plane depends on the object's distance from the pin-hole.

If a point $M$ in the 3D space is characterized by 3 coordinates $(x, y, z)$, the image point in the image plane is characterized by 2 coordinates $(u, v)$. The two sets of coordinates are related by the geometrical equations

$$u = \frac{fx}{z} \qquad \frac{fy}{z}$$

All the light rays are considered to be parallel to the optical axis and orthogonal to the image plane.

We define $\Delta z$ as the size of the object with respect to its distance from the camera, so as its thickness. Then, if $2\Delta z$ is small with respect to $z_0$, we have that

$$\frac{f}{z_0 + \Delta z} \approx \frac{f}{z_0 - \Delta z} \approx \frac{f}{z_0}$$

which means that

$$u \approx \frac{f}{z_0}x \qquad v \approx \frac{f}{z_0}y$$

This approximation of course is only valid for small objects, or objects that are close to the optical axis.

The quality of the image depends heavily on the size of the hole.

- If the hole is too big, we have that a point of the object becomes a spot in the image. This means that the image is going to be blurry.

- If the hole is too small, we can have diffraction effects, that again, end up blurring the image.

The solution for balancing the effects of big and small pinholes and settling for a middle ground is in the use of lenses.

Lenses have two main strengths: They allow to gather light coming from a point on the object and focus it into a single point on the image; since the aperture size of the lense is larger than that of a pinhole, the exposure times can be reduced as well.
Lenses solve another problem as well: In a pinhole camera, we have that many points from the object space are mapped into a single in the image plane. So, an image on the image plane can come from several objects in the real space.
On the other hand, a lens brings into focus only those object points that lie within one particular plane parallel to the image plane.
So, when the distance between lens and image plane is equal to $v$, only those points that are at a distance $u$ are brought into focus, with $u$ give by

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f} -> u = \frac{vf}{v - f}$$

In other words, if we bring an object into focus at a distance $u$, we must set the distance $v$ between the image plane and the lens to

$$v = \frac{uf}{u - f}$$

The object points that do not lie within this plane end up being blurred.

In a real camera ther is an object called diaphragm, whose purpose is to control the opening of the lens. By reducing the size of the aperture, it reduces the amount of light that reaches the sensor, and also reduces the size of the circle of confusion, thus increasing the depth of field.
The *field of view* is defined as the portion of space that actually projects onto the camera. It describes then the cone of viewing directions of the device.
The FOV depends of the effective area of the image sensor, the width $w$ and the height $h$:

$$FOV_v = 2\arctan\frac{w}{2f} \quad FOV_h = 2\arctan\frac{h}{2f}$$

The *magnification factor* is then defined as

$$M = \frac{x}{X} = \frac{v}{u} = \frac{f}{u}$$

where $x$ is the size of the image whereas $X$ is the size of the real object.
We see clearly that the magnification factor is proportional to the focal length.
Since the FOV depends on the focal length as well, we can say that the magnification factor and the FOV are linked. In particular:

- if $f$ is small -¿ large FOV, small M

- if $f$ is large -¿ small FOV, large M

## 2.1 Radiometry

Radiometry enables us to know what a pixel value implies about surface lighness and illumination. So radiometry links the effective brightness of an object point with the respective image point's pixel value.

The amount of object coming out of an object, $f(x, y)$, can be expressed as

$$f(x, y) = i(x, y)r(x, y)$$

with $0 < f(x, y) < \infty$, $0 < i(x, y) < \infty$ and $0 < r(x, y) < 1$, where $i(x, y)$ is the light coming from the source and $r(x, y)$ is the object's reflectance.
$i(x, y)$ is determined by the light source, whereas $r(x, y)$ depends on the surface of the object.

An acquisition device can be describd as a system that, given an input $f(\xi, \eta)$, produces an output $g(x, y)$, which represents the acquired image.
The two functions are usually related by

$$g(x, y) = \int_{-\infty}^{\infty} h(x, y, \xi, \eta)f(\xi, \eta)d\xi d\eta$$

where $h(x, y, \xi, \eta)$ is the system's response when presented with the unit impulse as unit, so it defines how a point $(\xi, \eta)$ in the object space contributes to the formation of the image in a particular point $(x, y)$.
In other words, $h$ describes the distorsions introduced by the system.
For linear and shift invariant processes, the previous relation can be simplified and written as

$$g(x, y) = \int_{-\infty}^{\infty} h(x - \xi, y - \eta)f(\xi, \eta)d\xi d\eta$$

Acquiring a series of static images, one would expect that the intensity value $g$ in one point remains the same for all the images, but this doesn't always happen, and we can see fluctuations of the $g$ value around a certain level.
In this cases we say that noise is present and the value's fluctuations are a way to measure its value.

Noise can be defined as the uncertainty or imprecision with which a signal is recorded. Images can have many sources of noise, like variations in the source's emission, electronic noise, interferences and so on. To take this effects into account in the previous formula we simply add a term representing the noise.

$$g(x,y) = \int_{-\infty}^{\infty} h(x - \xi, y - \eta)f(\xi, \eta)d\xi d\eta + n(x,y)$$

where in many situations we can model $n(x,y)$ to be gaussian.

## 2.2  Production of images

Acquisition devices register the amount of radiation impinging each point of the system as analog signals, and this produces digital images, which are basically numerical representations of an object. This means that analog signals need to be converted, through a process called digitization, and after that the digital images are made available for computer processing.

So the production of a digital image can be divided into three steps:

- Measure of the analog signal.

- Sampling, the process of measuring the grey levels for each pixel location.

- Quantization, the process of dividing the scale into a discrete set.

So digitization, that convers an image from its original form into digital form, is the combination of this three step, that convers an image from its original form into digital forms.

The number inserted into the digital image at each pixel location (which is the point's grey level) reflects the brightness of the image at the corresponding point, which as we have just said is sampled and quantizied.
This means that a digital image is basically a matrix, that for each element contains an integer, corresponding to the pixel's grey level.
Digital images are characterized by two resolutions: The spatial resolution (sampling density) and the grey-level resolution. The first is the pixel spacing, so the number of sample points per unit of measure, whereas the grey-level resolution is the number of grey-levels.
Spatial sampling can be described as a multiplication of the function $f(x,y)$ and the function $s(x,y)$, which is defined as

$$s(x,y) = \sum_i \sum_j \delta(x - i\Delta x, y - j\Delta y)$$

that basically defines the sample grid.
So the sample image can be defined as

$$f_c(x, y) = s(x, y)f(x, y) = \sum_i \sum_j f(i\Delta x, j\Delta y)\delta(x - i\Delta x, y - j\Delta y)$$

So what we do is discretize the input, and in the end the sampled image only consists of the samples acquired at each node of the sampling grid.

At this point we can ask ourselves, how many samples and grey levels do we need to represent the object in an acceptable or good way?
Intuitively, we can say that the pixel size should be of comparable size with with the smalles detail that we want to perceive from the image.
More quantitatively, the *Nyquist rate* says that the sampling interval must not be greater that half the size of the smallest resolvable feature of the image.
If the sampling interval is too large, we can have aliasing.
If the spatial resolution is too low, what we get is that the image becomes very pixellated, so we aren't able to distinguish small details anymore. On the other hand, if the grey-level resolution is too low, we lose the colour difference of neighbouring points, and this leads again to a loss of detail.

When choosing the resolution, we must balance pros and cons to find the right middle ground.
This is because, while it is true that high resolution images contain more information, all that information might not be needed, but it would still require more storage space and execution time for the various acquisition and processing steps. Furthermore, high definition images require all the acquisition, processing and visualization devices to support that level of definition. And finally, low risolution images are less affected by statistical noise.

A digital image can be described as a matrix

$$f(x, y) = \begin{pmatrix} f(0,0) & f(0,1) & ... & f(0, N-1) \\ f(1,0) & f(1,1) & ... & f(1, N-1) \\ ... & ... & ... & ... \\ f(M-1, 0) & f(M-1, 1) & ... & f(M-1, N-1) \end{pmatrix}$$

or also as a vector.

## 2.3 Quality of images

When we apply a certain operation to an image, we are obviously trying to improve its quality. In order to be able to say if we have managed to do so, we need a criteria for assessing the quality of an image.

There are a lot of sources for image degradation, so it is important to find a way to quantify image quality. The quality assessment can be subjective or objective:

- subjective: involving human observers. The best way to find the quality of an image is to look at it, because the human eye is going to be the final observer of any image.

- objective: the goal of objective evaluation is to develop a quantitative measure that can assess the distortions in the images. A possible way to evaluate objective image quality is through the use of the mean squared error

$$E = \frac{\sum_n (g_i(n) - g(n)^2)}{\sum_n g_i(n)^2}$$

where $g_i$ represents the ideal image and $g$ represents the reconstructed image.

# Chapter 3

# Digital image processing

Geometric transformations are common in computer graphics, and are often used in image analysis. They basically consist of rearranging pixels in the image plane.
A geometric transform consists of two basic steps:

- determining the pixel coordinate transformation mapping of the coordinates of the moving image pixel to the point in the fixed image.

- determining the brightness of the points in the digital grid of the transformed image.

The most common geometric transformations are rotations, reflections, translations and scaling (shrink or zoom).

Following a geometric transformation, a point might not fall on the grid points in the new space. This is very likely actually, because the image grid is discrete. So, the point has a certain starting grey level, and we need to decide where that value will fall in the discrete grid. This is called interpolation. The easiest way is to put the point in the nearest neighbour (nearest neighbour grey level interpolation).
Another way is to share the point to 4 pixels, expressing it as a linear combination.
So in nearest neighbour grey level interpolation we are just moving grey-levels, but their values do not change.

The most common applications of geometric operations are:

- elimination of geometric distortions

- scaling the image

- rotating the image

- alignment of images

A great tool for studying the distribution of grey levels in an image are image histograms.

Image histograms show how many times a particular grey-level appears in an image. With this histograms I can see how much I'm using the intensity level, but we don't know where those pixels are, so we are completely losing the spatial information. In addition to this, histograms are not unique: we can have several images all with the same histogram (you can't reconstruct the image starting from the histogram).

When the constrast is low, the number of grey levels used is low, so the histogram must be narrow.

On grey-level operation is thresholding, which converts pixel into black or white depending on whether the original color value is within the threshold range. This is very useful when we want to discriminate foreground and background.