Chapter 1

MULTIMEDIA SIGNAL PROCESSING

Introduction

Introduction

- Digital image & video processing : a new domain and technology
  - Mathematics of information
  - Signal processing
  - Electronic & optronics systems
  - Computer sciences and micro-processors (VLSI, DSP)

- Lots of methods
  - ad'hoc & founded as well
  - Needs of methodology

- Widespectrum of applications

When we speak of *digital image processing*, we mean the set of techniques used to modify a digital image in order to improve it (in terms of quality), or to reduce its size (in terms of bits compression encoding) or to get information out of it. Processing digital images is a new sector of knowledge, which has quickly developed thanks to the emergence of new information technologies. It relies mainly on the mathematics linked to information, signal processing, electronic systems and the advance in microprocessor computation capacities, particularly those that have been exclusively developed for signal processing and which offer high computation speed and capacity (DSP, etc.).

As digital image processing is in its early stages and its scope of application is quite spread, it has quickly become apparent that a methodology must be created and the domains of application separated out. Image processing can now be seen as four distinct fields of action: analysis, synthesis, coding and quality enhancement.

To start with, a description of these four domains will allow us to better understand what image processing really is. Following this, we will look at how to obtain the digital images (by digitization of analog images) that we wish to process.

Introduction

A picture is worth one thousand words
and a video is worth one thousand sentences

Rich info. from visual data.

Examples of images around us

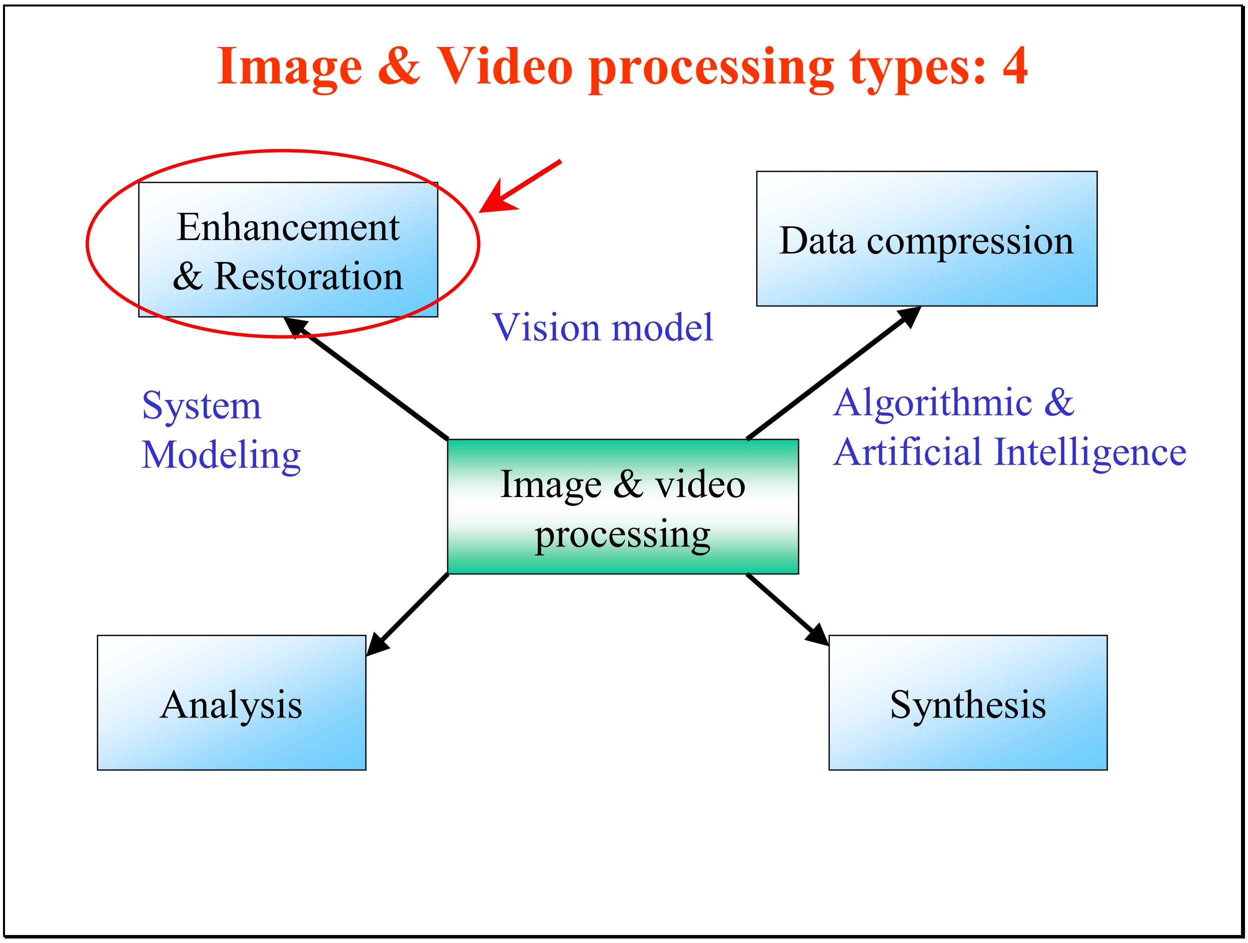
- natural photographic images
- artistic and engineering drawings
- scientific images (satellite, medical, etc.)

"Motion pictures" => video

- movie, TV program, news
- family video
- surveillance and highway/ferry camera

We come in contact with all sorts of images in our daily environment: photographs of landscapes, of people, computer-generated drawings and paintings, images from medical radiology, satellite images, and so on. Some of these images (such as via satellites and for medicine) cannot be directly observed, while others have characteristics that can be extracted automatically, stored and sent. The processing that could be carried out on these images is highly varied, as the images we meet in our environment are also varied, by their nature and properties as well as the scenes they describe. They are all different but, evidently, it is not conceivable to create a specific type of processing for each one. This has led to a classification of image processing that does not rely on image characteristics, but rather on the objective of the processing. We can distinguish four types of domains of application for the digital image processing:

- restoration and enhancement image,
- image analysis,
- image coding with data compression,
- image synthesis.



Let's now look in detail at the four fields of action linked to digital image processing.

Enhancement and restoration:

Let's consider an observed image I_0 with which we associate a signal s_0 that we model as: $\mathbf{s_0 = \mathit{f}(s_u, d, b)}$

where: - s_u is the usable image signal obtained from an ideal image (without any loss);
- d is the distortion function which operates on the ideal image (geometric distortions, blur, etc.);
- b is noise;
- f is an observation function dependant of these two signals and of the distortion function.

The processing carried out on I_0, which will output the transformed image I_T, must enable the information contained in I_T to be used in a more efficient manner than the information in the directly observed image I_0. If a change occurs in the presentation characteristics of the image, we can talk of an *enhancement process*, whereas if there is a partial inversion in the quality loss, we can talk of a *restoration process* (for example, 2D linear filters, 2D adaptive filters etc.).

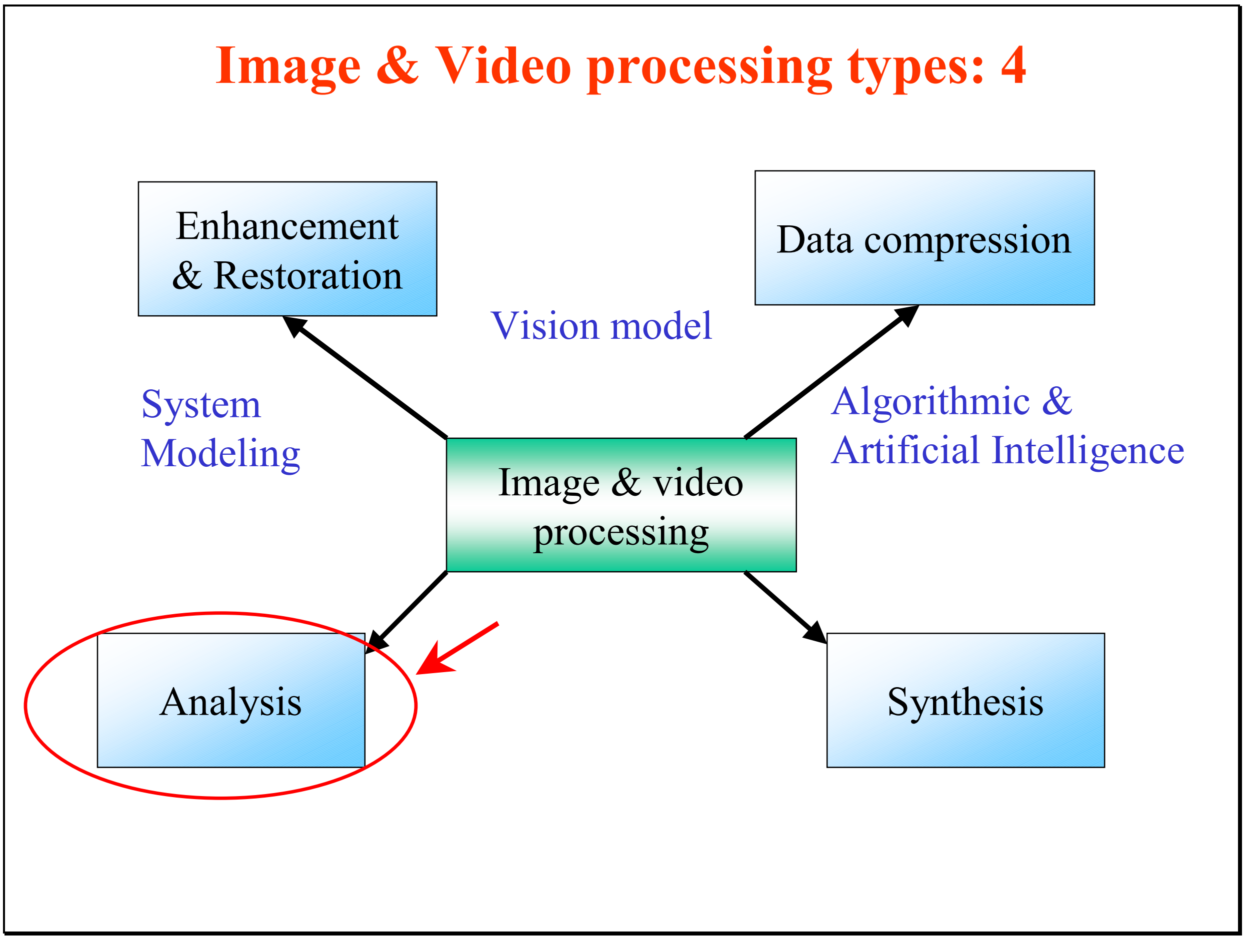


Image analysis:

This refers to partially or fully describing the scene from the observed image (objects detection, object dimensions, position in space, etc.).
Classically, this analysis process takes place in 3 successive stages:

- pre-processing and extraction of the characteristic traits,
- classification and pattern recognition,
- description and possibly interpretation of the image content.

Note however that image analysis also varies depending on the support medium: the problems raised by the analysis of 2D images, of 3D images or of moving images are quite numerous, with a special attention to moving images due to the techniques used in that case and the nature of the required objectives.

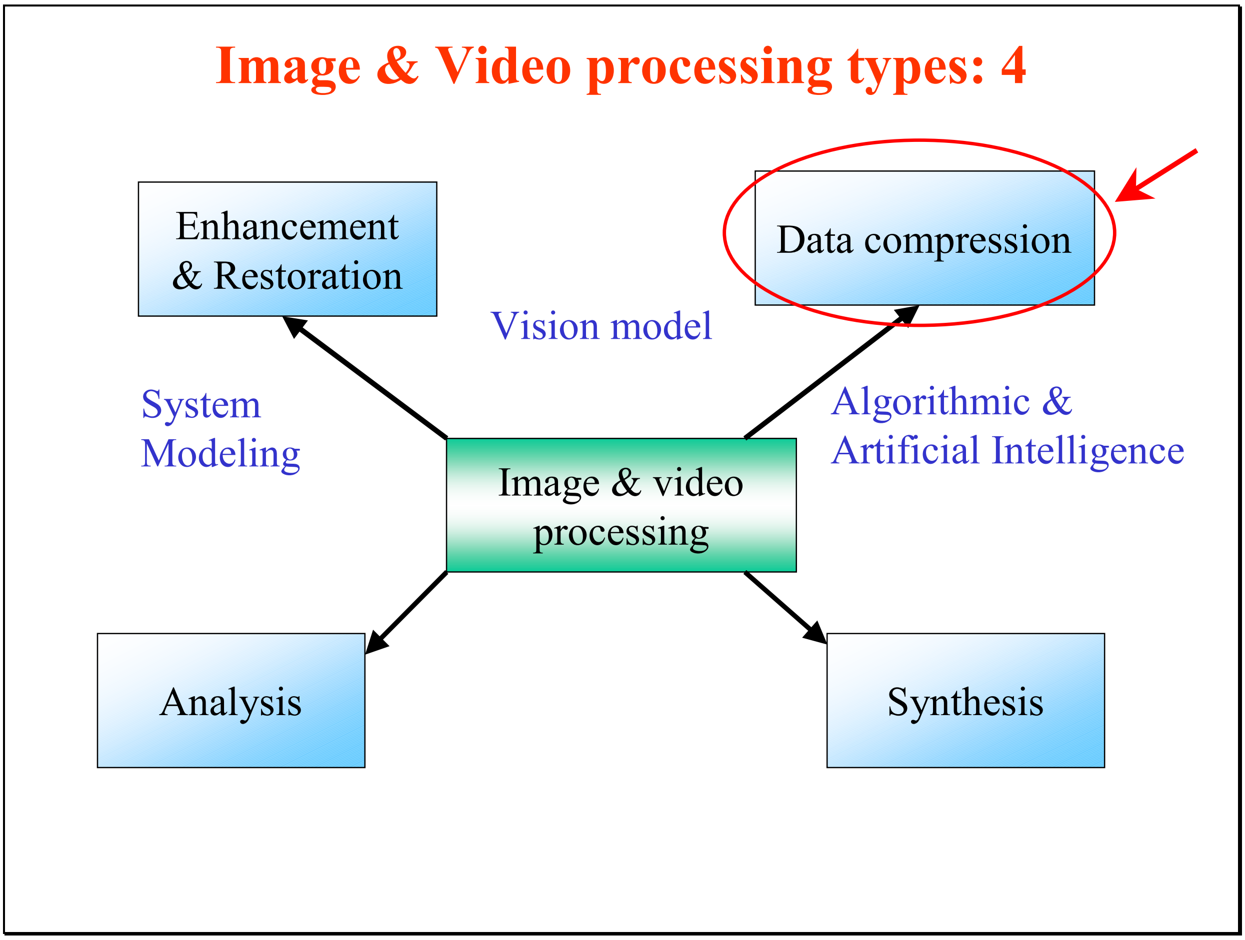


Image coding with data compression:

The basis for representing a digital image is a rectangular 2D table of elements called pixels. This implies to handle and memorize a large amount of pixels. For a simple grayscale image, typically 512×512 pixels (256 000 pixels) and 8 bits per pixel must be memorized to obtain a good resolution (8 bits to code one pixel, which gives per pixel a number of $2^8 = 256$ possible values). For high-resolution color images, or for moving images, the number of bits necessary for image representation quickly becomes enormous to store or send it.
Encoding an image aims to obtain a representation of that image that requires a greatly reduced number of bits in comparison with the original image representation.
To measure this reduction, we use a compression rate τ defined as:

$$\tau = \frac{\text{number of bits representing the basic image representation}}{\text{number of bits representing the image after encoding}}$$

This rate must be higher than 1 for real compression to take place.

As the number of bits for the basic image representation is fixed, the compression rate is in fact inversely proportional to the number of bits representing the image after encoding. If you want to make an exact reconstruction of the image after decoding, you need to use reversible encoding. This creates a constraint that means that the compression rate is often rather low. To increase the compression rate significantly, you need only to rely on a representation that is visually exact. In this case, the human eye will perceive no difference between the original image and the image that is reconstituted after decoding. In addition, the complexity of the encoding/decoding must be limited. Encoding a digital image involves finding a healthy balance between the compression rate that will be high enough to make data storage and transmission easier but that will not unduly affect the picture quality, and simultaneously keeping in mind that decoding complexity must be restrained.

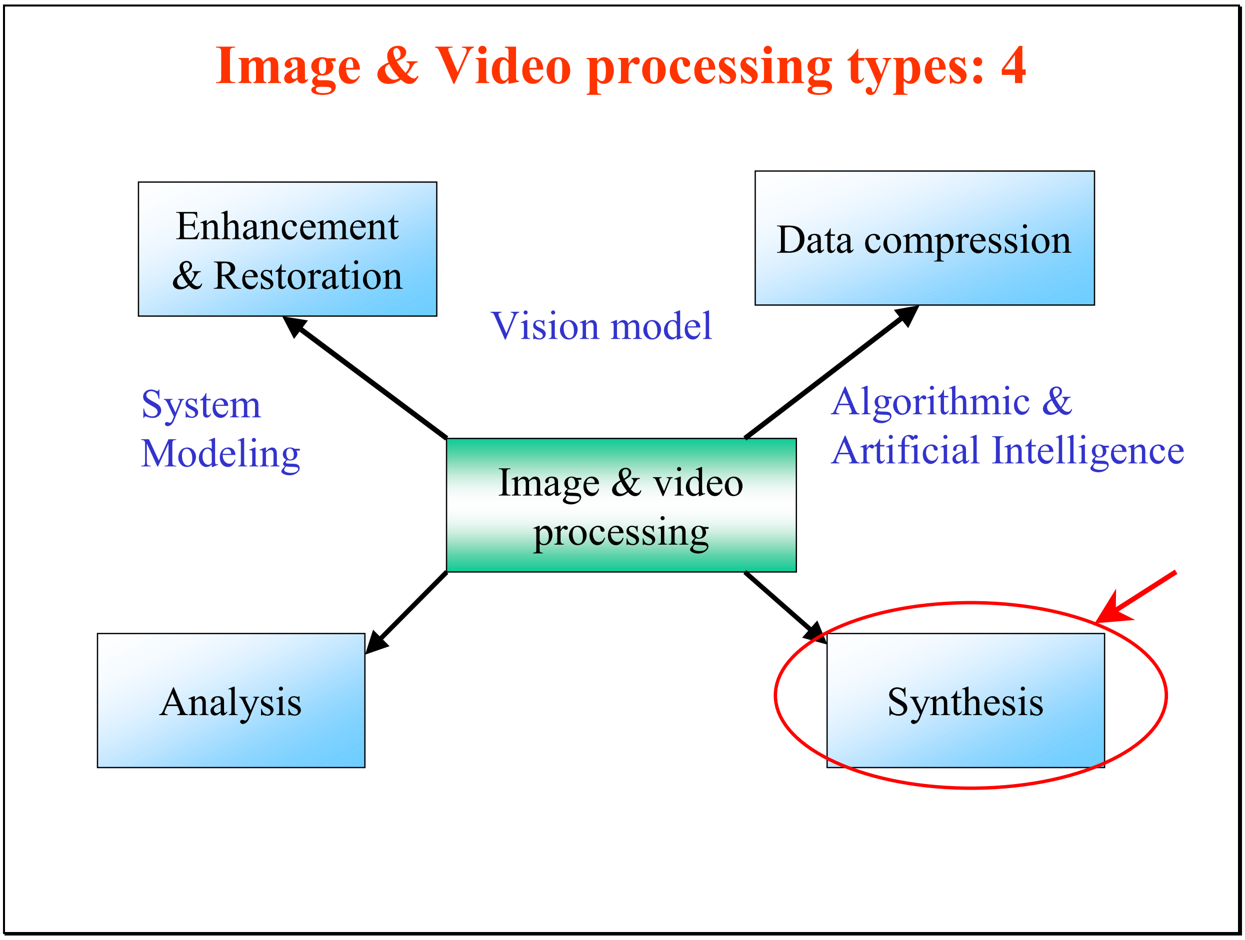
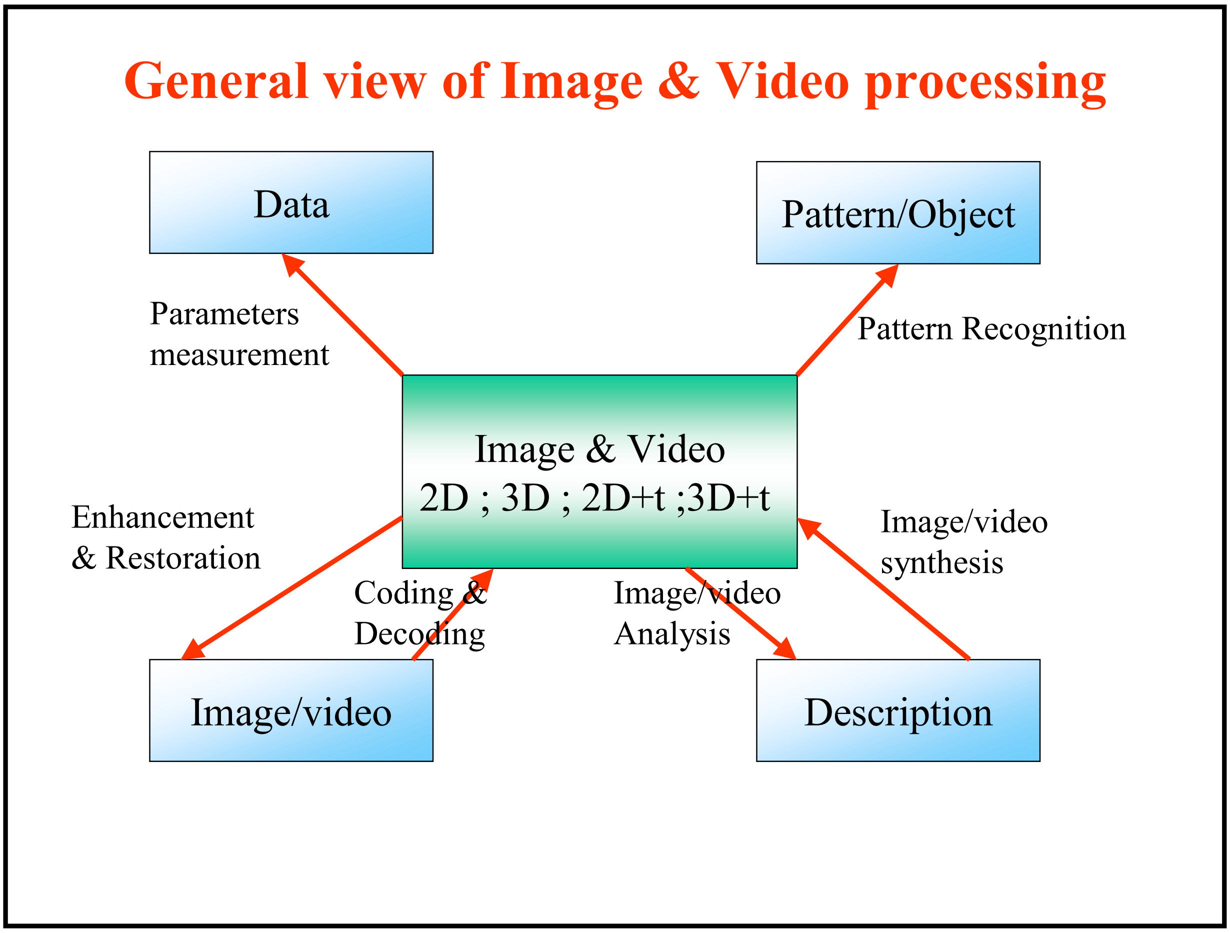


Image synthesis:

The goal of this is to reconstruct an image that resembles an image from the simulated scene, from a description of the scene, the objects making it up, its lighting characteristics (e.g. lightening orientation, intensity) as well as the capture device (CCD or CMOS camera, etc.). This reconstructed scene may resemble reality or be purely fictional. The first applications concerning image synthesis were oriented towards training simulators (flight and vehicle simulators) before involving out into other domains (audiovisual, cinema, art etc.).

We should point out that image processing also relies on studies linked to **the structure of processing machines**. An image contains a significant amount of data. In fact, for a moving image there are $N \times M \times P$ samples per second to process (N dots per line, M lines and P images per second). Image processing requires powerful calculation capacity. It needs high performance architectures with high degrees of parallelism and significant processing speeds.

Digital image processing has just been presented in accordance with the four main domains of application. We can now look at this from another angle, by concentrating of the nature of the processing results.



We can characterize image processing not simply in terms of its domains of application, but also according to the nature of the results that will be put out. There can be two types of input: an image or a description.

- <u>From an image input</u>

The output may be:

- image type
  This is the case with image coding for data compression, image enhancement and restoration of poor quality images; these three have already been presented.

- data type
  This is the case when you make an elementary image analysis. You are interested in the spatial dimensions of an object in the scene, its position or its color.

- pattern type
  This is also a case of image analysis, but a more elaborate one. This involves extracting and recognizing the objects observed in the scene.

- scene description type

  This is also a possible output for an image analysis, but in the most advanced version. The image is entirely broken up so that each object present in the scene can be recognized. The scene is described in its totality and can be interpreted.

- From a description input

For the output, the only expected type is an image. The domain of application involved is *image synthesis*. We wish to reconstruct the image according to a given description. What objects are present? What are their dimensions? Where are they in the scene? How are they lit? What are the parameters (focal length, viewing angle etc.) of the camera doing the filming?

You have now studied image processing from two different aspects: these aspects are nevertheless strongly interconnected. We can now go on to look at the characteristics of the signal that we wish to process: the image.

Basic Image/Video Properties

- Image/video signal is a 2-D/2-D+t scalar/vectorial signal which is complex:
  - Non stationary
  - Non Gaussian
  - Non isotropic
- Two (*image*) or three (*video*) main features
  - Contours: abrupt change in some important characteristics
  - Texture: spatial variation of the 2-D signal apart its local mean value (mean value, texture description)

    Edges are locally 1-D signals
  - Motion: in case of video (time-varying image sequence)

An image is a 2D scalar (grayscale image) or vectorial (color image for example) signal. Video which is a succession of images ordered temporally is a 2-D+t (t: time) scalar (gray-scale video) or vectorial (color video) signal.
The "Image" signal is complex because it is:

- non-stationary: its contents in space frequencies change with the space coordinates;
- non-Gaussian: its statistical properties do not follow a Gaussian probability law;
- non-isotropic: the properties of the image signal are not the same ones with the orientation (e.g. in the images taken on the ground, the horizontal and vertical directions are more frequent for contours than the oblique directions).

The classical methods and tools used in signal processing are often designed for stationary, Gaussian or isotropic signals (e.g. Discrete Fourier Transform). They cannot be directly applied to images.

Incidentally, images are mainly characterized by two types of element, namely ***contour*** and ***texture***:

- Contours are abrupt change of important characteristics from an area A to an area B of a scene (average value, texture description). The edges can be locally considered as 1D signals;
- Textures are spatial variation of the 2D signal apart its local mean value;
- Motion of the objects in a scene involves temporal modifications in the successive frames of a video.

Now you have seen a general overview of images and digital image processing. In the next part of this chapter, we will look at methods for representing images (digitizing and encoding), which must be used in order to carry out digital processing. We will also see some concrete examples of results relating to the domains of application that we explored earlier.