**Fourier transformation of bitmap images (2DFT):**  it uses harmonic function for spectral decomposition. The 2D Fourier transform for the continuous image f is defined by the integral:



Parameters (u,v) are called spatial frequencies. The function f on the left-hand side of the inverse equation can be interpreted analogously to the 1D case, as a linear combination of simple periodic patterns. The real and imaginary components of the pattern are cosine and sine functions. The complex spectrum F(u, v) is a weight function which represents the influence of the elementary patterns.







**Fourier transformation of vector images (1DFT):** it transforms a function f(t) (e.g., dependent on time) into a frequency domain representation, is a frequency and is an angular frequency. The complex function F is called the (complex) frequency spectrum in which it is easy to visualize relative proportions of different frequencies. For instance, the sine wave has a simple spectrum consisting of a single spike for positive frequencies, indicating that only a single frequency is present. Let i be the usual imaginary unit. The continuous Fourier transform T is given by:





The Fourier transform always exists for digital signals (including images) as they are bounded and have a finite number of discontinuities. Attempting to understand the 1DFT equation, it is useful to express inverse Fourier transform as a Riemannian sum:



The inverse formula shows that any 1D function can be decomposed as a weighted sum (integral) of many different complex exponentials. Those exponentials can be decomposed into sines and cosines, because . The decomposition of f(t) into sines and cosines starts with some basic frequency . Other sines and cosines have frequencies obtained by multiplying by increasing natural numbers. The coefficients F() are complex numbers in general and give both magnitude and phase of the elementary waves.



**Application for text & graphic recognition:**

Application for degraded text recognition:



Characters cannot be segmented by simple thresholding, and the color, size, font, and orientation of the text are unknown. The main design choice is the kind of text occurrences, between scene text and document text.

Text is considered as a scene text when the text is recorded from a part of a scene (e.g., road signs, posters on the street, street names). Unlike document text, characters in scene images originally exist in 3D space, and can therefore be distorted by a slant or a tilt, and by the shape of objects on which they are printed. Text extraction from a natural scene has been studied, in projects such as vehicle license plates detection.

**Texture segmentation:** Text detection technique is based on a texture segmentation approach. Text in a document is considered as a textured region to isolate; non-text contents in the image, such as blanks, pictures, graphics, and other objects in the image, must be considered as regions with different textures. The human vision can quickly identify text regions without having to recognize individual characters because text has textural properties that differentiate it from the rest of a scene. Instinctively, text has the following distinguishing characteristics.

1. Characters contrast with their background.
2. Text possesses some frequencies and orientation information.
3. Text shows spatial cohesion: characters appear in clusters at a regular distance aligned to a virtual line.

**Text characterization:** By processing text as a distinctive texture, we propose a text characterization based on a bank of Gabor filters associated with an edge density measure. The features are designed to identify text paragraphs. None of them will uniquely identify text regions. Each individual feature will still confuse text with non-text regions but a collection of features will complement each other and allow identifying text unambiguously. Physically interpreted, the Gabor transform acts like the Fourier transform but only for a small Gaussian window over the image. In spatial domain, the two-dimensional Gabor filter *h*(*x*, *y*) is given by:



where *σx* and *σy* are the standard deviations of the Gaussian envelope along the *x* and *y* directions, and *wx* and *wy* are the centered frequencies of the filter. One important characteristic of Gabor filter is its orientation selectivity, which can be understood when the expression of 2D Gabor filter is rewritten in polar coordinates as:





**Text region clustering:** use of an adapted *K*-means clustering algorithm to cluster feature vectors. In order to reduce computational time, we apply the standard *K*-means clustering to a reduced number of pixels and a minimum distance classification is used to categorize all surrounding non-clustered pixels. Empirically, the number of clusters (value of *K*) was set to three, value that works well with all test images. The cluster whose center is closest to the origin of feature vector space is labeled as background while the furthest one is labeled as text. Text boxes rotation is applied after the estimation of document skew. The angle is estimated from the shape and the centroids of all text boxes. The final stage of text detection module is a validation module that confirms text boxes. It identifies false text boxes by using heuristic rules about aspect ratio, global intensity indicators, and so forth. We have applied text detection module on a set of 100 test images where there are one or two text areas per image. A text zone is correct when all the lines of the text area are included. A partly detected text zone is considered as an error. On the contrary, a false detection occurs when the detected zone does not contain any text information.

**Applications for graphic recognition:**