8.5 Vector spaces and unitary transforms

8.5.1 Introduction

An $N \times M$ digital image has NM individual pixels that can take arbitrary values. Thus it has NM degrees of freedom. Without mentioning it explicitly, we thought of an image as being composed of individual pixels. Thus, we can compose each image of *basis images* ${}^{m,n}\boldsymbol{P}$ where just one pixel has a value of one while all other pixels are zero:

$$ {}^{m,n}P_{m',n'} = \delta_{m-m'}\delta_{n-n'} = \begin{cases} 1 & \text{if} \quad m = m' \wedge n = n' \\ 0 & \text{otherwise} \end{cases} \tag{8.40} $$

Any arbitrary image can then be composed of all basis images in Eq. (8.40) by

$$ \boldsymbol{G} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} G_{m,n}\, {}^{m,n}\boldsymbol{P} \tag{8.41} $$

where $G_{m,n}$ denotes the gray value at the position $[m, n]$. The *inner product* (also known as *scalar product*) of two "vectors" in this space can be defined similarly to the scalar product for vectors and is given by

$$ (\boldsymbol{G}, \boldsymbol{H}) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} G_{m,n} H_{m,n} \tag{8.42} $$

where the parenthesis notation $(\cdot, \cdot)$ is used for the inner product in order to distinguish it from matrix multiplication. The basis images ${}^{m,n}\boldsymbol{P}$ form an *orthonormal base* for an $N \times M$-dimensional vector space. From Eq. (8.42), we can immediately derive the *orthonormality relation* for the basis images ${}^{m,n}\boldsymbol{P}$:

$$ \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} {}^{m',n'}P_{m,n}\, {}^{m'',n''}P_{m,n} = \delta_{m'-m''}\delta_{n'-n''} \tag{8.43} $$

This states that the inner product between two base images is zero if two different basis images are taken. The scalar product of a basis image with itself is one. The MN basis images thus span an $M \times N$-dimensional vector space $\mathbb{R}^{N \times M}$ over the set of real numbers.

An $M \times N$ image represents a point in the $M \times N$ vector space. If we change the coordinate system, the image remains the same but its coordinates change. This means that we just observe the same piece of information from a different point of view. All these representations are equivalent to each other and each gives a complete representation of the image. A coordinate transformation leads us from one representation to the other and back again. An important property of such a transform is that the *length* or (*magnitude*) of a vector

$$\|\boldsymbol{G}\|_2 = (\boldsymbol{G}, \boldsymbol{G})^{1/2} \tag{8.44}$$

is not changed and that orthogonal vectors remain orthogonal. Both requirements are met if the coordinate transform preserves the inner product. A transform with this property is known as a *unitary transform*.

Physicists will be reminded of the theoretical foundations of *quantum mechanics*, which are formulated in an inner product vector space of infinite dimension, the *Hilbert space*.

8.5.2 Basic properties of unitary transforms

The two most important properties of a unitary transform are [4]:

Theorem 8.2 (Unitary transform) *Let V be a finite-dimensional inner product vector space. Let $\boldsymbol{U}$ be a one-one linear transformation of V onto itself. Then*

1. *$\boldsymbol{U}$ preserves the inner product, that is, $(\boldsymbol{G}, \boldsymbol{H}) = (\boldsymbol{UG}, \boldsymbol{UH})$, $\forall \boldsymbol{G}, \boldsymbol{H} \in V$.*
2. *The inverse of $\boldsymbol{U}$, $\boldsymbol{U}^{-1}$, is the adjoin $\boldsymbol{U}^{*T}$ of $\boldsymbol{U}$: $\boldsymbol{U}\boldsymbol{U}^{*T} = \boldsymbol{I}$.*

Rotation in $\mathbb{R}^2$ or $\mathbb{R}^3$ is an example of a transform where the preservation of the length of vectors is obvious.

The product of two unitary transforms $\boldsymbol{U}_1\boldsymbol{U}_2$ is unitary. Because the identity operator $\boldsymbol{I}$ is unitary, as is the inverse of a unitary operator, the set of all unitary transforms on an inner product space is a *group* under the operation of composition. In practice, this means that we can compose/decompose complex unitary transforms of/into simpler or elementary transforms.

8.5.3 Significance of the Fourier transform (FT)

A number of unitary transforms have gained importance for digital signal processing including the cosine, sine, Hartley, slant, Haar, and Walsh transforms [5, 6, 7]. But none of these transforms matches the *Fourier transform* in importance.

The uniqueness of the Fourier transform is related to a property expressed by the *shift theorem*. If a signal is shifted in space, its Fourier transform does not change in amplitude but only in phase, that is, it is multiplied with a complex phase factor. Mathematically, this means that all base functions of the Fourier transform are *eigenvectors* of the *shift operator* $S(s)$:

$$S(s)\exp(-2\pi \mathrm{i}kx) = \exp(-2\pi \mathrm{i}ks)\exp(-2\pi \mathrm{i}kx) \tag{8.45}$$

The phase factor $\exp(-2\pi \mathrm{i}ks)$ is the *eigenvalue* and the complex exponentials $\exp(-2\pi \mathrm{i}kx)$ are the base functions of the Fourier transform spanning the infinite-dimensional vector space of the square integrable complex-valued functions over $\mathbb{R}$. For all other transforms, various base functions are mixed with each other if one base function is shifted. Therefore, the base functions of all these transforms are not an eigenvector of the shift operator.

The base functions of the Fourier space are the eigenfunctions of *all linear shift-invariant operators* or *convolution* operators. If an operator is shift-invariant, the result is the same at whichever point in space it is applied. Therefore, a periodic function such as the complex exponential is not changed in period and does not become an aperiodic function. If a convolution operator is applied to a periodic signal, only its phase and amplitude change, which can be expressed by a complex factor. This complex factor is the (wave-number dependent) eigenvalue or transfer function of the convolution operator.

At this point, it is also obvious why the Fourier transform is complex valued. For a real periodic function, that is, a pure sine or cosine function, it is not possible to formulate a shift theorem, as both functions are required to express a shift. The complex exponential $\exp(\mathrm{i}kx) = \cos kx + \mathrm{i}\sin kx$ contains both functions and a shift by a distance s can simply be expressed by the complex phase factor $\exp(\mathrm{i}ks)$.

Each base function and thus each point in the Fourier domain contains two pieces of information: the *amplitude* and the *phase*, that is, relative position, of a periodic structure. Given this composition, we ask whether the phase or the amplitude contains the more significant information on the structure in the image, or whether both are of equal importance.

In order to answer this question, we perform a simple experiment. Figure 8.9 shows two images of a street close to Heidelberg University taken at different times. Both images are Fourier transformed and then the phase and amplitude are interchanged as illustrated in Fig. 8.9c, d. The result of this interchange is surprising. It is the phase that determines the content of an image. Both images look somewhat patchy but the significant information is preserved.

From this experiment, we can conclude that the phase of the Fourier transform carries essential information about the image structure. The amplitude alone implies only *that* such a periodic structure is contained in the image but not *where*.

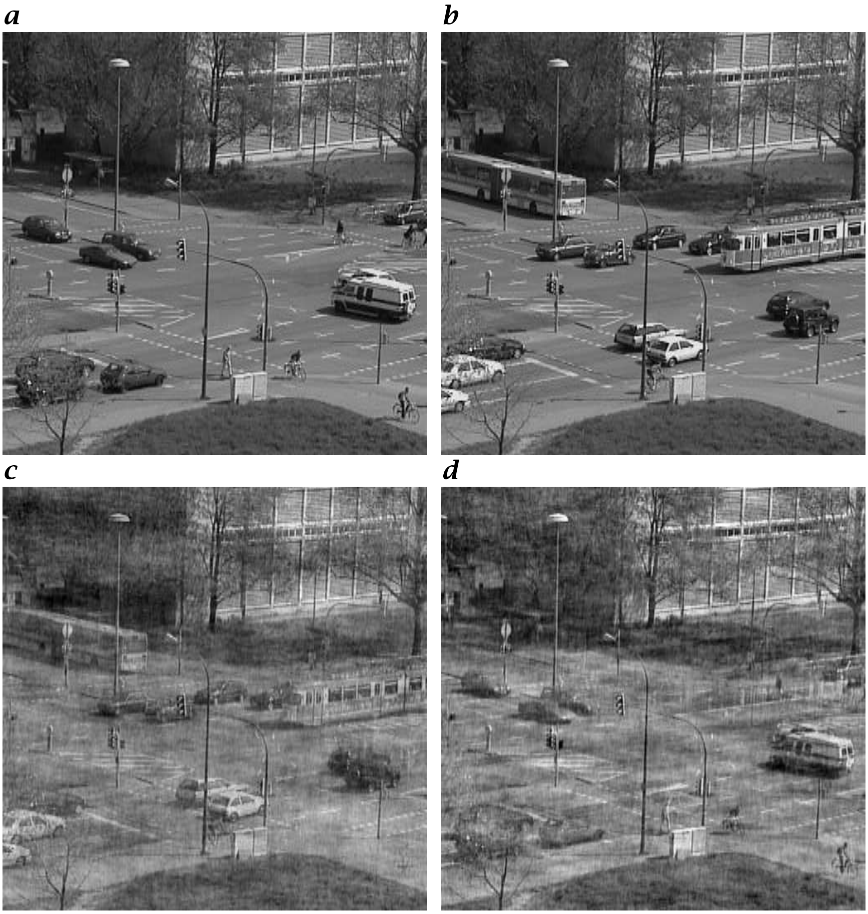

Figure 8.9: *Importance of phase and amplitude in Fourier space for the image content:* ***a, b*** *two images of a traffic scene taken at different times;* ***c*** *composite image using the phase from image* ***b*** *and the amplitude from image* ***a***; ***d*** *composite image using the phase from image* ***a*** *and the amplitude from image* ***b***.

8.5.4 Dynamical range and resolution of the FT

While in most cases it is sufficient to represent an image with rather few quantization levels, for example, 256 values or one byte per pixel, the Fourier transform of an image needs a much larger dynamical range. Typically, we observe a strong decrease of the Fourier components with the magnitude of the wave number, so that a dynamical range of at least 3–4 decades is required. Consequently, at least 16-bit integers or 32-bit floating-point numbers are necessary to represent an image in the Fourier domain without significant rounding errors.

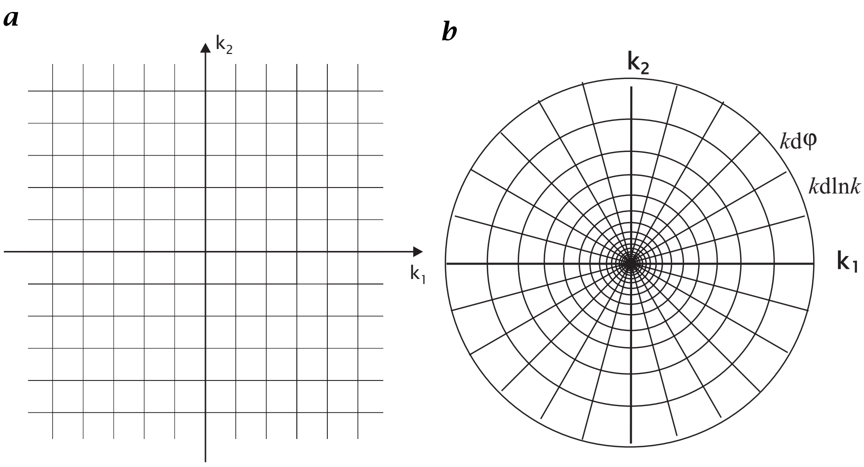

Figure 8.10: *Tessellation of the 2-D Fourier domain into:* ***a*** *Cartesian; and* ***b*** *logarithmic-polar lattices.*

The reason for this behavior is not the insignificance of high wave numbers in images. If we simply omitted them, we would blur the image. The decrease is caused by the fact that the *relative* resolution is increasing with the wave number. With the discrete Fourier transform (see Section 8.7), the Fourier transform contains only wave numbers that fit exactly integer times into the image:

$$k_{vp} = \frac{v}{d_p} \tag{8.46}$$

where $\boldsymbol{d} = [d_1, \ldots, d_D]^T$ is the size of the D-dimensional signal. Therefore, the absolute wave number resolution $\Delta k = 1/\Delta x$ is constant, equivalent to a Cartesian tessellation of the Fourier space (Fig. 8.10a). Thus the smallest wave number ($v = 1$) has a wavelength of the size of the image, and the next coarse wave number has a wavelength of half the size of the image. This is a very low resolution for large wavelengths. The smaller the wavelength, the better the resolution.

This ever increasing *relative resolution* is not natural. We can, for example, easily see the difference of 10 cm in 1 m, but not in 1 km. It is more natural to think of relative resolutions, because we are better able to distinguish relative distance differences than absolute ones. If we apply this concept to the Fourier domain, it seems to be more natural to tessellate the Fourier domain in intervals increasing with the wave number, a *log-polar coordinate system*, as illustrated in Fig. 8.10b. Such a lattice partitions the space into angular and $\ln k$ intervals. Thus, the cell area is proportional to k^2. In order to preserve the norm, or—physically speaking—the energy, of the signal in this representation, the increase in the area of the cells proportional to k^2 must be considered:

$$\int_{-\infty}^{\infty} |\hat{g}(\boldsymbol{k})|^2 \, \mathrm{d}k_1 \, \mathrm{d}k_2 = \int_{-\infty}^{\infty} k^2 |\hat{g}(\boldsymbol{k})|^2 \, \mathrm{d} \ln k \, \mathrm{d}\varphi \tag{8.47}$$

Thus, the *power spectrum* $|\hat{g}(\boldsymbol{k})|^2$ in the log-polar representation is multiplied by k^2 and falls off much less steep than in the Cartesian representation. The representation in a log-polar coordinate system allows a much better evaluation of the directions of the spatial structures and of the smaller scales. Moreover, a change in scale or orientation just causes a shift of the signal in the log-polar representation. Therefore, it has gained importance in representation object for shape analysis ([CVA3, Chapter 8]).

8.6 Continuous Fourier transform (FT)

In this section, we give a brief survey of the continuous Fourier transform and we point out the properties that are most important for signal processing. Extensive and excellent reviews of the Fourier transform are given by Bracewell [8], Poularikas [7, Chapter 2], or Madisetti and Williams [9, Chapter 1]

8.6.1 One-dimensional FT

Definition 8.1 (1-D FT) *If $g(x) : \mathbb{R} \mapsto \mathbb{C}$ is a square integrable function, that is,*

$$\int_{-\infty}^{\infty} |g(x)| \, \mathrm{d}x < \infty \tag{8.48}$$

then the Fourier transform *of $g(x)$, $\hat{g}(k)$ is given by*

$$\hat{g}(k) = \int_{-\infty}^{\infty} g(x) \exp(-2\pi \mathrm{i} k x) \, \mathrm{d}x \tag{8.49}$$

The Fourier transform maps the vector space of absolutely integrable functions onto itself. The inverse Fourier transform *of $\hat{g}(k)$ results in the original function $g(x)$:*

$$g(x) = \int_{-\infty}^{\infty} \hat{g}(k) \exp(2\pi \mathrm{i} k x) \, \mathrm{d}k \tag{8.50}$$