9 Digitization, Sampling, Quantization

9.1 Definition and Effects of Digitization

The final step of digital image formation is the *digitization*. This means sampling the gray values at a discrete set of points, which can be represented by a matrix. Sampling may already occur in the sensor that converts the collected photons into an electrical signal. In a conventional tube camera, the image is already sampled in lines, as an electron beam scans the imaging tube line by line. A CCD camera already has a matrix of discrete sensors. Each sensor is a sampling point on a 2-D grid. The standard video signal, however, is again an analog signal. Consequently, we lose the horizontal sampling, as the signal from a line of sensors is converted back to an analog signal.

At first glance, digitization of a continuous image appears to be an enormous loss of information, because a continuous function is reduced to a function on a grid of points. Therefore the crucial question arises as to which criterion we can use to ensure that the sampled points are a valid representation of the continuous image, i. e., there is no loss of information. We also want to know how and to which extent we can reconstruct a continuous image from the sampled points. We will approach these questions by first illustrating the distortions that result from improper sampling.

Intuitively, it is clear that sampling leads to a reduction in resolution, i. e., structures of about the scale of the sampling distance and finer will be lost. It might come as a surprise to know that considerable distortions occur if we sample an image that contains fine structures. Figure 9.1 shows a simple example. Digitization is simulated by overlaying a 2-D grid on the object comprising two linear grids with different grid constants. After sampling, both grids appear to have grid constants with different periodicity and direction. This kind of image distortion is called the *Moiré effect*.

The same phenomenon, called *aliasing*, is known for one-dimensional signals, especially time series. Figure 9.2 shows a signal with a sinusoidal oscillation. It is sampled with a sampling distance, which is slightly smaller than its wavelength. As a result we will observe a much larger wavelength. Whenever we digitize analog data, these problems occur. It is a general phenomenon of signal processing. In this respect, image processing is only a special case in the more general field of signal theory.

B. Jähne, Digital Image Processing
ISBN 3-540-24035-7

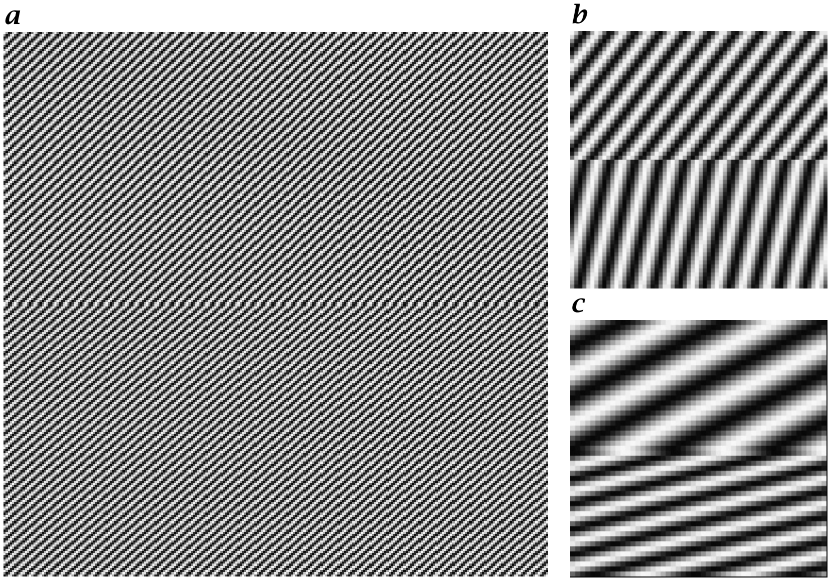

***Figure 9.1:** The Moiré effect. **a** Original image with two periodic patterns: top $\boldsymbol{k} = [0.21, 0.22]^T$, bottom $\boldsymbol{k} = [0.21, 0.24]^T$. **b** Each fourth and **c** each fifth point are sampled in each direction, respectively.*

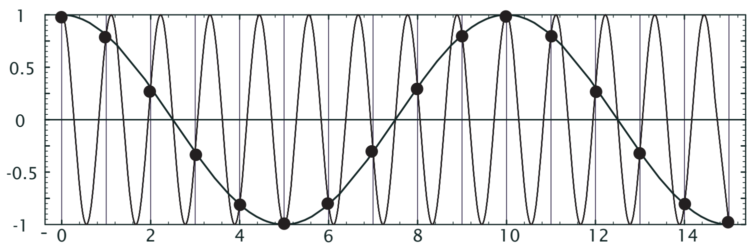

***Figure 9.2:** Demonstration of the aliasing effect: an oscillatory signal is sampled with a sampling distance Δx equal to 9/10 of the wavelength. The result is an aliased wavelength which is 10 times the sampling distance.*

Because the aliasing effect has been demonstrated with periodic signals, the key to understand and thus to avoid it is to analyze the digitization process in Fourier space. In the following, we will perform this analysis step by step. As a result, we can formulate the conditions under which the sampled points are a correct and complete representation of the continuous image in the so-called *sampling theorem*. The following considerations are not a strict mathematical proof of the sampling theorem but rather an illustrative approach.

9.2 Image Formation, Sampling, Windowing

Our starting point is an infinite, continuous image $g(\boldsymbol{x})$, which we want to map onto a matrix $\boldsymbol{G}$. In this procedure we will include the image formation process, which we discussed in Section 7.6. We can then distinguish three separate steps: image formation, sampling, and the limitation to a finite image matrix.

9.2.1 Image Formation

Digitization cannot be treated without the image formation process. The optical system, including the sensor, influences the image signal so that we should include this process.

Digitization means that we sample the image at certain points of a discrete grid, $\boldsymbol{r}_{m,n}$ (Section 2.2.3). If we restrict our considerations to rectangular grids, these points can be written according to Eq. (2.2):

$$\boldsymbol{r}_{m,n} = [m\,\Delta x_1, n\,\Delta x_2]^T \quad \text{with} \quad m, n \in \mathbb{Z}. \tag{9.1}$$

Generally, we do not collect the illumination intensity exactly at these points, but in a certain area around them. As an example, we take an ideal CCD camera, which consists of a matrix of photodiodes without any light-insensitive strips in between. We further assume that the photodiodes are equally sensitive over the whole area. Then the signal at the grid points is the integral over the area of the individual photodiodes:

$$g(\boldsymbol{r}_{m,n}) = \int\limits_{(m-1/2)\Delta x_1}^{(m+1/2)\Delta x_1} \int\limits_{(n-1/2)\Delta x_2}^{(n+1/2)\Delta x_2} g'(\boldsymbol{x})\,\mathrm{d}x_1\,\mathrm{d}x_2. \tag{9.2}$$

This operation includes *convolution* with a rectangular box function and sampling at the points of the grid. These two steps can be separated. We can perform first the continuous convolution and then the sampling. In this way we can generalize the image formation process and separate it from the sampling process.

Because convolution is an associative operation, we can combine the averaging process of the CCD sensor with the PSF of the optical system (Section 7.6.1) in a single convolution process. Therefore, we can describe the image formation process in the spatial and Fourier domain by the following operation:

$$g(\boldsymbol{x}) = \int\limits_{-\infty}^{\infty} g'(\boldsymbol{x}')h(\boldsymbol{x}-\boldsymbol{x}')\mathrm{d}^2x' \quad \circ\!\!-\!\!\bullet \quad \hat{g}(\boldsymbol{k}) = \hat{g}'(\boldsymbol{k})\hat{h}(\boldsymbol{k}), \tag{9.3}$$

where $h(\boldsymbol{x})$ and $\hat{h}(\boldsymbol{k})$ are the resulting PSF and OTF, respectively, and $g'(\boldsymbol{x})$ can be considered as the gray value image that would be obtained by a perfect sensor, i. e., an optical system (including the sensor) whose OTF is identically 1 and whose PSF is a δ-function.

Generally, the *image formation* process results in a blurring of the image; fine details are lost. In Fourier space this leads to an attenuation of high wave numbers. The resulting gray value image is said to be *band-limited*.

9.2.2 Sampling

Now we perform the *sampling*. Sampling means that all information is lost except at the grid points. Mathematically, this constitutes a multiplication of the continuous function with a function that is zero everywhere except for the grid points. This operation can be performed by multiplying the image function $g(\boldsymbol{x})$ with the sum of δ functions located at the grid points $\boldsymbol{r}_{m,n}$ Eq. (9.1). This function is called the two-dimensional δ comb, or "*bed-of-nails function*". Then sampling can be expressed as

$$g_s(\boldsymbol{x}) = g(\boldsymbol{x}) \sum_{m,n} \delta(\boldsymbol{x} - \boldsymbol{r}_{m,n}) \quad \circ\!\!-\!\!\bullet \quad \hat{g}_s(\boldsymbol{k}) = \sum_{u,v} \hat{g}(\boldsymbol{k} - \hat{\boldsymbol{r}}_{u,v}), \tag{9.4}$$

where

$$\hat{\boldsymbol{r}}_{u,v} = \begin{bmatrix} u\square k_1 \\ v\square k_2 \end{bmatrix} \quad \text{with} \quad u, v \in \mathbb{Z} \quad \text{and} \quad \square k_w = \frac{1}{\Delta x_w} \tag{9.5}$$

are the points of the so-called *reciprocal grid*, which plays a significant role in solid state physics and crystallography. According to the convolution theorem (Theorem 2.4, p. 54), multiplication of the image with the 2-D δ comb corresponds to a convolution of the Fourier transform of the image, the image spectrum, with another 2-D δ comb, whose grid constants are reciprocal to the grid constants in x space (see Eqs. (9.1) and (9.5)). A dense sampling in x space yields a wide mesh in the k space, and vice versa. Consequently, sampling results in a reproduction of the image spectrum at each grid point $\hat{\boldsymbol{r}}_{u,v}$ in the Fourier space.

9.2.3 Sampling Theorem

Now we can formulate the condition where we get no distortion of the signal by sampling, known as the *sampling theorem*. If the image spectrum is so extended that parts of it overlap with the periodically repeated copies, then the overlapping parts are alternated. We cannot distinguish whether the spectral amplitudes come from the original spectrum at the center or from one of the copies. In order to obtain no distortions, we must avoid overlapping.

A safe condition to avoid overlapping is as follows: the spectrum must be restricted to the area that extends around the central grid point

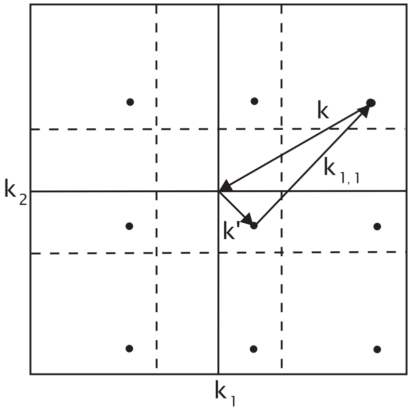

Figure 9.3: *Explanation of the Moiré effect with a periodic structure that does not meet the sampling condition.*

up to the lines parting the area between the central grid point and all other grid points. In solid state physics this zone is called the first Brillouin zone [108].

On a rectangular W-dimensional grid, this results in the simple condition that the maximum wave number at which the image spectrum is not equal to zero must be restricted to less than half of the grid constants of the reciprocal grid:

Theorem 9.1 (Sampling theorem) *If the spectrum $\hat{g}(\boldsymbol{k})$ of a continuous function $g(\boldsymbol{x})$ is band-limited, i. e.,*

$$\hat{g}(\boldsymbol{k}) = 0 \;\; \forall \; |k_w| \geq \square k_w/2, \tag{9.6}$$

then it can be reconstructed exactly from samples with a distance

$$\Delta x_w = 1/\square k_w. \tag{9.7}$$

In other words, we will obtain a periodic structure correctly only if we take at least two samples per wavelength. The maximum wave number that can be sampled without errors is called the *Nyquist* or *limiting* wave number. In the following, we will often use dimensionless wave numbers, which are scaled to the limiting wave number. We denote this scaling with a tilde:

$$\boxed{\tilde{k}_w = \frac{k_w}{\square k_w/2} = 2k_w \Delta x_w.} \tag{9.8}$$

In this scaling all the components of the wave number $\tilde{k}_w$ fall into the $]-1, 1[$ interval.

Now we can explain the Moiré and aliasing effects. We start with a periodic structure that does not meet the sampling condition. The original spectrum contains a single peak, which is marked with the long vector $\boldsymbol{k}$ in Fig. 9.3.

Because of the periodic replication of the sampled spectrum, there is exactly one peak, at $\boldsymbol{k}'$, which lies in the central cell. Figure 9.3 shows that this peak has not only another wavelength but in general another direction, as observed in Fig. 9.1.

The observed wave number $\boldsymbol{k}'$ differs from the true wave number $\boldsymbol{k}$ by a grid translation vector $\hat{\boldsymbol{r}}_{u,v}$ on the reciprocal grid. The indices u and v must be chosen to meet the condition

$$
\begin{array}{rcl}
|k_1 + u \,\square k_1| & < & \square k_1/2 \\
|k_2 + v \,\square k_2| & < & \square k_2/2.
\end{array}
\tag{9.9}
$$

According to this condition, we obtain an aliased wave number

$$
k_1' = k_1 - \square k_1 = 9/10 \,\square k_1 - \square k_1 = -1/10 \,\square k_1 \tag{9.10}
$$

for the one-dimensional example in Fig. 9.2, as we just observed.

The sampling theorem, as formulated above, is actually too strict a requirement. A sufficient and necessary condition is that the periodic replications of the image spectra must not overlap.

9.2.4 Limitation to a Finite Window

So far, the sampled image is still infinite in size. In practice, we can only work with finite image matrices. Thus the last step is the limitation of the image to a finite window size. The simplest case is the multiplication of the sampled image with a box function. More generally, we can take any *window function* $w(\boldsymbol{x})$ which is zero for sufficiently large $\boldsymbol{x}$ values:

$$
g_l(\boldsymbol{x}) = g_s(\boldsymbol{x}) \cdot w(\boldsymbol{x}) \quad \circ\!\!-\!\!\!-\!\!\bullet \quad \hat{g}_l(\boldsymbol{k}) = \hat{g}_s(\boldsymbol{k}) * \hat{w}(\boldsymbol{k}). \tag{9.11}
$$

In Fourier space, the spectrum of the sampled image will be convolved with the Fourier transform of the window function. Let us consider the example of the box window function in detail. If the window in the x space includes $M \times N$ sampling points, its size is $M\Delta x_1 \times N\Delta x_2$. The Fourier transform of the 2-D box function is the 2-D sinc function (≻ R5). The main peak of the sinc function has a half-width of $1/(M\Delta x_1) \times 1/(N\Delta x_2)$. A narrow peak in the spectrum of the image will become a 2-D sinc function. Generally, the resolution in the spectrum will be reduced to the order of the half-width of the sinc function.

In summary, sampling leads to a limitation of the wave number, while the limitation of the image size determines the wave number resolution. Thus the scales in space and wave number domains are reciprocal to each other. The resolution in the space domain determines the size in the wave number domain, and vice versa.

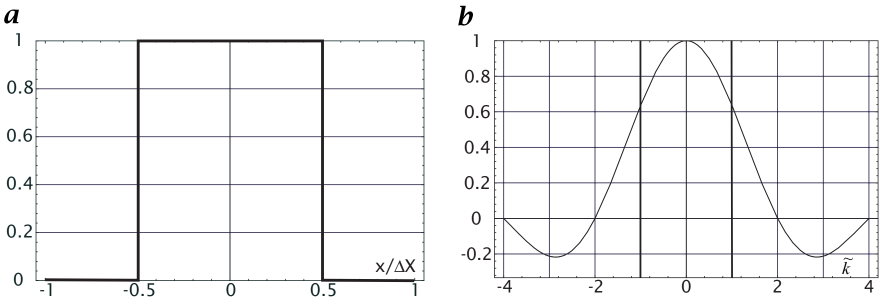

Figure 9.4: ***a*** *PSF and* ***b*** *transfer function of standard sampling.*

9.2.5 Standard Sampling

The type of sampling discussed in Section 9.2.1 using the example of the ideal CCD camera is called *standard sampling*. Here the mean value of an elementary cell is assigned to a corresponding sampling point. It is a kind of *regular* sampling, since each point in the continuous space is equally weighted. We might be tempted to assume that standard sampling conforms to the sampling theorem. Unfortunately, this is not the case (Fig. 9.4). To the Nyquist wave number, the Fourier transform of the box function is still $1/\sqrt{2}$. The first zero crossing occurs at double the Nyquist wave number. Consequently, Moiré effects will be observed with CCD cameras. The effects are even more pronounced as only a small fraction — typically 20% of the chip area for interline transfer cameras — are light sensitive [120].

Smoothing over larger areas with a box window is not of much help as the Fourier transform of the box window only decreases with k^{-1} (Fig. 9.4). The ideal window function for sampling is identical to the ideal interpolation formula Eq. (9.15) discussed in Section 9.3, as its Fourier transform is a box function with the width of the elementary cell of the reciprocal grid. However, this windowing is impracticable. A detailed discussion of interpolation can be found in Section 10.5.

9.3 Reconstruction from Samples

9.3.1 Perfect Reconstruction

The sampling theorem ensures the conditions under which we can reconstruct a continuous function from sampled points, but we still do not know how to perform the reconstruction of the continuous image from its samples, i. e., the inverse operation to sampling.

Reconstruction is performed by a suitable *interpolation* of the sampled points. Generally, the interpolated points $g_r(\boldsymbol{x})$ are calculated from the sampled values $g(\boldsymbol{r}_{m,n})$ weighted with suitable factors depending on the distance from the interpolated point:

$$g_r(\boldsymbol{x}) = \sum_{m,n} h(\boldsymbol{x} - \boldsymbol{r}_{m,n}) g_s(\boldsymbol{r}_{m,n}). \tag{9.12}$$

Using the integral properties of the δ function, we can substitute the sampled points on the right side by the continuous values:

$$\begin{aligned} g_r(\boldsymbol{x}) &= \sum_{m,n} \int_{-\infty}^{\infty} h(\boldsymbol{x} - \boldsymbol{x}') g(\boldsymbol{x}') \delta(\boldsymbol{r}_{m,n} - \boldsymbol{x}') \mathrm{d}^2 x' \\ &= \int_{-\infty}^{\infty} h(\boldsymbol{x} - \boldsymbol{x}') \sum_{m,n} \delta(\boldsymbol{r}_{m,n} - \boldsymbol{x}') g(\boldsymbol{x}') \mathrm{d}^2 x'. \end{aligned}$$

The latter integral is a convolution of the weighting function h with the product of the image function g and the 2-D δ-comb. In Fourier space, convolution is replaced by complex multiplication and vice versa:

$$\hat{g}_r(\boldsymbol{k}) = \hat{h}(\boldsymbol{k}) \sum_{u,v} \hat{g}(\boldsymbol{k} - \hat{\boldsymbol{r}}_{u,v}). \tag{9.13}$$

The interpolated function cannot be equal to the original image if the periodically repeated image spectra are overlapping. This is nothing new; it is exactly what the sampling theorem states. The interpolated image function is only equal to the original image function if the weighting function is a box function with the width of the elementary cell of the reciprocal grid. Then the effects of the sampling — all replicated and shifted spectra — are eliminated and only the original band-limited spectrum remains, and Eq. (9.13) becomes:

$$\hat{g}_r(\boldsymbol{k}) = \Pi(k_1 \Delta x_1, k_2 \Delta x_2) \hat{g}(\boldsymbol{k}). \tag{9.14}$$

Then the interpolation function is the inverse Fourier transform of the box function, a sinc function ($\succ$ R5):

$$h(\boldsymbol{x}) = \operatorname{sinc}(x_1/\Delta x_1) \operatorname{sinc}(x_2/\Delta x_2). \tag{9.15}$$

9.3.2 Oversampling

Unfortunately, this function decreases only with $1/x$ towards zero. Therefore, a correct interpolation requires a large image area; mathematically, it must be infinitely large. This condition can be weakened if we "overfill" the sampling theorem, i. e., ensure that $\hat{g}(\boldsymbol{k})$ is already zero before we reach the Nyquist wave number. According to Eq. (9.13), we can then choose $\hat{h}(\boldsymbol{k})$ arbitrarily in the region where $\hat{g}$ vanishes. We can use this freedom to construct an interpolation function that decreases more quickly in the spatial domain, i. e., has a minimum-length interpolation mask. We can also start from a given interpolation formula. Then the deviation of its Fourier transform from a box function tells us to what extent structures will be distorted as a function of the wave number. Suitable interpolation functions will be discussed in detail in Section 10.5.

The principle of *oversampling* is not only of importance for the construction of effective interpolation functions. It is also essential for the design of any type of precise filter with small filter masks (see Chapters 11 and 12). Generally, we must find a balance between the rate of oversampling, which increases the number of data points, and the requirements of the filter design. Practical experience shows that a sample rate between 3 and 6 samples per wavelength, i. e., a 1.5-3-fold oversampling is a good compromise.

9.4 *Multidimensional Sampling on Nonorthogonal Grids*

So far, sampling has only been considered for rectangular 2-D grids. Here we will see that it can easily be extended to higher dimensions and nonorthogonal grids. Two extensions are required.

First, W-dimensional grid vectors must be defined using a set of W not necessarily orthogonal basis vectors $\boldsymbol{b}_w$ that span the W-dimensional space. Then a vector on the lattice is given by

$$\boldsymbol{r_n} = [n_1\boldsymbol{b}_1, n_2\boldsymbol{b}_2, \ldots, n_W\boldsymbol{b}_W]^T \quad \text{with} \quad \boldsymbol{n} = [n_1, n_2, \ldots, n_W],\ n_w \in \mathbb{Z}. \tag{9.16}$$

In image sequences one of these coordinates is the time. Second, for some types of lattices, e. g., a triangular grid, more than one point is required. Thus for general regular lattices, P points per elementary cell must be considered. Each of the points of the elementary cell is identified by an offset vector $\boldsymbol{s}_p$.

Therefore an additional sum over all points in the elementary cell is required in the sampling integral, and Eq. (9.4) extends to

$$g_s(\boldsymbol{x}) = g(\boldsymbol{x})\sum_p\sum_{\boldsymbol{n}}\delta(\boldsymbol{x} - \boldsymbol{r_n} - \boldsymbol{s}_p). \tag{9.17}$$

In this equation, the summation ranges have been omitted.

The extended sampling theorem directly results from the Fourier transform of Eq. (9.17). In this equation the continuous signal $g(\boldsymbol{x})$ is multiplied by the sum of delta combs. According to the convolution theorem (Theorem 2.4, p. 54), this results in a convolution of the Fourier transform of the signal and the sum of the delta combs in Fourier space. The Fourier transform of a delta comb is again a delta comb (≻R5). As the convolution of a signal with a delta distribution simply replicates the function value at the zero point of the delta functions, the Fourier transform of the sampled signal is simply a sum of shifted copies of the Fourier transform of the signal:

$$\hat{g}_s(\boldsymbol{k}, \nu) = \sum_p\sum_{\boldsymbol{v}}\hat{g}(\boldsymbol{k} - \hat{\boldsymbol{r}}_{\boldsymbol{v}})\exp(-2\pi \mathrm{i}\boldsymbol{k}^T\boldsymbol{s}_p). \tag{9.18}$$

The phase factor $\exp(-2\pi i \boldsymbol{k}^T \boldsymbol{s}_p)$ results from the shift of the points in the elementary cell by $\boldsymbol{s}_p$ according to the shift theorem (Theorem 2.3, p. 54). The vectors $\hat{\boldsymbol{r}}_{\boldsymbol{v}}$

$$\hat{\boldsymbol{r}}_{\boldsymbol{v}} = v_1 \hat{\boldsymbol{b}}_1 + v_2 \hat{\boldsymbol{b}}_2 + \ldots + v_D \hat{\boldsymbol{b}}_D \quad \text{with} \quad v_d \in \mathbb{Z} \tag{9.19}$$

are the points of the *reciprocal lattice*. The fundamental translation vectors in the space and Fourier domain are related to each other by

$$\boldsymbol{b}_d^T \hat{\boldsymbol{b}}_{d'} = \delta_{d-d'}. \tag{9.20}$$

This basically means that a fundamental translation vector in the Fourier domain is perpendicular to all translation vectors in the spatial domain except for the corresponding one. Furthermore, the magnitudes of the corresponding vectors are reciprocally related to each other, as their scalar product is one. In 3-D space, the fundamental translations of the reciprocial lattice can therefore be computed by

$$\hat{\boldsymbol{b}}_d = \frac{\boldsymbol{b}_{d+1} \times \boldsymbol{b}_{d+2}}{\boldsymbol{b}_1^T (\boldsymbol{b}_2 \times \boldsymbol{b}_3)}. \tag{9.21}$$

The indices in the preceding equation are computed modulo 3, and $\boldsymbol{b}_1^T (\boldsymbol{b}_2 \times \boldsymbol{b}_3)$ is the volume of the primitive elementary cell in the spatial domain. All these equations are familiar to solid state physicists or crystallographers [108]. Mathematicians know the lattice in the Fourier domain as the *dual base* or *reciprocal base* of a vector space spanned by a nonorthogonal base. For an orthogonal base, all vectors of the dual base show into the same direction as the corresponding vectors and the magnitude is given by $\left|\hat{\boldsymbol{b}}_d\right| = 1/\left|\boldsymbol{b}_d\right|$. Then often the length of the base vectors is denoted by Δx_d, and the length of the reciprocal vectors by $\Box k_d = 1/\Delta x_d$. Thus an orthonormal base is dual to itself.

Reconstruction of the continuous signal is performed again by a suitable *interpolation* of the values at the sampled points. Now the interpolated values $g_r(\boldsymbol{x})$ are calculated from the values sampled at $\boldsymbol{r_n} + \boldsymbol{s}_p$, weighted with suitable factors that depend on the distance from the interpolated point:

$$g_r(\boldsymbol{x}) = \sum_p \sum_{\boldsymbol{n}} g_s(\boldsymbol{r_n} + \boldsymbol{s}_p) h(\boldsymbol{x} - \boldsymbol{r_n} - \boldsymbol{s}_p). \tag{9.22}$$

Using the integral property of the δ distributions, we can substitute the sampled points on the right-hand side by the continuous values and then interchange summation and integration:

$$\begin{aligned} g_r(\boldsymbol{x}) &= \sum_p \sum_{\boldsymbol{n}} \int_{-\infty}^{\infty} g(\boldsymbol{x}') h(\boldsymbol{x} - \boldsymbol{x}') \delta(\boldsymbol{r_n} + \boldsymbol{s}_p - \boldsymbol{x}') \mathrm{d}^W x' \\ &= \int_{-\infty}^{\infty} h(\boldsymbol{x} - \boldsymbol{x}') \sum_p \sum_{\boldsymbol{n}} \delta(\boldsymbol{r_n} + \boldsymbol{s}_p - \boldsymbol{x}') g(\boldsymbol{x}') \mathrm{d}^W x'. \end{aligned}$$

The latter integral is a convolution of the weighting function h with a function that is the sum of the product of the image function g with shifted δ combs. In Fourier space, convolution is replaced by complex multiplication and vice versa.

If we further consider the shift theorem and that the Fourier transform of a δ comb is again a δ comb, we finally obtain

$$\hat{g}_r(\boldsymbol{k}) = \hat{h}(\boldsymbol{k}) \sum_p \sum_{\boldsymbol{v}} \hat{g}(\boldsymbol{k} - \hat{\boldsymbol{r}}_{\boldsymbol{v}}) \exp(-2\pi \mathrm{i} \boldsymbol{k}^T \boldsymbol{s}_p). \tag{9.23}$$

The interpolated signal $\hat{g}_r$ can only be equal to the original signal $\hat{g}$ if its periodical repetitions are not overlapping. This is exactly what the sampling theorem states. The Fourier transform of the ideal interpolation function is a box function which is one within the first Brillouin zone and zero outside, eliminating all replications and leaving the original band-limited signal $\hat{g}$ unchanged.

9.5 *Quantization*

9.5.1 *Equidistant Quantization*

After digitization (Section 9.2), the pixels still show continuous gray values. For use with a computer we must map them onto a limited number Q of discrete gray values:

$$[0, \infty[\xrightarrow{Q} \{g_0, g_1, \ldots, g_{Q-1}\} = G.$$

This process is called *quantization*, and we have already discussed some aspects thereof in Section 2.2.4. In this section, we discuss the errors related to quantization. Quantization always introduces errors, as the true value g is replaced by one of the quantization levels g_q. If the quantization levels are equally spaced with a distance Δg and if all gray values are equally probable, the variance introduced by the quantization is given by

$$\sigma_q^2 = \frac{1}{\Delta g} \int_{g_q - \Delta g/2}^{g_q + \Delta g/2} (g - g_q)^2 \mathrm{d}g = \frac{1}{12} (\Delta g)^2. \tag{9.24}$$

This equation shows how we select a quantization level. We take the level g_q for which the distance from the gray value g, $|g - g_q|$, is smaller than the neighboring quantization levels q_{k-1} and q_{k+1}. The standard deviation σ_q is about 0.3 times the distance between the quantization levels Δg.

Quantization with unevenly spaced quantization levels is hard to realize in any image processing system. An easier way to yield unevenly spaced levels is to use equally spaced quantization but to transform the intensity signal, before quantization, with a nonlinear amplifier, e. g., a logarithmic amplifier. In case of a logarithmic amplifier we would obtain levels whose widths increase proportionally with the gray value.

9.5.2 *Accuracy of Quantized Gray Values*

With respect to the quantization, the question arises of the accuracy with which we can measure a gray value. At first glance, the answer to this question seems to be trivial and given by Eq. (9.24): the maximum error is half a quantization level and the mean error is about 0.3 quantization levels.

But what if we measure the value repeatedly? This could happen if we take many images of the same object or if we have an object of a constant gray value and want to measure the mean gray value of the object by averaging over many pixels.

From the laws of statistical error propagation (Section 3.3.3), we know that the error of the mean value decreases with the number of measurements according to

$$\sigma_{\text{mean}} \approx \frac{1}{\sqrt{N}}\sigma, \tag{9.25}$$

where σ is the standard deviation of the individual measurements and N the number of measurements taken. This equation tells us that if we take 100 measurements, the error of the mean should be just about 1/10 of the error of the individual measurements.

Does this law apply to our case? Yes and no — it depends, and the answer appears to be a paradox. If we measure with a perfect system, i. e., without any noise, we would always get the same quantized value and, therefore, the result could not be more accurate than the individual measurements. However, if the measurements are noisy, we would obtain different values for each measurement. The probability for the different values reflects the mean and variance of the noisy signal, and because we can measure the distribution, we can estimate both the mean and the variance.

As an example, we take a standard deviation of the noise equal to the quantization level. Then, the standard deviation of an individual measurement is about 3 times larger than the standard deviation due to the quantization. However, already with 100 measurements, the standard deviation of the mean value is only 0.1, or 3 times lower than that of the quantization.

As in images we can easily obtain many measurements by spatial averaging, there is the potential to measure mean values with standard deviations that ar much smaller than die standard deviation of quantization in Eq. (9.24).

The accuracy is also limited, however, by other, systematic errors. The most significant source is the unevenness of the quantization levels. In a real quantizer, such as an analog to digital converter, the quantization levels are not equally distant but show systematic deviations which may be up to half a quantization interval. Thus, a careful investigation of the analog to digital converter is required to estimate what really limits the accuracy of the gray value measurements.

9.6 *Exercises*

9.1: Sampling theorem

Interactive illustration of the sampling theorem (dip6ex09.01)

9.2: Standard sampling

Interactive illustration of standard sampling (dip6ex09.02)

9.3: Moiré effect

Interactive illustration of the Moiré effect with periodic signals (dip6ex09.03)

9.4: ****Discrete sampling**

What happens with the discrete Fourier transform of a 1-D signal $\boldsymbol{g}$ if you use only every second point of the signal? Try to express a discrete sampling theorem for this case and prove it. Compare it with the theorem for sampling of a continuous signal.

9.5: Quantization, noise, and averaging

Interactive demonstration of systematic and statistical errors when estimating mean values with quantized signals at different noise levels (dip6ex09.04)

9.7 *Further Readings*

Sampling theory is detailed in Poularikas [156, Section 1.6]. A detailed account on sampling of random fields — also with random distances is given by Papoulis [149, Section 11.5]. Section 9.5 discusses only quantization with even bins. Quantization with uneven bins is expounded in Rosenfeld and Kak [172].