Chapter 5

Sampling and Quantization

5.1 I - The effect of sampling and quantization

In order for a computer to process an image, this image has to be described as a series of numbers, each of finite precision. This calls for two kinds of discretisation:

1. sampling, and
2. quantisation.

By "sampling" is meant that the brightness information is only stored at a discrete number of locations. "Quantisation" indicates the discretisation of the brightness levels at these positions.

Sampling is best performed following a regular tessellation of the image. Brightness will then be integrated over the corresponding cells, yielding the sampled brightness values. For one thing, we would like these cells to cover the whole image. Moreover, such cells would best correspond to equal areas and similarly shaped regions. All cells, usually referred to as picture elements or "pixels" (sometimes also called "pels"), are then basically equivalent. For one-dimensional arrays the situation is simple: this means equidistant positions for subsequent sample points. For the two-dimensional case, these requirements lead us to the problem of covering the image with regular polygons of one type. The latter restriction is important, because out of a total of 11 possible tessellations with regular polygons (e.g. mixing squares and triangles), only 3 remain. These regular tessellations (the mixed ones are referred to as "semi-regular") are rectangular, triangular, or hexagonal, as shown in fig. 5.1 [1]. Rectangular tessellations, in particular squares, are by far the most popular in computer vision. They reflect the underlying sensor architecture and directly support operations in Cartesian coordinate frames. Nevertheless, hexagonal schemes offer some substantial advantages (e.g. no ambiguities in defining connectedness, higher degree of isotropy, etc.) and come closer to the spatial organization as found in mammalian retinas. Although other sampling schemes are available (e.g. log-polar sensor architectures), these will not be discussed here.

[1] This topic is related to the symmetry or "wallpaper" groups of the plane.

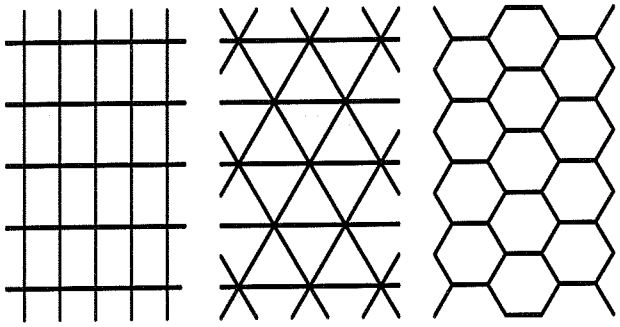

Figure 5.1: *(a) rectangular tessellation, (b) triangular tessellation, (c) hexagonal tessellation.*

As mentioned, most algorithms will be based on square tessellations. It is often important to have resolutions that are powers of 2: e.g. 16×16, $32 \times 32, \ldots$, 256×256, 512×512 images. Having powers of 2 can be useful for hardware implementations and for algorithms such as the fast Fourier transform and quadtree decompositions (see chapter 12). CCD cameras are often specially designed to yield these resolutions. Fig. 5.2 shows the effect of changing image resolution.

Similarly, quantisation will discretise brightness. Usually incoming images are discretised over 256 levels (1 byte for each pixel). Industrial vision will often reduce this information to a binary image. Objects and background are given two different gray levels, e.g. 1 for objects and 0 for background. The corresponding gray level transformation is based on simple thresholding (pixels above some threshold are "object" and those below are "background" or v.v.). It is clear that such binary image requires less storage space, although the rationale behind the operation usually has more to do with the simplicity of further processing. The technique is only feasible when there exist large contrasts between objects and their backgrounds, e.g. as a result of the backlighting technique (cf. chapter 4). An example of an image with different numbers of quantisation levels is given in fig. 5.3.

Sometimes non-uniform sampling and quantisation are considered. Regions in an image that carry little detail can be sampled coarsely, but require fine quantisation (otherwise "contouring" effects will be visible, i.e. gradual brightness variations seem to occur over sharp, clearly visible edges, see e.g. fig. 5.3's bottom right image). On the other hand, regions with fine detail require fine sampling, but quantisation can be rather coarse (e.g. the cogwheel's contour in fig. 5.3)

5.2 I - More on sampling

5.2.1 A sampling model

As we have seen, the intensity value attributed to a pixel corresponds to the integration of incoming irradiance over a cell of the tessellation. Moreover, the pixels are only located

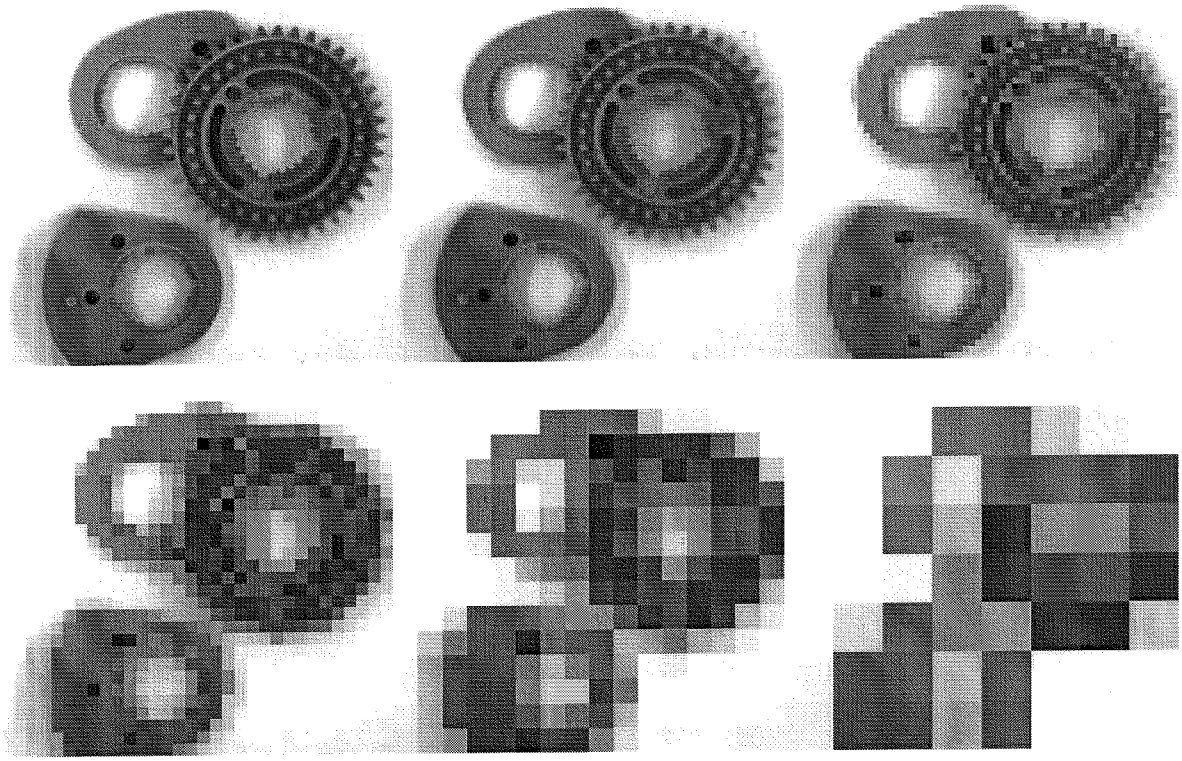

Figure 5.2: *Image at different resolutions. From top left to bottom right: 256 × 256, 128 × 128, 64 × 64, 32 × 32, 16 × 16, 8 × 8.*

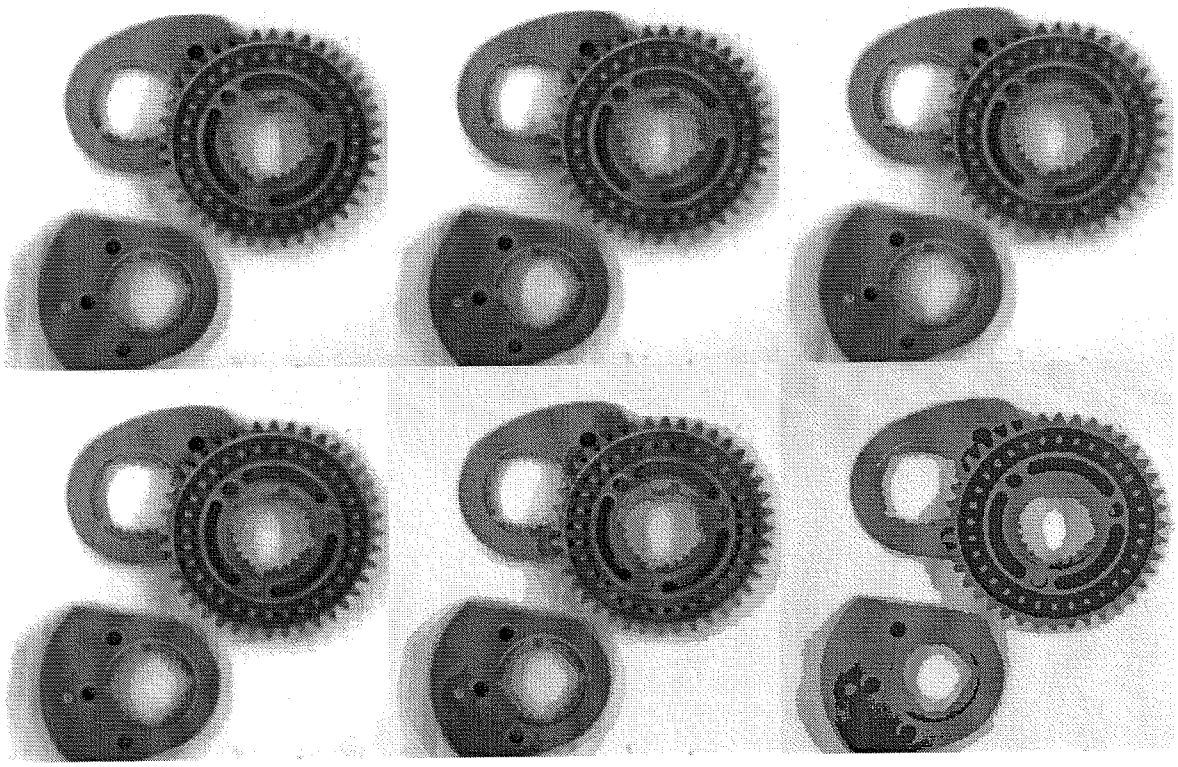

Figure 5.3: *Image with different quantization levels. From top left to bottom right: 256, 64, 32, 16, 8, 4.*

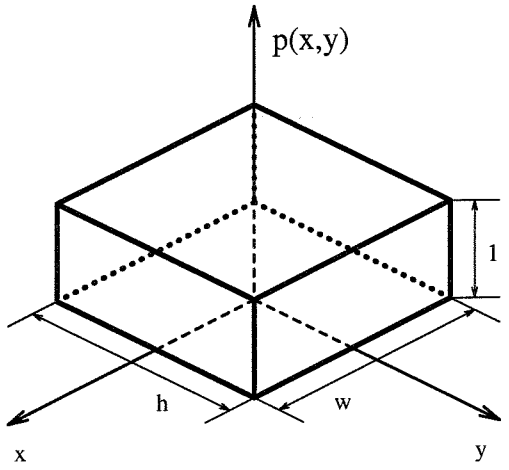

Figure 5.4: *The function $p(x, y)$ for the case of square pixels.*

at discrete positions. We can model this situation in a 2-step process:

1. integrate brightness over regions of the pixel size,
2. read out values only at the pixel positions.

Note that for the sampling schemes that are used the most (rectangular – mostly square – and to some extent hexagonal), the regions of integration for the pixels are exactly the same everywhere. This will be seen to also simplify the analysis.

Now define a neighbourhood function $p(x, y)$ that is 1 inside a region with the shape of a pixel and 0 outside that region. Fig. 5.4 shows such a function for the case of square pixels. Integrating incoming intensity $i(x, y)$ over such regions then yields the output intensity

$$o(x', y') = \int\int i(x, y)\ p(x - x', y - y')\ dxdy,$$

i.e. at each point the new value is obtained by integrating over a neighbourhood around that point. Rewriting this expression as

$$o(x', y') = \int\int i(x, y)\ p(-(x' - x), -(y' - y))\ dxdy,$$

we recognize it as the *convolution* of $i(x, y)$ with $p(-x, -y)$, in this course written as $i(x, y) * p(-x, -y)$. Note that in our example p is a symmetric function, such that we can equally well write the previous expression as $i(x, y) * p(x, y)$.

After this convolution, our next step is to select only values at the pixel positions (typically the centers of the tessellation cells). This can be modelled as a *multiplication with a two-dimensional pulse train*, i.e. multiplying the convolution output with a function consisting of Dirac impulses at the pixel positions. Without going into the details of distribution theory, we would like to remind the reader of the *sifting theorem*:

$$f(a, b) = \int\int f(x, y)\ \delta(x - a, y - b)\ dxdy.$$

This second step thus selects values at the pulse positions.

Both steps have important implications, that are easier explained when introducing the Fourier transform. For the first step, this will facilitate the interpretation of the effects induced by the integration over pixel areas. The *convolution theorem* plays a central role here, as we will see. On the other hand, the second step should also be well understood in order to avoid so-called aliasing problems and the *sampling theorem* is central to this part of the discussion. The necessary mathematical tools will be studied in the next few sections. The actual discussion of the implications for the sampling model will resume with section 5.2.5. The reader who is well acquainted with Fourier analysis in the two-dimensional case can skip the intermediate sections.

5.2.2 Linear, shift-invariant operators

The operation of convolution is not only important for the problem of image acquisition, it will also play a crucial role at several other occasions. In order to fully benefit from the convolution theorem, we therefore dwell a little bit more on this type of operation. In fact, as will now be explained, every linear, shift-invariant operation boils down to a convolution and v.v.

Consider a two-dimensional system that produces outputs $o_1(x, y)$ and $o_2(x, y)$ when given inputs $i_1(x, y)$ and $i_2(x, y)$, respectively. The system is called linear if the output $a\ o_1(x, y)\ +\ b\ o_2(x, y)$ is produced when the input is $a\ i_1(x, y)\ +\ b\ i_2(x, y)$ with a and b arbitrary numbers. Now consider a system that produces output $o(x, y)$ when given input $i(x, y)$. The system is called shift-invariant if it produces the shifted output $o(x - \alpha, y - \beta)$ when given the shifted input $i(x - \alpha, y - \beta)$, for arbitrary α and β.

Suppose a process can be modelled as a linear, shift-invariant operation $\mathcal{O}$. The concept of a point source (Dirac impulse) is very convenient in that case. Any arbitrary image i can be considered to be a sum of point sources (Dirac impulses). Then knowledge of the operation's output for a point source input could be used to determine the output for i. The output of $\mathcal{O}$ for a point source input is called the *point spread function of* $\mathcal{O}$. We denote the point spread function for our image formation operation $\mathcal{O}$ as $r(x, y)$. By assuming shift-invariance, we assume that the output to such a Dirac pulse input will always be the same, irrespective of its position, up to the translation given to the input. In terms of image acquisition, we assume that the light coming from a point (source) will be distributed over the image following a fixed spatial distribution pattern. The projection formed of a point will therefore always be deformed (blurred) in the same way, regardless where the point is shown in the image. As noted before in our discussion of optical aberrations (chapter 4), this is a simplification. Thus, the response to $\delta(x - \alpha, y - \beta)$, a shifted Dirac impulse, is assumed to be given by $r(x - \alpha, y - \beta)$. Now let us consider an arbitrary input picture $i(x, y)$. Writing $i(x, y)$ as a linear combination of point sources yields

$$i(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} i(\alpha, \beta)\ \delta(x - \alpha, y - \beta)\ d\alpha d\beta.$$

Now, for the (linear and shift-invariant) operation $\mathcal{O}$ we obtain

$$\begin{aligned} o(x,y) &= \mathcal{O}[i(x,y)] \\ &= \mathcal{O}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} i(\alpha,\beta)\ \delta(x-\alpha,y-\beta)\ d\alpha d\beta\right] \\ &= \int\int i(\alpha,\beta)\ \mathcal{O}[\delta(x-\alpha,y-\beta)]\ d\alpha d\beta \\ &= \int\int i(\alpha,\beta)\ r(x-\alpha,y-\beta)\ d\alpha d\beta \end{aligned}$$

The latter expression is the *convolution* of i and r, $i * r$. We see how the linear, shift-invariant operation $\mathcal{O}$ has led to a convolution operation. This is true in general and every linear, shift-invariant operation can be written as a convolution, just as every convolution can be checked to be a linear and shift-invariant operation.

A simple change of variables shows that the above expression can also be written as

$$o(x,y) = \int\int i(x-\alpha,y-\beta)\ r(\alpha,\beta)\ d\alpha d\beta$$

so that $i*r = r*i$, i.e. convolution is commutative. Convolution is also associative, but we will not prove this here. As we will see, the Fourier transform will make the proof almost trivial. Both characteristics of convolution can be highly practical at times. Suppose we would like to process our image by first convolving with r_1, followed by a convolution with r_2. We thus have

$$r_2 * (r_1 * i) = (r_2 * r_1) * i$$

and we see that the global operation can be interpreted as applying a single (normally larger) filter $r_2 * r_1$. In particular, the reverse analysis might be useful, i.e. if a filter can be decomposed as a convolution of two simpler filters. Efficiency can be increased by applying such smaller filters sequentially. In particular, the filter might be the convolution of two filters, one of which operates on pixels on a single row and the other on pixels of the same column. The filter is said to be *separable*. We will come back to this issue in chapter 8.

5.2.3 The two-dimensional Fourier transform

It is harder to visualize the effect of convolution than to visualize the multiplication of two functions. Because convolution in the spatial domain becomes multiplication in the frequency domain (a feature of the 1D Fourier transform which the reader is supposed to be familiar with, and which will be generalized to two dimensions), a transformation to the spatial frequency domain is attractive in the case of linear, shift-invariant operations.

Before we can explore these ideas, however, we must understand what frequency means for two-dimensional systems. For the one-dimensional case, the Fourier transform is based on a decomposition into as set of functions $e^{i2\pi ft} = \cos 2\pi ft + i \sin 2\pi ft$, that form an orthogonal basis. Similarly, for the two-dimensional case, we can decompose images into a weighted sum of two-dimensional orthogonal basis functions

$$e^{i2\pi(ux+vy)} = \cos 2\pi(ux+vy) + i \sin 2\pi(ux+vy)\ .$$

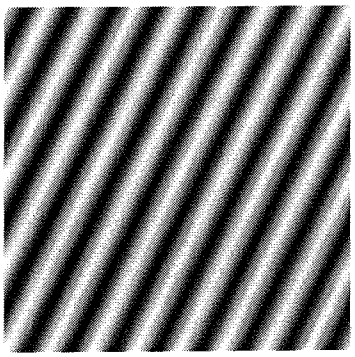

Figure 5.5: *Example of a two-dimensional (co)sine pattern.*

Note that frequency now has two components, u and v. We refer to the uv-plane as the (spatial) *frequency domain*, in contrast to the xy-plane, which is referred to as the *spatial domain*. The waveforms $\cos 2\pi(ux + vy)$ and $\sin 2\pi(ux + vy)$ correspond to waves in two dimensions. Such a pattern is shown in fig. 5.5. The maxima and minima of $\cos 2\pi(ux+vy)$ lie on parallel equidistant ridges along the lines

$$2\pi(ux + vy) = k\pi$$

for integer k. Taking a cut through the surface at right angles to these lines, that is, in the direction (u, v), gives us cosinusoidal waves with wavelength

$$\lambda = \frac{1}{\sqrt{u^2 + v^2}}.$$

The two-dimensional Fourier transform $\mathcal{F}$ will decompose a function as a weighted sum of such basis functions. The Fourier transform of a function $f(x, y)$ is given by

$$\mathcal{F}\{f(x,y)\} = F(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\; e^{-i2\pi(ux+vy)}\; dxdy.$$

$F(u, v)$ is a complex function $F_R(u, v) + iF_I(u, v)$. The magnitude

$$|F(u,v)| = \sqrt{F_R^2(u,v) + F_I^2(u,v)}$$

indicates the amplitude of the corresponding cosine patterns and $|F(u, v)|^2$ is generally referred to as the *power spectrum*. The *phase angle* is given by $\arctan(F_I(u, v)/F_R(u, v))$.

The fact that the Fourier transform of our (real) image yields a complex output might give the impression that information has somehow doubled. This of course cannot be true. In fact, for real input (such as an image) a number of important properties hold for the Fourier transform: As follows from table 5.2.3 relationships between Fourier coefficients are such that the total number of independent variables remains the same.

spatial domain	frequency domain
real	real part even imaginary part odd
real, even	real, even
real, odd	imaginary, odd

5.2.4 The convolution theorem

As we have seen, pixel intensity is derived from integrated power and this process can be modeled on the basis of convolution. As we will now show, the effect of convolution can better be assessed in the frequency domain. Let $c(x,y) = a(x,y) * b(x,y)$ and let $A(u,v)$ and $B(u,v)$ be the Fourier transforms of $a(x,y)$ and $b(x,y)$, resp.; then the Fourier transform $C(u,v)$ of $c(x,y)$ is

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} a(x-\alpha,y-\beta)\; b(\alpha,\beta)\; d\alpha d\beta\right] e^{-i2\pi(ux+vy)}\; dxdy$$

or

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} a(x-\alpha,y-\beta)\; e^{-i2\pi(ux+vy)}\; dxdy\right] b(\alpha,\beta)\; d\alpha d\beta.$$

That is,

$$C(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(u,v)\; e^{-i2\pi(u\alpha+v\beta)}\; b(\alpha,\beta)\; d\alpha d\beta = A(u,v)B(u,v).$$

This expression was obtained by changing variables, i.e. by using $x' = x-\alpha$ and $y' = y-\beta$. Convolution in the spatial domain becomes multiplication in the frequency domain. This is the *convolution theorem*. The commutativity and associativity of convolution follow directly from the corresponding properties of multiplication.

We thus have $C(u,v) = A(u,v)B(u,v)$ which means that every component basis function of $a(x,y)$ and $b(x,y)$ can be looked at independently, i.e. calculating $A(u,v)B(u,v)$ suffices to know what the result for the frequency component (u,v) will be. More in particular, if we have an input $i(x,y)$ which undergoes a linear, shift-invariant operation with point spread function $r(x,y)$, then the Fourier transform of the output is

$$O(u,v) = \mathcal{F}\{i(x,y) * r(x,y)\} = I(u,v)R(u,v).$$

$R(u,v)$ is called the *modulation transfer function*, and is the Fourier transform of the point spread function. This function is complex and specifies how each frequency component behaves under the operation, i.e. how the component's amplitude and phase are changed. Indeed, each frequency component is only changed in these respects. No new frequencies emerge. For each component $I(u,v)$ the corresponding output $O(u,v)$ clearly has the same frequency. We say that the frequency components $e^{-i2\pi(ux+vy)}$ are eigenfunctions of our

linear, shift-invariant operator and the eigenvalues are given by $R(u,v)$. Indeed, the output of the system for such eigenfunction is

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i2\pi(u(x-\alpha)+v(y-\beta))}\, r(\alpha,\beta)\, d\alpha d\beta,$$

or

$$e^{i2\pi(ux+vy)}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-i2\pi(u\alpha+v\beta)}\, r(\alpha,\beta)\, d\alpha d\beta.$$

5.2.5 Applying the convolution theorem to the sampling model

Let us return to the specific point spread function $p(x,y)$ as shown in fig. 5.4. In order to use the convolution theorem and thereby gain a deeper understanding of the sampling model's first step, we need its Fourier transform $P(u,v)$:

$$\begin{aligned}
P(u,v) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-i2\pi(ux+vy)}\, p(x,y)\, dxdy \ , \\
&= \int_{-w/2}^{w/2} e^{-i2\pi ux}dx \int_{-h/2}^{h/2} e^{-i2\pi vy}dy \ , \\
&= \left[\frac{e^{-i2\pi ux}}{-i2\pi u}\right]_{-w/2}^{w/2} \left[\frac{e^{-i2\pi vy}}{-i2\pi v}\right]_{-h/2}^{h/2} \ , \\
&= -\frac{1}{4\pi^2 uv}(-2i\sin(2\pi u\tfrac{w}{2}))(-2i\sin(2\pi v\tfrac{h}{2})) \ , \\
&= wh\left(\frac{\sin\pi wu}{\pi wu}\right)\left(\frac{\sin\pi hv}{\pi hv}\right) \ .
\end{aligned}$$

$P(u,v)$ is shown in fig. 5.6. Because $p(x,y)$ is real and even, its Fourier transform $P(u,v)$ is too. An important consequence is that our neighbourhood "filter" $p(x,y)$ won't change the phase of the frequency components, but only their amplitude. One should beware of amplitude sign changes though, since $P(u,v)$ is seen to become negative for some (u,v). In those cases, frequencies in fact undergo a complete phase reversal (shift over π). This effect is illustrated in fig. 5.7.

5.2.6 Discretising the image

As the second step in our sampling model, we select the neighbourhood filtered values only at discrete pixel positions. This is modelled as a multiplication with a two-dimensional pattern ("train") of Dirac impulses at these positions. The situation is shown in fig. 5.8's left column for a one-dimensional example [2]. In order to see what happens in the frequency domain, we need the Fourier transform of our 2D impulse train. Indeed, multiplication in the spatial domain corresponds to convolution in the frequency domain. If

$$\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\delta(x-kw,y-lh)$$

[2]Notice that the input in the figure is even (and real), so that also its Fourier transform is real (and even). As a result, we can give this example on the basis of a single curve, i.e. the real component only.

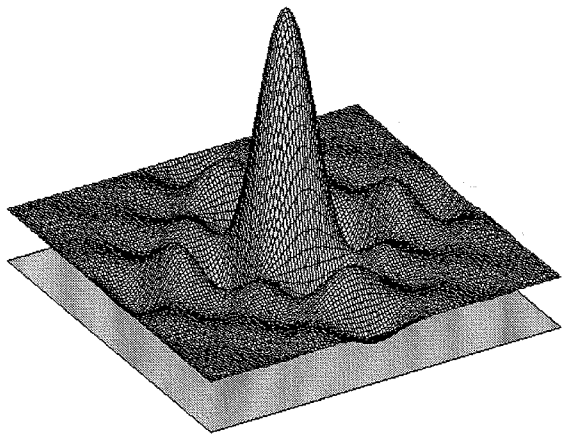

Figure 5.6: $P(u,v)$, *the Fourier transform of the neighbourhood function* $p(x,y)$ *is seen to be even. Also notice the negative values.*

is our pulse train pattern, distribution theory shows that

$$\frac{1}{wh}\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\delta(u-k\frac{1}{w},v-l\frac{1}{h}) \tag{5.1}$$

is its Fourier transform, i.e. a pulse train yields a pulse train, with inversely spaced pulses. We thus find that closer sampling in the spatial domain will further separate the pulses in the frequency domain. The importance of this observation can be seen from fig. 5.8, where the reader this time is invited to scrutinize the frequency domain. Notice that our original signal has been selected as to be bandlimited, i.e. there is a cutoff frequency beyond which no higher frequencies are present. This certainly is not the case for an arbitrary input signal. The original Fourier transform is seen to be simply repeated under the influence of the convolution that corresponds to the multiplication in the spatial domain. In fact, the spacing between the frequency domain pulses (and hence the sampling rate in the spatial domain) has been chosen so as to make the copies adjacent. It is clear that if we would sample at higher rates in the spatial domain, then these copies would move apart. On the other hand, choosing lower sampling rates would clearly lead to copies overlapping. This is bad news, because in such situation we would no longer be able to recover the true Fourier transform and hence the true image. We therefore see that the situation with contiguous copies as drawn in fig. 5.8 represents a limit case, i.e. the corresponding sampling rate is the lowest one may apply without image degradation. This sampling rate is usually referred to as the *Nyquist sampling rate*. Below the Nyquist [3] rate, the copies

[3]Harry Nyquist came to the American Telephone and Telegraph Company in 1917. In 1924 and 1928 he published his papers "Certain factors affecting telegraph speed" and "Certain topics in telegraph

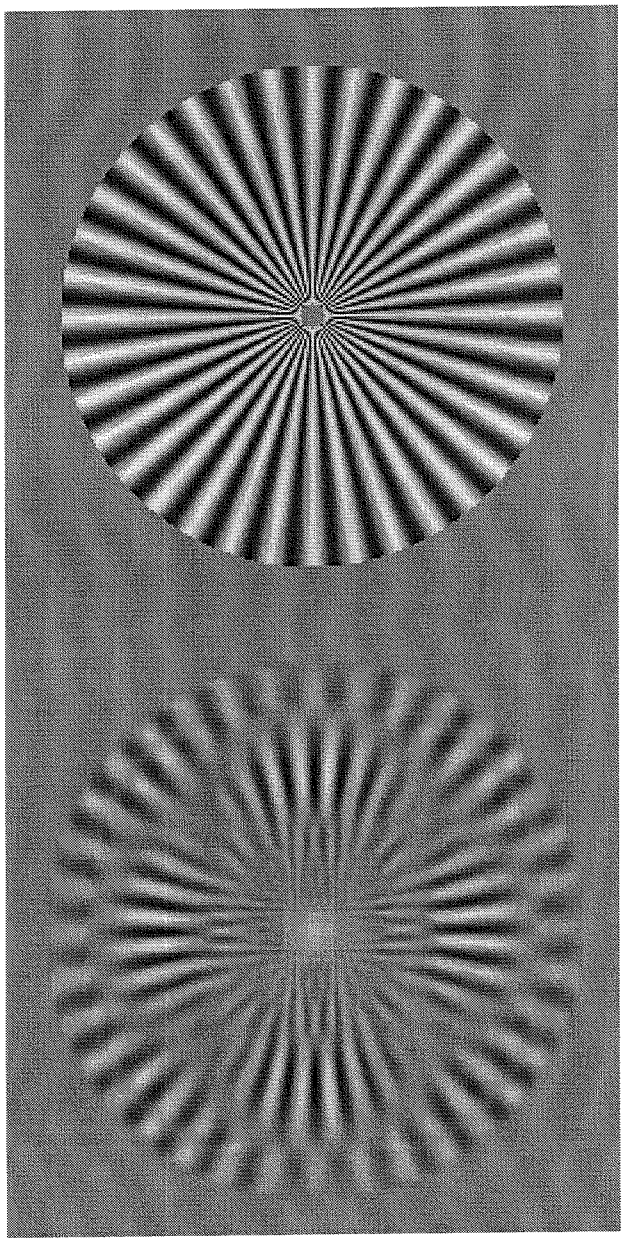

Figure 5.7: *When the top figure undergoes convolution with a square neighbourhood function (sinc function in the frequency domain), phase reversal will be observed at higher frequencies.*

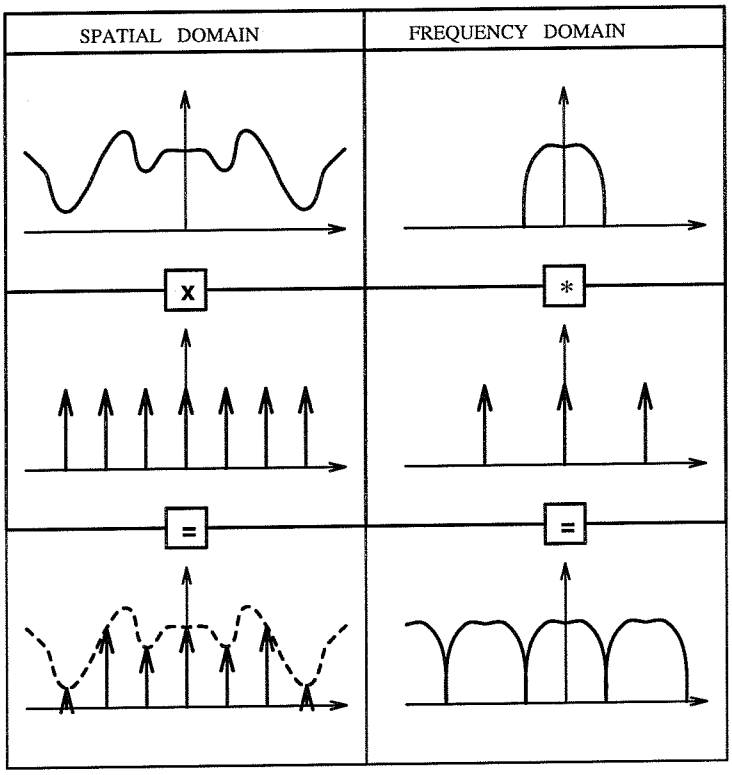

Figure 5.8: *Discretisation is modelled as a multiplication with an impulse train in the spatial domain. This operation corresponds with a convolution in the frequency domain, resulting in a periodic Fourier transform.*

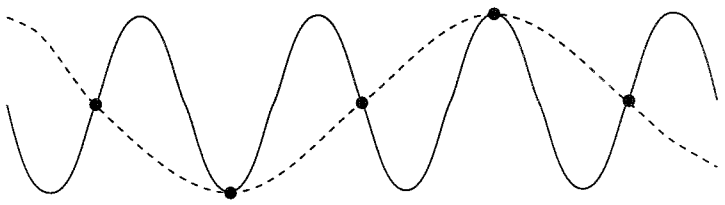

Figure 5.9: *One-dimensional example of aliasing. There are too few samples to unambiguously identify the highest frequency, since the lower frequency is consistent with all the data.*

overlap and a period of the Fourier domain will no longer be the original Fourier transform. Higher frequencies interfer (are folded back onto lower frequencies) and the resulting image degradation is called *aliasing*. If the input signal were not band-limited, aliasing would become unavoidable. Fig. 5.9 gives a one-dimensional, pictorial example of how a lack of samples might lead to misinterpretation if both frequencies lie within the input bandwith. It is only when the higher frequency lies outside this range that the interpretation is unambiguous. Before we illustrate these effects in two-dimensional images, we first dwell somewhat further on the reconstruction of images from their samples.

5.2.7 The sampling theorem

The above result is referred to as the *sampling theorem*. The sampling theorem states that if the Fourier transform of a function $f(x, y)$ is zero for all frequencies beyond u_b and v_b, i.e. if the Fourier transform is *band-limited*, then the continuous function $f(x, y)$ can be completely reconstructed from its samples as long as the sampling distances w and h along the x and y directions are such that $w < \frac{1}{2u_b}$ and $h < \frac{1}{2v_b}$.

The proof, which is kept rather sketchy, is yet another illustration of the importance of the convolution theorem. When the image is digitized, the brightness is known only at a discrete set of locations. As noted earlier, the result can be thought of as defined by a discrete grid of impulses, resulting from the multiplication of the original function $f(x, y)$ with a two-dimensional pulse train. This function, which we will write as $\tilde{f}(x, y)$, is given by

$$\tilde{f}(x, y) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} f_{kl}\, \delta(x - kw, y - lh),$$

where w and h are the horizontal and vertical sampling intervals, respectively. As noted before, the corresponding operation in the frequency domain is the convolution with the original (band-limited) transform with a pulse train. The Fourier transform can also be

transmission theory" resp. By that time telegraph and telephone signals were sent on the same wires but on different frequencies. Nyquist showed how telegraph signals could be shaped as to have no sinusoidal components of high enough frequency to be heard as interference in telephone conversations.

written as

$$\begin{aligned}\tilde{F}(u,v) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} f_{kl}\,\delta(x-kw,y-lh)\,e^{-i2\pi(ux+vy)}\,dxdy \\ &= \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} f_{kl}\,e^{-i2\pi(ukw+vlh)}.\end{aligned}$$

This is a periodic function, as predicted from fig. 5.8. The period in u is $1/w$ and that in v is $1/h$. Thus a discrete function transforms into a periodic one. This means that we can forget the part of $F(u,v)$ for $|u| > 1/2w$ and $|v| > 1/2h$ if we have sampled sufficiently densely so as to make sure that the copies of the original Fourier transform don't overlap (cf. the aliasing effect). In that case we can extract the original Fourier transform $F(u,v)$ by simply taking one period. To this end it suffices to multiply $\tilde{F}(u,v)$ with a window filter that is wh high in this region and zero outside. The factor wh is to compensate the scaling factor $1/wh$ of the pulse train in the frequency domain (cf. equation 5.1). Once we have $F(u,v)$ we clearly have all information required to completely reconstruct $f(x,y)$. Multiplication with the window function the frequency domain amounts to convolution with a two-dimensional sinc function. (Remember that in section 5.2.5 we have derived that a rectangular window and a two-dimensional sinc form a Fourier pair.) The two-dimensional sinc is the optimal interpolation function in the spatial domain:

$$\begin{aligned}f(x,y) &= \int_{u=-1/(2w)}^{1/(2w)}\int_{v=-1/(2h)}^{1/(2h)} wh\,\left(\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} f_{kl}\,e^{-i2\pi(ukw+vlh)}\right)\,e^{i2\pi(ux+vy)}\,dudv \\ &= \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} wh\,f_{kl}\int_{u=-1/(2w)}^{1/(2w)}\int_{v=-1/(2h)}^{1/(2h)} e^{i2\pi(u(x-kw)+v(y-lh))}\,dudv \\ &= \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} f_{kl}\,\frac{\sin(\pi(x/w-k))}{\pi(x/w-k)}\,\frac{\sin(\pi(y/h-l))}{\pi(y/h-l)}.\end{aligned}$$

One can easily reformulate this reconstruction scheme directly in terms of the spatial domain. The corresponding operations in the spatial domain are

1. sample by multiplication with a pulse train, and
2. convolve with the same sinc function that is the inverse Fourier transform of the window function used to select one period in the frequency domain but that can be determined directly for the selected spatial sampling rate just the same.

The reader will recognize the last equality above as the direct outcome of this alternative procedure.

The sampling theorem might be counter-intuitive. More in particular, one could wonder how discrete samples can ever give the same information as the continuous (albeit band-limited) Fourier transform. One can show, however, that a function cannot be band-limited in both the spatial and the frequency domain. Thus the number of samples in the spatial domain will be infinite.

It is useful to note that the sampling theorem can be generalized. Instead of sampling at the Nyquist rate with regular intervals, one might also use N_s regular sampling sequences with N_s times larger sampling intervals and applied with arbitrary but different phases. The result is the same number of samples, but spread less orderly. This is e.g. useful for studying sampling effects together with interlacing.

5.2.8 Aliasing

No information is lost, as long as the image contains no frequencies outside the region demarcated by the *Nyquist frequencies* $\pm\frac{1}{2w}$ and $\pm\frac{1}{2h}$. The sampling theorem makes clear that if higher frequencies do occur, information will be lost. In fact, these higher frequencies are folded back within the acceptable frequency range and cause aliasing. Fig. 5.10 shows an input image containing regions with clearly different frequency content. We see that going from the center to the boundary, the frequency increases. The figure shows that once the Nyquist rate is higher than the actual sampling, aliasing occurs. As can be seen, the higher frequencies are replaced by lower frequencies and false structures emerge. The emergence of such false structure can be reduced by adding some Poisson noise on the sample point positions. It has been argued that deviations from perfect hexagonal sampling in the human retina might serve the purpose of replacing false structure by noise this way. This modification is beyond the scope of this introduction and quite unfeasible in most applications, however.

As a second example, we will look at a sampled version of the function

$$f(x,y) = \cos 2\pi(u_0 x + v_0 y)$$

which has as its Fourier transform two Dirac impulses

$$F(u,v) = \frac{1}{2}[\delta(u-u_0, v-v_0) + \delta(u+u_0, v+v_0)].$$

Fig. 5.11 shows the reconstruction of this rolling pattern for a sampling rate that is well above the Nyquist rate. Fig. 5.12 shows a similar reconstruction, this time with a sampling rate below the Nyquist rate. The reconstruction is no longer faithful. It is interesting to note that not only has the frequency changed, but even the global orientation of the pattern is not reproduced. In order to understand how this is possible, fig. 5.13 shows schematic drawings corresponding to the two given situations. As can be seen for the second case, a false lower frequency with a wrong orientation is introduced indeed.

5.2.9 The discrete Fourier transform

Analogous to sampling the spatial domain, one could consider sampling the frequency domain. As a matter of fact, one will have to if the frequency domain is to be analysed by computer. Also here a similar sampling theorem holds. If the function $f(x,y)$ is zero outside the interval $(-x_b, x_b)$ for x and $(-y_b, y_b)$ for y, then its Fourier domain can be completely reconstructed from samples not further apart than $1/2x_b$ for the u direction and $1/2y_b$ for the v direction:

$$F(u,v) = \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} F(\frac{k}{2x_b}, \frac{l}{2y_b}) \frac{\sin(2\pi x_b(u-k/2x_b))}{\pi 2x_b(u-k/2x_b)} \frac{\sin(2\pi y_b(v-l/2y_b))}{\pi 2y_b(v-l/2y_b)}.$$

Discrete versions of both the spatial and the frequency domain can be considered together according to what is called the discrete Fourier transform. Sometimes, this transform

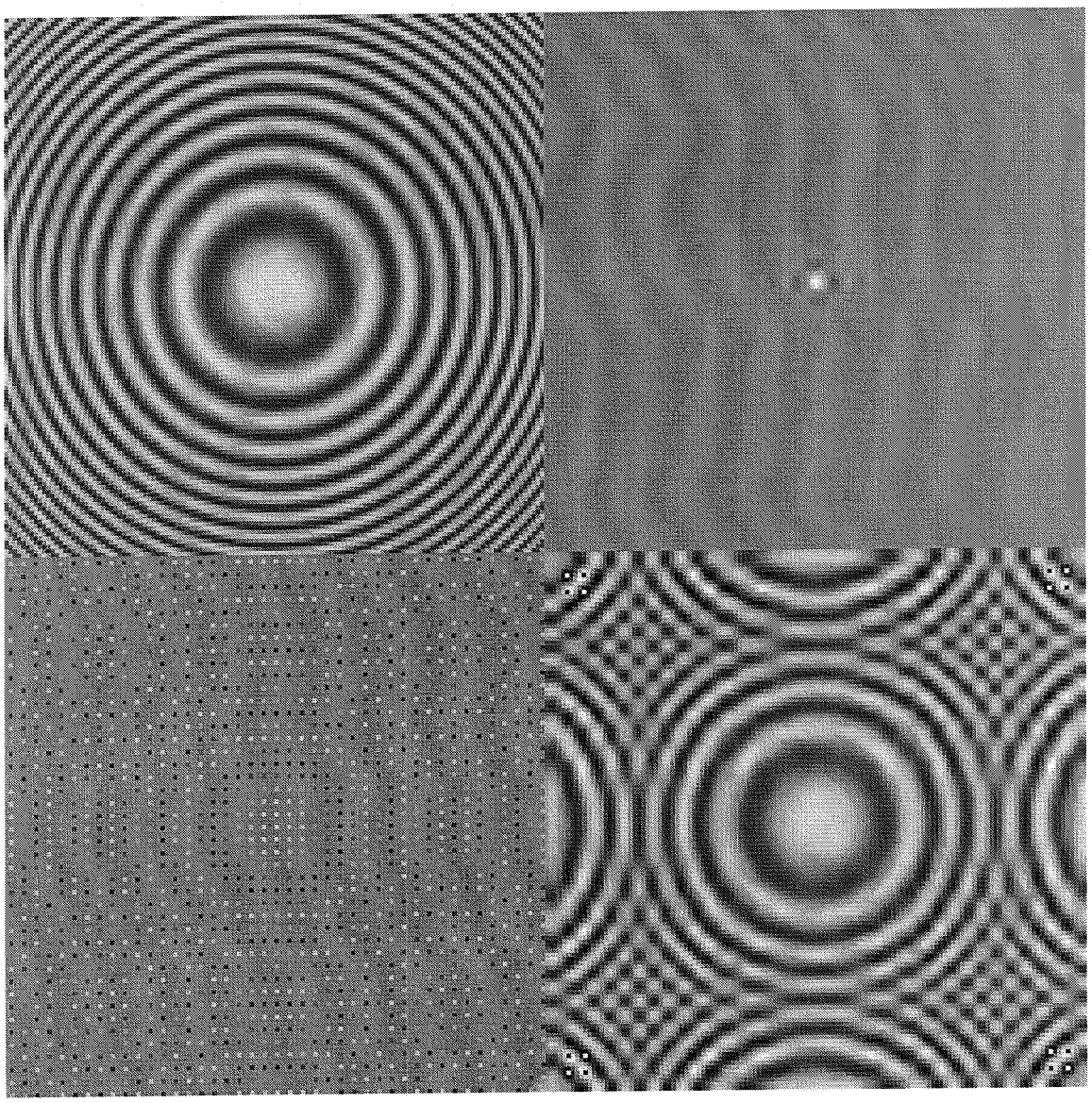

Figure 5.10: *Grey as in most part of the top right figure stands for zero values. Brighter means positive values, darker means negative values. Top left: concentric circles generating pattern of increasing frequencies when going to the boundaries; Top right: optimal sinc function for reconstruction (interpolation) according to the sampling rate indicated in the bottom left figure; Bottom right: result of convolving the samples of the bottom left figure with the sinc shown top right. Notice the aliasing effect at the higher frequencies.*

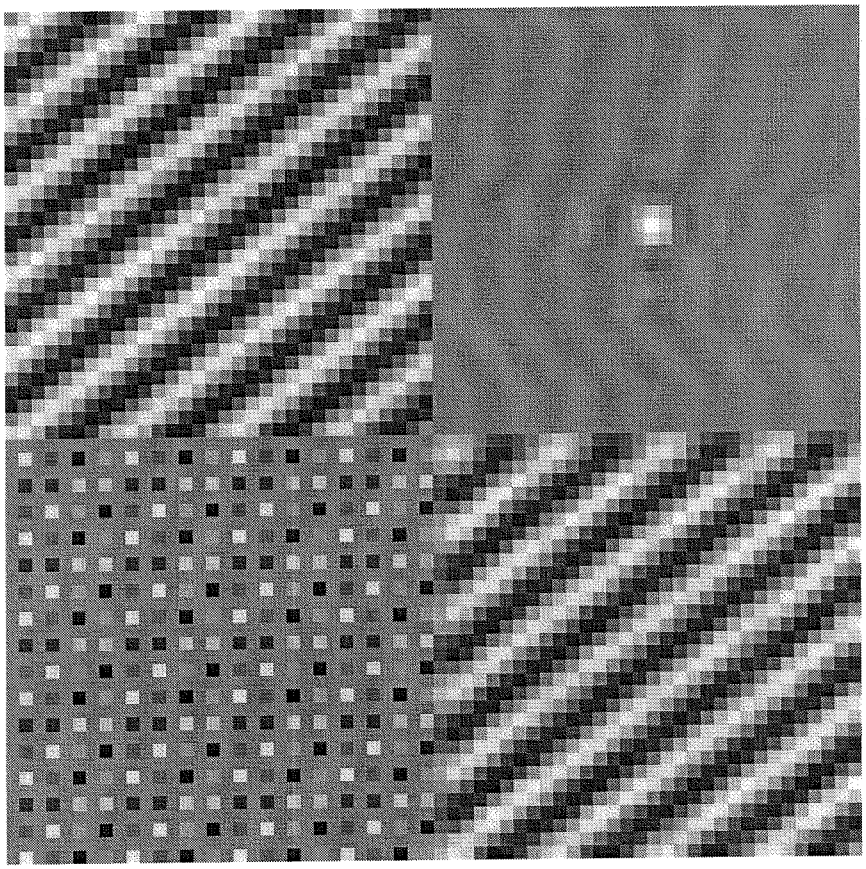

Figure 5.11: *Rolling pattern in the upper left corner is reconstructed on the basis of the sampling points in the left bottom image and the corresponding sinc function in the upper right. Because sampling is well above the Nyquist rate, the reconstruction is perfect (except for boundary effects), as shown in the lower right picture.*

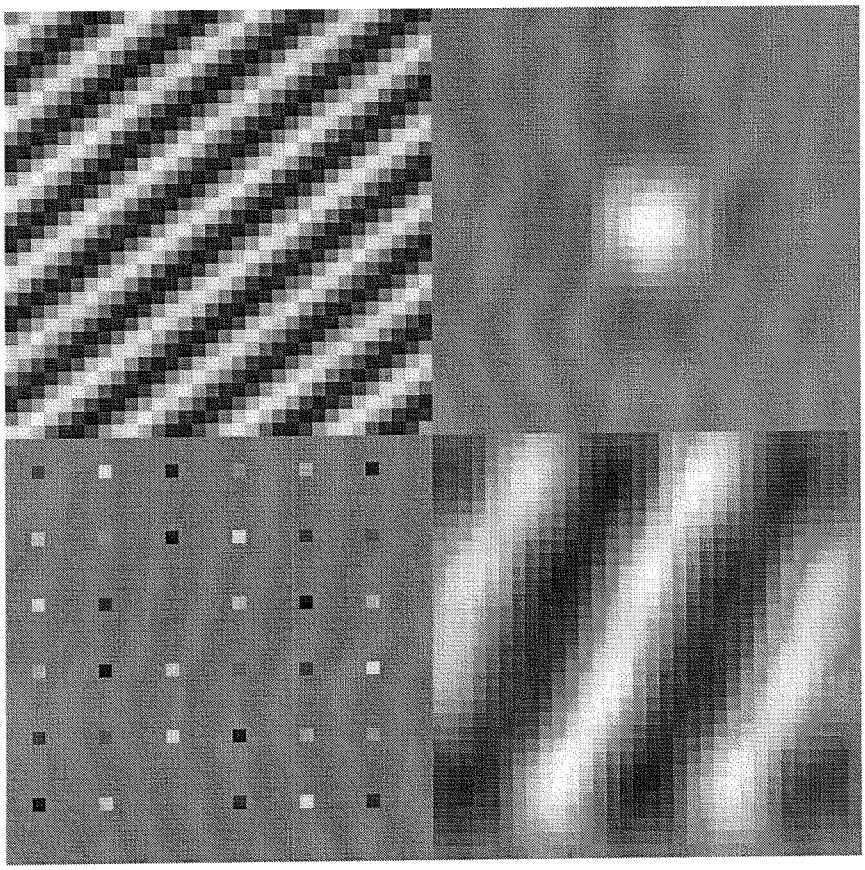

Figure 5.12: *Rolling pattern in the upper left corner is reconstructed on the basis of the sampling points in the left bottom image and the corresponding sinc function in the upper right. Because sampling is below the Nyquist rate, the reconstruction fails.*

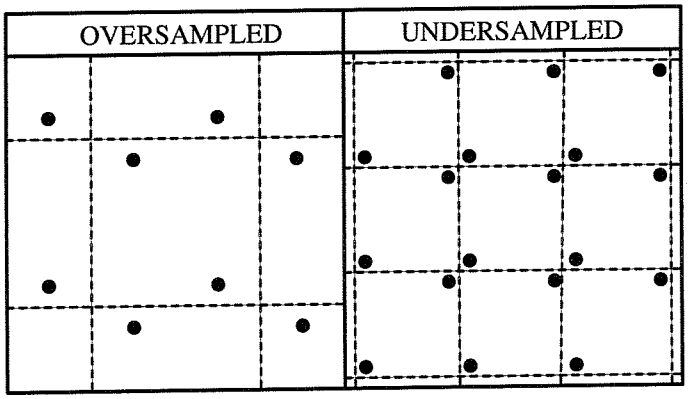

Figure 5.13: *Schematic representations of the Fourier power spectra of a sampled cosine pattern. At the left, the resulting period in the frequency domain is a faithful representation of the power spectrum of the continuous cosine. At the right, spatial sampling is too coarse and the period of the Fourier spectrum contains a wrong cosine component.*

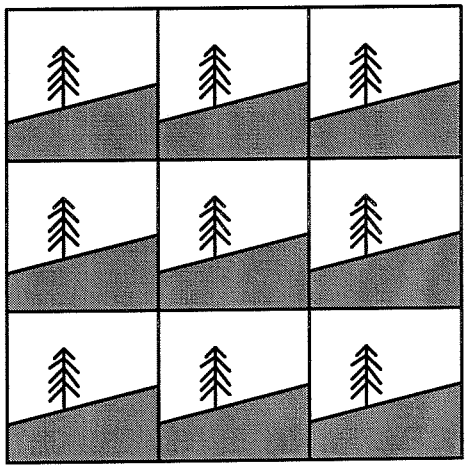

Figure 5.14: *The periodically extended image shows sharp discontinuities that introduce high frequencies in the frequency domain.*

pair is defined directly, but it can be shown to yield exactly the values of the continuous transform pair at the sample positions. The pair is given by

$$F(k,l) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m,n)\, e^{-2\pi i\left(\frac{km}{M}+\frac{ln}{N}\right)} ,$$

$$f(m,n) = \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} F(k,l)\, e^{2\pi i\left(\frac{mk}{M}+\frac{nl}{N}\right)} .$$

Substantial research has gone into the development of efficient implementations for this transform (FFT, Primfactor, Winograd, split-radix, etc.). These methods are not part of the course. The reader should nevertheless be aware of the fact that these efficient implementations of the Fourier transform exist. The computational cost for an $N \times N$ image by using the FFT instead of a straightforward implementation is proportional with $N \log_2 N$ multiply/add operations instead of N^2. The gain in efficiency is dramatic for large N.

As a final remark on the discrete Fourier transform, we have to put out a warning. As already noted, sampling in one domain implies periodicity in the other. Here, both domains have been discretised and thus should both the original image and its Fourier transform be interpreted as periods of periodic signals. The discrete Fourier transform is therefore not the Fourier transform of the image as such, but rather of the periodic signal created by repeating the image data both horizontally and vertically. In order to see the bad news here, assume we have taken a picture of an outdoor scene. The sky will probably be brighter than the bottom part of the image, showing part of the landscape. The situation is drawn in fig. 5.14. Now, building our periodic signal we will append copies above and below the picture. Sky parts will therefore border landscape parts, leading to sharp edges. These edges will introduce high frequencies in the Fourier spectrum. These frequencies are not inherent to the image data, however. A way to overcome this caveat is to build a larger image from the original one (the new image is four times as large as the original one) that

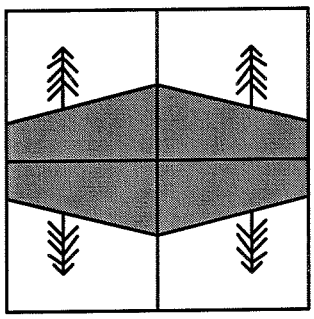

Figure 5.15: *When the image is repeated four times as shown, the periodic extension of the resulting larger image will be free of sharp discontinuities that do not find their origin in the original data.*

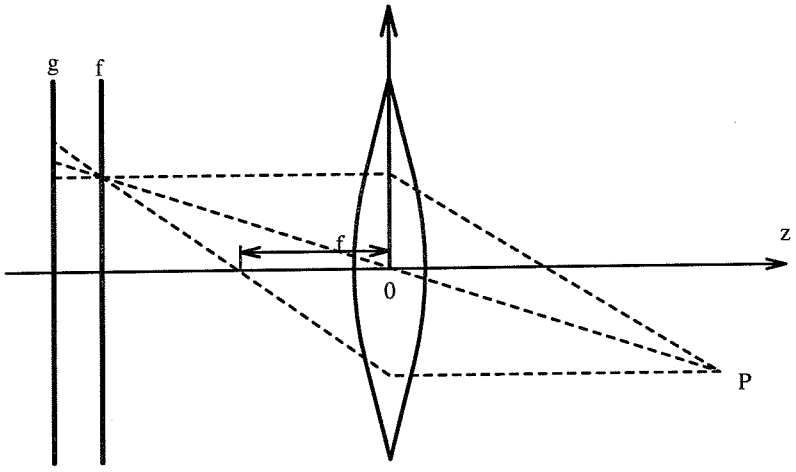

Figure 5.16: *An out-of-focus image $g(x, y)$ can be thought of as a processed version of the ideal image $f(x, y)$.*

no longer yields discontinuities when periodically extended. The procedure is shown in fig. 5.15.

5.2.10 A short note on degradations and deformations

*** Degradations**

We have discussed convolution and the convolution theorem because of the sampling model. As already suggested, convolution is a highly useful tool to model several other image operations, however. Indeed, if e.g. a degradation can be approximated as being linear and shift-invariant then there must be a convolution 'kernel' to model its effect. Consider an out-of-focus image system (fig. 5.16). We can think of the image $g(x, y)$ produced by the defocused system as the result of a convolution of the ideal (sharp) image $f(x, y)$ with some blurring filter. The linearity and shift-invariance should in principle be assessed carefully. Some simple test can at least give a first impression. If the lighting is changed so as to

double the brightness of the ideal image, the brightness of the out-of-focus image should also be doubled. It is clear that any real system will show saturation effects that limit its linearity. Further, if the imaging system is moved slightly, so that the ideal image is shifted in the image plane, the out-of-focus image should similarly be shifted. In practice, images are limited in area, so that shift invariance only holds for limited displacements. Moreover, aberrations in optical imaging systems vary with the distance from the the optical axis; such systems are therefore only approximately shift-invariant.

In fact, the filter we have used in our sampling model, i.e. a rectangular window, is primarily a blurring filter, in that it has a dominant low-pass lobe, sided by much smaller band-pass lobes. In half of these band-pass lobes phase reversal did occur. A blurring filter that doesn't introduce such phase reversal is the Gaussian, i.e. a convolution kernel with the profile of a two-dimensional Gaussian function. Gaussian blurring is often a good model for out-of-focus systems. A Gaussian filter $g(x, y)$ with standard deviation σ is given by

$$g(x, y) = e^{-\frac{r^2}{2\sigma^2}}.$$

where $r = \sqrt{x^2 + y^2}$. The Fourier transform $G(u, v)$ is found to be

$$G(u, v) = 2\pi\sigma^2 e^{-\frac{\rho^2}{2\alpha^2}}.$$

where $\rho = \sqrt{u^2 + v^2}$ and $\alpha = 1/(2\pi\sigma)$. We see that $G(u, v)$ also is a Gaussian, but its standard deviation is inversely proportional to that in the spatial domain. A second interesting characteristic of the Gaussian is that it minimizes joint spatial domain and frequency domain uncertainty. Define

$$x_{av} = \frac{\int_{-\infty}^{\infty} x f \bar{f} dx}{\int_{-\infty}^{\infty} f \bar{f} dx} \qquad (\Delta x)^2 = \frac{\int_{-\infty}^{\infty} (x - x_{av})^2 f \bar{f} dx}{\int_{-\infty}^{\infty} f \bar{f} dx}$$

$$u_{av} = \frac{\int_{-\infty}^{\infty} u F \bar{F} du}{\int_{-\infty}^{\infty} F \bar{F} du} \qquad (\Delta u)^2 = \frac{\int_{-\infty}^{\infty} (u - u_{av})^2 F \bar{F} du}{\int_{-\infty}^{\infty} F \bar{F} du}$$

with $f = f(x, y)$ the original function and $F = F(u, v)$ its Fourier transform as usual. Bars indicate complex conjugate values. The *uncertainty principle* states that $\Delta x \Delta u \geq 1/4\pi$. A similar relation holds for y and v. Now, it can be shown that Gaussian filters minimize the uncertainty product, i.e. $\Delta x \Delta u = 1/4\pi$ (as do the more general class of Gabor filters, i.e. cosine modulated Gaussians). These filters therefore realise the best possible simultaneous localisation in the spatial ánd the frequency domain.

Deformations

Before ending this part on sampling, we succinctly describe the effects deformations in the spatial domain have on the Fourier transform. Suppose we deform our original image by

applying an affine transformation, characterised by the matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

How does the Fourier transform of our image change ?

$$\mathcal{F}\{f(ax+by, cx+dy)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(ax+by, cx+dy)e^{-i2\pi(ux+vy)}dxdy.$$

By changing variables x, y to x', y' with

$$\begin{aligned} x' &= ax+by \\ y' &= cx+dy \end{aligned}$$

and applying the inverse

$$\begin{aligned} x &= (dx'-by')/|A| \\ y &= (-cx'+ay')/|A| \end{aligned}$$

We then have

$$\mathcal{F}\{f(ax+by, cx+dy)\} = \int\int f(x', y')\, e^{-i2\pi((du-cv)\frac{x'}{|A|}+(-bu+av)\frac{y'}{|A|})}\, \frac{1}{|A|}\, dx'dy'.$$

We may conclude that

$$\mathcal{F}\{f(ax+by, cx+dy)\} = \frac{1}{ad-bc}F(\frac{du-cv}{ad-bc}, \frac{-bu+av}{ad-bc}).$$

It is interesting to consider a number of special cases. First, take a simple rotation, which is equiaffine (i.e. area preserving) such that $ad-bc=1$. We find

$$\mathcal{F}\{f(\cos\theta x - \sin\theta y, \sin\theta x + \cos\theta y)\} = F(\cos\theta u - \sin\theta v, \sin\theta u + \cos\theta v)$$

where θ is the rotation angle. We conclude that the Fourier transform undergoes the same rotation. Similarly, scaling along x and y with factors k and l, respectively, yields

$$\mathcal{F}\{f(kx, ly)\} = \frac{1}{kl}F(\frac{u}{k}, \frac{v}{l}).$$

The scaling is reversed along both directions.

These effects are also illustrated in figure 5.18.

5.3 * I - More on quantization

5.3.1 Lloyd-Max quantization

In the first section of this chapter, we have already considered the effect of using more or less quantization levels, here we will concern ourselves with optimal placement of these levels.

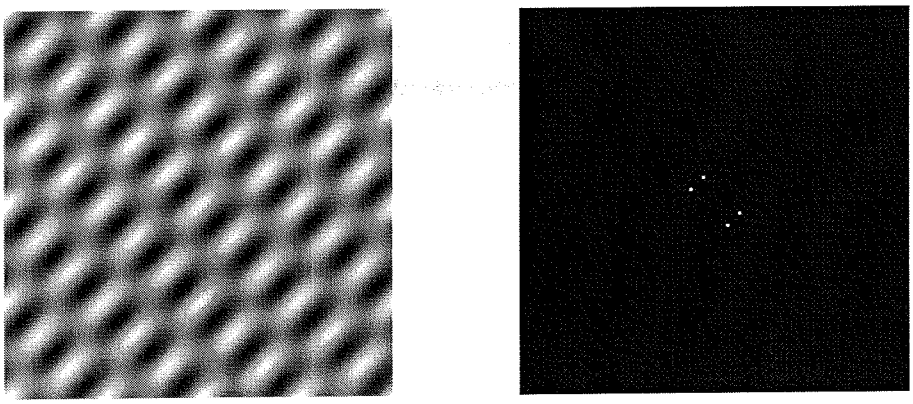

Figure 5.17: *Test pattern consisting of two superimposed cosine patterns (left) and its Fourier spectrum (right).*

Suppose we create K intervals in the range of possible intensities, defined by the decision levels $z_1, z_2, \cdots, z_{K+1}$. To all intensities in the interval $[z_k, z_{k+1}]$ we assign the new gray level q_k, i.e. the corresponding output of the quantizer. Let δ denote the mean-square quantization error between the input and the output of the quantizer for a given choice of the output levels q and the interval boundaries z. Thus

$$\delta = \sum_{k=1}^{K} \int_{z_k}^{z_{k+1}} (z - q_k)^2 p(z) dz := \sum_{k=1}^{K} \delta_k$$

where $p(z)$ is the probability density function for the input sample value which is represented by the continuous variable z. For a given number K of output levels, we would now like to determine the output levels q and interval boundaries z minimizing δ. Demanding that the partial derivatives of δ with respect to the q's and z's vanish yields:

$$\begin{aligned} \frac{\partial \delta}{\partial z_k} &= (z_k - q_{k-1})^2 p(z_k) - (z_k - q_k)^2 p(z_k) = 0, \quad k = 2, 3, \cdots, K \\ \frac{\partial \delta}{\partial q_k} &= -2 \textstyle\int_{z_k}^{z_{k+1}} (z - q_k) p(z) dz = 0, \qquad\qquad k = 1, 2, \cdots, K \end{aligned}$$

Note that we have not differentiated with respect to z_1 and z_{K+1}, as these correspond to the fixed intensity interval boundaries. For $p(z_k)$ not equal to zero we obtain

$$\begin{aligned} z_k &= (q_{k-1} + q_k)/2, \qquad\qquad k = 2, 3, \cdots, K \\ q_k &= \textstyle\int_{z_k}^{z_{k+1}} z p(z) dz / \int_{z_k}^{z_{k+1}} p(z) dz, \quad k = 1, 2, \cdots, K \end{aligned}$$

These equations are sometimes referred to as the "Lloyd-Max quantizer equations". We see that for an optimum quantizer the decision levels z_k's are located halfway between the output levels q_k, while each q_k is the centroid of the portion of $p(z)$ between z_k and z_{k+1}.

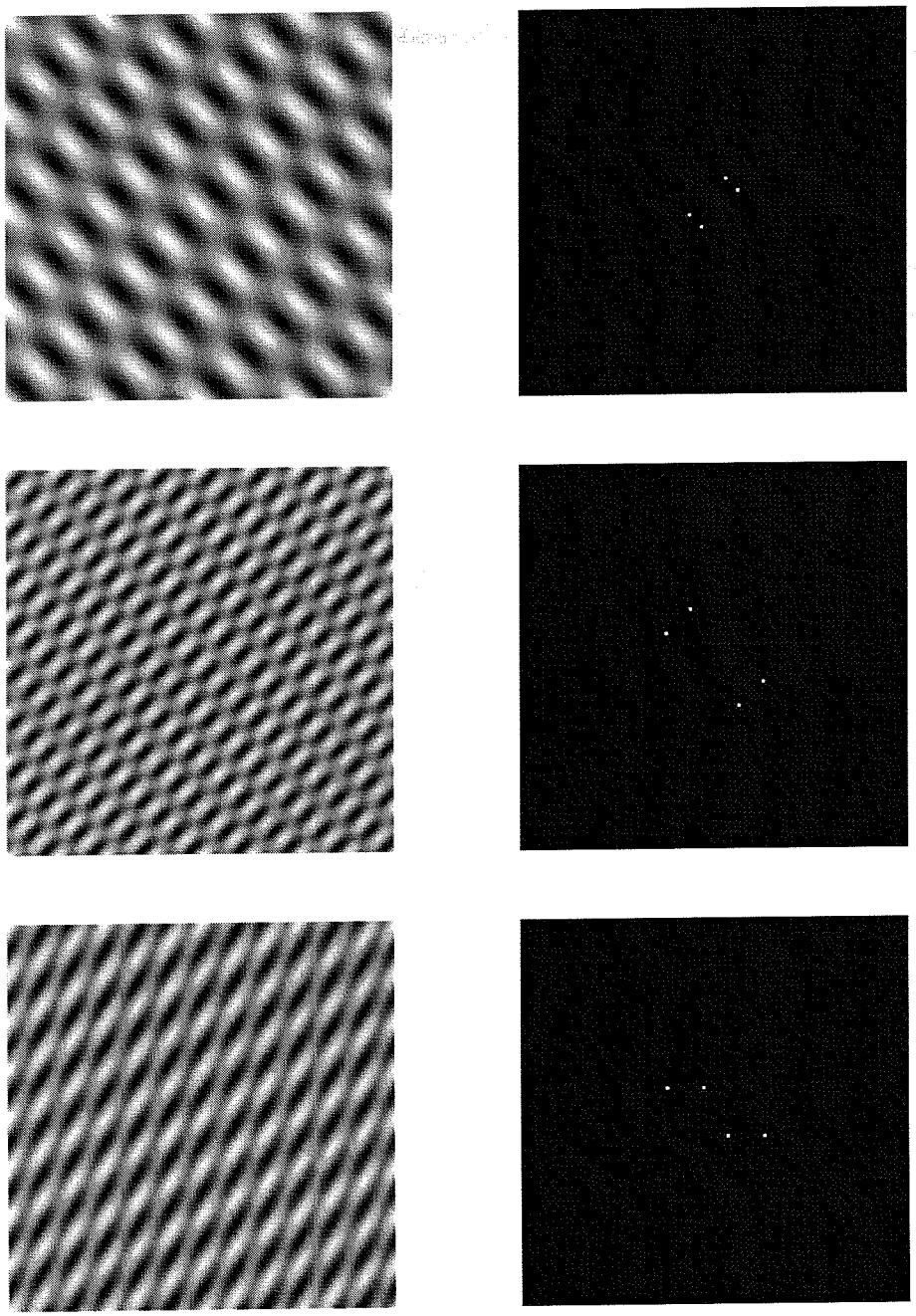

Figure 5.18: *Top: if the test pattern is rotated, its Fourier spectrum undergoes the same rotation. Middle: If the test pattern is scaled, inverse scaling is applied to the Fourier spectrum. Bottom: The test pattern is deformed by an affine transformation. This yields skewing in the frequency domain also.*



Figure 5.19: *TV presenter image, 256 grey levels.*

If the sample values occur equally frequently, then the optimal quantizer will spread the values z_k and q_k uniformly:

$$\begin{aligned} z_k &= (q_{k-1} + q_k)/2, \quad k = 2, 3, \cdots, K \\ q_k &= (z_k + z_{k+1})/2, \quad k = 1, 2, \cdots, K \end{aligned}$$

If the sample values occur more frequently in some part of the intensity range than in others, this "uniform quantization" will in general no longer be optimal. In that case, we have to solve the nonlinear, simultaneous Lloyd-Max equations by some iterative scheme.

An example of Lloyd-Max quantisation is given for fig. 5.19. Uniform quantisation (8 grey levels) is demonstrated in fig. 5.20. The original probability distribution of the grey levels (i.e. the *histogram*) with the uniform quantisation boundaries z_k superimposed are shown in fig. 5.21. Lloyd-Max quantisation gives the result of fig. 5.22 with the quantisation boundaries shown in fig. 5.23 As can be seen from this example, improvement can be disputed. Whereas the background is now more homogeneous, just as in the original image, the face shows some increased contouring. The latter part of the image is the more interesting one, however. Lloyd-Max quantisation doesn't take local image structure or interpretation into account.



Figure 5.20: *TV presenter image, requantised uniformly with 11 grey levels.*

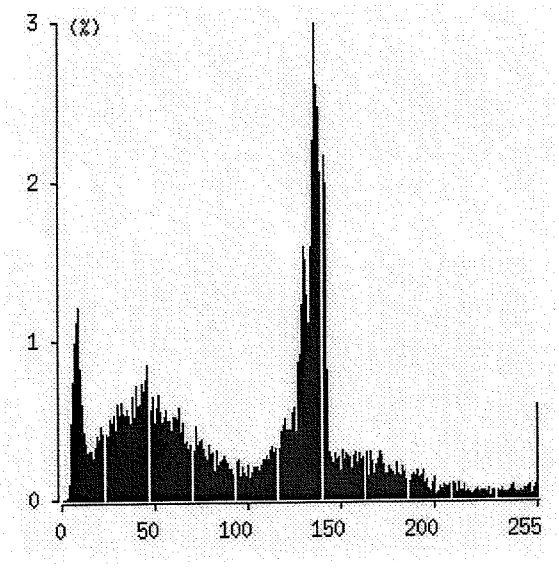

Figure 5.21: *Uniform quantisation boundaries with 11 grey levels.*



Figure 5.22: *TV presenter image, requantised following the Lloyd-Max equations with 11 grey levels.*

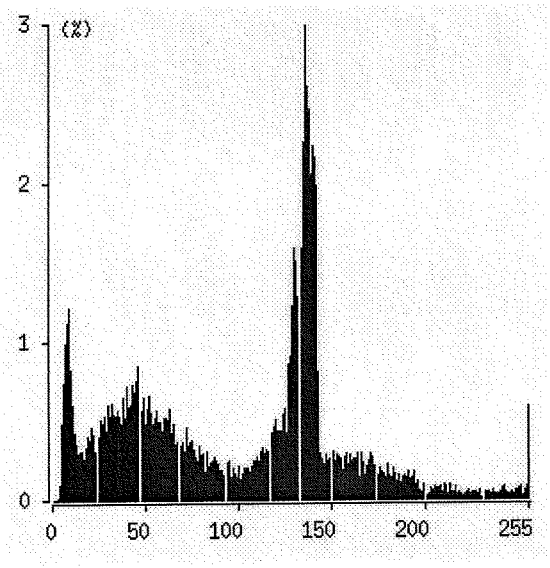

Figure 5.23: *Lloyd-Max quantisation boundaries with 11 grey levels superimposed on the original intensity distribution.*

5.3.2 Panter-Dite quantization

When the number of quantization levels is large, an approximate solution can be obtained by modelling the probability density $p(z)$ as a piecewise constant function [4],

$$p(z) \simeq p(\hat{z}_k), \qquad \hat{z}_k := (z_k + z_{k+1})/2$$

for $z_k \leq z < z_{k+1}$. We then again have that

$$q_k = \hat{z}_k = (z_k + z_{k+1})/2, \qquad k = 1, 2, \cdots, K.$$

Further,

$$\Delta_k := q_k - z_k = z_{k+1} - q_k.$$

Then for the quantization interval errors we find

$$\delta_k = \frac{2}{3}\, p(q_k)\, \Delta_k^3.$$

The total mean-square error δ is obtained as

$$\delta = \sum_{k=1}^{K} \delta_k.$$

Now, it will be proved that δ is minimized if all δ_ks are equal. Furthermore,

$$2\sum_{k=1}^{K} p^{1/3}(q_k)\Delta_k \simeq \int_{z=z_1}^{z_{K+1}} p^{1/3}(z)\, dz = 2S,$$

where S is a constant, dependent on the fixed probability distribution $p(z)$. Let $\mu_k = p^{1/3}(q_k)\Delta_k$, then

$$\begin{aligned} \delta &= \tfrac{2}{3}\textstyle\sum_{k=1}^{K} \mu_k^3, \\ S &= \textstyle\sum_{k=1}^{K} \mu_k. \end{aligned}$$

The problem is now to minimize a sum of cubes, when the sum of the variables themselves should remain constant. Using the method of Lagrange multipliers leads us to

$$\mu_1 = \mu_2 = \cdots = \mu_K = \frac{S}{K},$$

or,

$$\delta_1 = \delta_2 = \cdots = \delta_K = \frac{2}{3}\left(\frac{S}{K}\right)^3.$$

From the μ_ks we can calculate the Δ_ks, thereby fully specifying the interval boundaries z_k and the quantization levels q_k. We have

$$\begin{aligned} z_k &= z_1 + 2\Delta_1 + 2\Delta_2 + \cdots + 2\Delta_{k-2} + 2\Delta_{k-1} \\ &= z_1 + \tfrac{2S}{K}\left[\tfrac{1}{p^{1/3}(q_1)} + \tfrac{1}{p^{1/3}(q_2)} + \cdots + \tfrac{1}{p^{1/3}(q_{k-2})} + \tfrac{1}{p^{1/3}(q_{k-1})}\right]. \end{aligned}$$

[4] This scheme has been derived by Panter and Dite

A continuous version of this scheme is given by

$$z_k \simeq \frac{KQ \int_{z_1}^{z_1+kQ} p(z)^{-1/3} dz}{\int_{z_1}^{z_{K+1}} p(z)^{-1/3} dz} + z_1,$$

with $Q = (z_{K+1} - z_1)/K$ and $k = 0, \cdots, K$. Notice that $k = 0$ yields z_1 and $k = K$ yields z_{K+1}, thus using the complete dynamic range.

sectionFurther reading

Not available yet