3 Digital Holography

3.1 General Principles

The concept of digital hologram recording is illustrated in figure 3.1(a) [131]. A plane reference wave and the wave reflected from the object are interfering at the surface of a Charged Coupled Device (CCD). The resulting hologram is electronically recorded and stored. The object is in general a three dimensional body with diffusely reflecting surface, located at a distance *d* from the CCD.

In optical reconstruction the virtual image appears at the position of the original object and the real image is formed at a distance *d* as well, but in the opposite direction from the CCD, see figure 3.1 (b).

The diffraction of a light wave at an aperture (in this case a hologram) which is mounted perpendicular to the incoming beam is described by the Fresnel-Kirchhoff integral, see Eq. (2.48):

$$\Gamma(\xi',\eta') = \frac{i}{\lambda}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)E_R(x,y)\frac{\exp\left(-i\frac{2\pi}{\lambda}\rho'\right)}{\rho'}dxdy \tag{3.1}$$

with

$$\rho' = \sqrt{(x-\xi')^2 + (y-\eta')^2 + d^2} \tag{3.2}$$

h(x,y) is the hologram function and ρ' is the distance between a point in the hologram plane and a point in the reconstruction plane. The geometrical quantities are explained in figure 3.2. The inclination factor is set to 1, because the angles θ and θ' (see chapter 2.4) are approximately 0. This is valid for all numerical reconstruction algorithms in this book.

For a plane reference wave $E_R(x,y)$ is simply given by the real amplitude:

$$E_R = a_R + i0 = a_R \tag{3.3}$$

The diffraction pattern is calculated at a distance *d* behind the CCD plane, which means it reconstructs the complex amplitude in the plane of the real image.

Eq. (3.1) is the basis for numerical hologram reconstruction. Because the reconstructed wave field $\Gamma(\xi',\eta')$ is a complex function, both the intensity as well as the phase can be calculated [128]. This is in contrast to the case of optical hologram reconstruction, in which only the intensity is made visible. This interesting

property of Digital Holography is used in Digital Holographic Interferometry, see chapter 4.

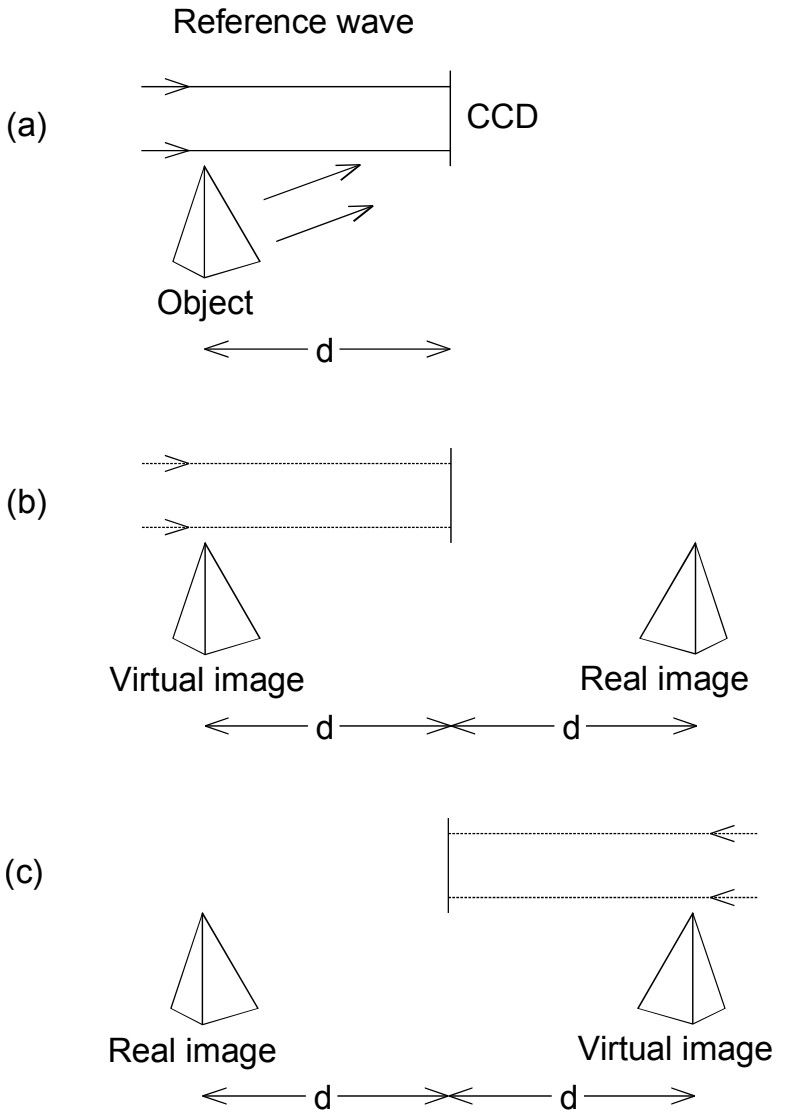

Fig. 3.1. Digital Holography
(a) Recording
(b) Reconstruction with reference wave E_R
(c) Reconstruction with conjugate reference wave E_R^*

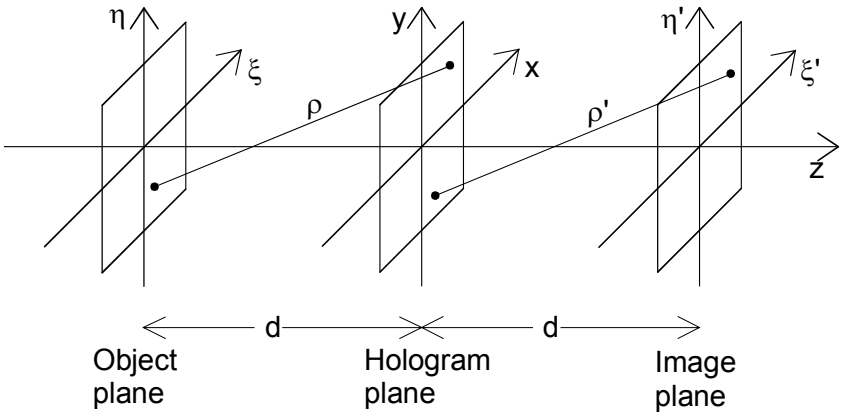

Fig. 3.2. Coordinate system for numerical hologram reconstruction

As mentioned in chapter 2.6 the real image could be distorted. According to Eq. (2.63) an undistorted real image can be produced by using the conjugate reference beam for reconstruction. To reconstruct an undistorted real image in Digital Holography it is therefore necessary to insert E_R^* instead of E_R in Eq. (3.1):

$$\Gamma(\xi,\eta) = \frac{i}{\lambda}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)E_R^*(x,y)\frac{\exp\left(-i\frac{2\pi}{\lambda}\rho\right)}{\rho}dxdy \tag{3.4}$$

with

$$\rho = \sqrt{(x-\xi)^2 + (y-\eta)^2 + d^2} \tag{3.5}$$

This reconstruction scheme is shown in figure 3.1(c). The real image emerges at that position, where the object was located during recording. It should be mentioned that for the plane reference wave defined in Eq. (3.3) both reconstruction formulas, Eq. (3.1) and (3.4), are equivalent because $E_R = E_R^* \equiv a_R$.

The set-up of figure 3.1 with a plane reference wave impinging perpendicularly onto the CCD is often used in Digital Holography. Other recording geometries are discussed later.

The reconstruction of the virtual image is also possible by introducing the imaging properties of a lens into the numerical reconstruction process [129]. This lens corresponds to the eye lens of an observer watching through an optically reconstructed hologram. In the simplest case this lens is located directly behind the hologram, figure 3.3. The imaging properties of a lens with focal distance *f* are considered by a complex factor:

$$L(x,y) = \exp\left[i\frac{\pi}{\lambda f}\left(x^2 + y^2\right)\right] \tag{3.6}$$

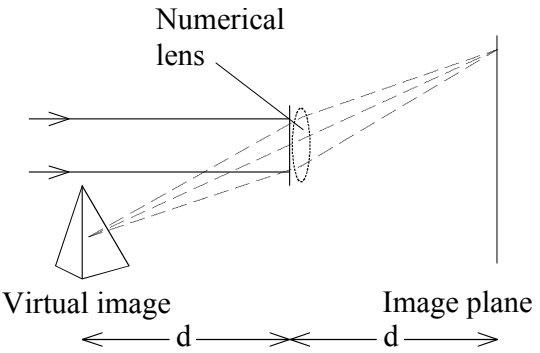

Fig. 3.3. Reconstruction of the virtual image

This factor is calculated in Annex B1. For a magnification of 1 a focal distance of $f = d/2$ has to be used.

The lens described by Eq. (3.6) causes phase aberrations, which can be corrected by multiplying the reconstructed wave field by a factor

$$P(\xi',\eta') = \exp\left[i\frac{\pi}{\lambda f}\left(\xi'^2 + \eta'^2\right)\right] \tag{3.7}$$

This correction factor is derived in Annex B2. The full formula for reconstruction via a virtual lens with $f = d/2$ is therefore

$$\Gamma(\xi',\eta') = \frac{i}{\lambda} P(\xi',\eta') \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y) E_R(x,y) L(x,y) \frac{\exp\left(-i\frac{2\pi}{\lambda}\rho'\right)}{\rho'} dxdy \tag{3.8}$$

3.2 Numerical Reconstruction

3.2.1 Reconstruction by the Fresnel Approximation

For x- and y-values as well as for ξ- and η-values, which are small compared to the distance d between the reconstruction plane and the CCD, the expression Eq. (3.5) can be expanded to a Taylor series:

$$\rho = d + \frac{(\xi - x)^2}{2d} + \frac{(\eta - x)^2}{2d} - \frac{1}{8}\frac{\left[(\xi - x)^2 + (\eta - x)^2\right]^2}{d^3} + \ldots \tag{3.9}$$

The fourth term can be neglected, if it is small compared to the wavelength [69]:

$$\frac{1}{8}\frac{\left[(\xi - x)^2 + (\eta - x)^2\right]^2}{d^3} << \lambda \tag{3.10}$$

or

$$d >> \sqrt[3]{\frac{1}{8}\frac{\left[(\xi - x)^2 + (\eta - x)^2\right]^2}{\lambda}} \tag{3.11}$$

Then the distance ρ consists of linear and quadratic terms:

$$\rho = d + \frac{(\xi - x)^2}{2d} + \frac{(\eta - x)^2}{2d} \tag{3.12}$$

With the additional approximation of replacing the denominator in (3.4) by d the following expression results for reconstruction of the real image:

$$\Gamma(\xi,\eta) = \frac{i}{\lambda d}\exp\left(-i\frac{2\pi}{\lambda}d\right) \times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E_R^*(x,y)h(x,y)\exp\left[-i\frac{\pi}{\lambda d}\left((\xi - x)^2 + (\eta - y)^2\right)\right]dxdy \tag{3.13}$$

If the multiplications in the argument of the exponential under the integral are carried out one gets

$$\Gamma(\xi,\eta) = \frac{i}{\lambda d}\exp\left(-i\frac{2\pi}{\lambda}d\right)\exp\left[-i\frac{\pi}{\lambda d}\left(\xi^2 + \eta^2\right)\right] \times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E_R^*(x,y)h(x,y)\exp\left[-i\frac{\pi}{\lambda d}\left(x^2 + y^2\right)\right]\exp\left[i\frac{2\pi}{\lambda d}(x\xi + y\eta)\right]dxdy \tag{3.14}$$

This equation is named *Fresnel approximation* or *Fresnel transformation* due to its mathematical similarity with the Fourier Transform (see below). It enables reconstruction of the wave field in a plane behind the hologram, in this case in the plane of the real image.

The intensity is calculated by squaring:

$$I(\xi,\eta) = |\Gamma(\xi,\eta)|^2 \tag{3.15}$$

The phase is calculated by

$$\varphi(\xi,\eta) = \arctan\frac{\mathrm{Im}[\Gamma(\xi,\eta)]}{\mathrm{Re}[\Gamma(\xi,\eta)]} \tag{3.16}$$

Re denotes the real part and Im the imaginary part.

The reconstruction formula for the virtual image in the Fresnel approximation is:

$$\begin{aligned}\Gamma(\xi',\eta') &= \frac{i}{\lambda d}\exp\left(-i\frac{2\pi}{\lambda}d\right)\exp\left[-i\frac{\pi}{\lambda d}\left(\xi'^2+\eta'^2\right)\right]P(\xi',\eta') \\ &\times\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E_R(x,y)L(x,y)h(x,y)\exp\left[-i\frac{\pi}{\lambda d}\left(x^2+y^2\right)\right]\exp\left[i\frac{2\pi}{\lambda d}\left(x\xi'+y\eta'\right)\right]dxdy \\ &= \frac{i}{\lambda d}\exp\left(-i\frac{2\pi}{\lambda}d\right)\exp\left[+i\frac{\pi}{\lambda d}\left(\xi'^2+\eta'^2\right)\right] \\ &\times\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E_R(x,y)h(x,y)\exp\left[+i\frac{\pi}{\lambda d}\left(x^2+y^2\right)\right]\exp\left[i\frac{2\pi}{\lambda d}\left(x\xi'+y\eta'\right)\right]dxdy\end{aligned} \tag{3.17}$$

For digitisation of the Fresnel transform Eq. (3.14) following substitutions are introduced [176]:

$$\nu = \frac{\xi}{\lambda d}; \qquad \mu = \frac{\eta}{\lambda d} \tag{3.18}$$

Herewith (3.14) becomes

$$\begin{aligned}\Gamma(\nu,\mu) &= \frac{i}{\lambda d}\exp\left(-i\frac{2\pi}{\lambda}d\right)\exp\left[-i\pi\lambda d\left(\nu^2+\mu^2\right)\right] \\ &\times\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E_R^{\,*}(x,y)h(x,y)\exp\left[-i\frac{\pi}{\lambda d}\left(x^2+y^2\right)\right]\exp\left[i2\pi\left(x\nu+y\mu\right)\right]dxdy\end{aligned} \tag{3.19}$$

A comparison of Eq. (3.19) with the definition of the two-dimensional Fourier transform (see Annex A) shows that the Fresnel approximation up to a spherical phase factor is the inverse Fourier transformation of the function $E_R^{\,*}(x,y)h(x,y)\exp\left[-i\pi/\lambda d\left(x^2+y^2\right)\right]$:

$$\begin{aligned}\Gamma(\nu,\mu) &= \frac{i}{\lambda d}\exp\left(-i\frac{2\pi}{\lambda}d\right)\exp\left[-i\pi\lambda d\left(\nu^2+\mu^2\right)\right] \\ &\times\Im^{-1}\left\{E_R^{\,*}(x,y)h(x,y)\exp\left[-i\frac{\pi}{\lambda d}\left(x^2+y^2\right)\right]\right\}\end{aligned} \tag{3.20}$$

The function Γ can be digitised if the hologram function *h(x,y)* is sampled on a rectangular raster of N×N points, with steps Δx and Δy along the coordinates. Δx and Δy are the distances between neighbouring pixels on the CCD in horizontal

and vertical direction. With these discrete values the integrals of (3.19) are converted to finite sums:

$$\Gamma(m,n) = \frac{i}{\lambda d}\exp\left(-i\frac{2\pi}{\lambda}d\right)\exp\left[-i\pi\lambda d\left(m^2\Delta\nu^2 + n^2\Delta\mu^2\right)\right] \tag{3.21}$$
$$\times\sum_{k=0}^{N-1}\sum_{l=0}^{N-1} E_R^{\,*}(k,l)h(k,l)\exp\left[-i\frac{\pi}{\lambda d}\left(k^2\Delta x^2 + l^2\Delta y^2\right)\right]\exp\left[i2\pi\left(k\Delta x m\Delta\nu + l\Delta y n\Delta\mu\right)\right]$$

for $m = 0,1,...,N-1$; $n = 0,1,...,N-1$

According to the theory of Fourier transform among Δx, Δy and $\Delta\nu$, $\Delta\mu$ the following relation exist, see Annex A:

$$\Delta\nu = \frac{1}{N\Delta x}; \qquad \Delta\mu = \frac{1}{N\Delta y} \tag{3.22}$$

After re-substitution:

$$\Delta\xi = \frac{\lambda d}{N\Delta x}; \qquad \Delta\eta = \frac{\lambda d}{N\Delta y} \tag{3.23}$$

Using these equations (3.21) converts to

$$\Gamma(m,n) = \frac{i}{\lambda d}\exp\left(-i\frac{2\pi}{\lambda}d\right)\exp\left[-i\pi\lambda d\left(\frac{m^2}{N^2\Delta x^2} + \frac{n^2}{N^2\Delta y^2}\right)\right] \tag{3.24}$$
$$\times\sum_{k=0}^{N-1}\sum_{l=0}^{N-1} E_R^{\,*}(k,l)h(k,l)\exp\left[-i\frac{\pi}{\lambda d}\left(k^2\Delta x^2 + l^2\Delta y^2\right)\right]\exp\left[i2\pi\left(\frac{km}{N} + \frac{ln}{N}\right)\right]$$

This is the discrete Fresnel transform. The matrix Γ is calculated by multiplying $E_R^{\,*}(k,l)$ with *h(k,l)* and $\exp\left[-i\pi/(\lambda d)\left(k^2\Delta x^2 + l^2\Delta y^2\right)\right]$ and applying an inverse discrete Fourier transform to the product. The calculation is done most effectively using the Fast Fourier Transform (FFT) algorithm. The factor in front of the sum only affects the overall phase and can be neglected, if just the intensity according to Eq. (3.15) is of interest. This is also the case if phase differences between holograms recorded with the same wavelength have to be calculated ($\Delta\varphi = \varphi_1 + const. - (\varphi_2 + const.) = \varphi_1 - \varphi_2$).

The corresponding discrete formula for reconstruction via a virtual lens with $f = d/2$ (Eq. 3.17) is

$$\Gamma(m,n) = \frac{i}{\lambda d}\exp\left(-i\frac{2\pi}{\lambda}d\right)\exp\left[+i\pi\lambda d\left(\frac{m^2}{N^2\Delta x^2}+\frac{n^2}{N^2\Delta y^2}\right)\right] \tag{3.25}$$
$$\times\sum_{k=0}^{N-1}\sum_{l=0}^{N-1}E_R(k,l)h(k,l)\exp\left[+i\frac{\pi}{\lambda d}\left(k^2\Delta x^2+l^2\Delta y^2\right)\right]\exp\left[i2\pi\left(\frac{km}{N}+\frac{ln}{N}\right)\right]$$

A typical digital hologram is shown in figure 3.4. It was recorded with the geometry of figure 3.1. The object is placed *d=1.054m* apart from the CCD-array of 1024×1024 pixels with pixel distances $\Delta x = \Delta y = 6.8\mu m$. The wavelength is 632.8nm. The numerical reconstruction according to Eq. (3.14) resp. (3.24) is demonstrated in figure 3.5. A real image of the dice used as object is noticeable. The bright square in the centre is the undiffracted reconstruction wave (zero order) and corresponds to the first term of the right side of Eq. (2.63). Because of the off-axis geometry the image is spatially separated from the zero order term. The other (virtual) image is out of focus in this reconstruction.

An interesting property of holography is that every part of a hologram contains the information about the entire object. This is illustrated by the holograms of figures 3.6 and 3.8, where black masks cover nearly half of the hologram areas. Nevertheless the entire cube is visible in the reconstructions without obstructions, figures 3.7 and 3.9. The masks are visible as shadows in the zero order terms. The reduction of the effective pixel number leads to a reduction of the resolution in the reconstructed images. This corresponds to the increase of the speckle size due to aperture reduction in optical hologram reconstruction.

Regarding Eq. (3.23) the pixel distances in the reconstructed image $\Delta\xi$ and $\Delta\eta$ are depending on the reconstruction distance *d* chosen for numerical reconstruction. This is because Eq. (3.23) corresponds to the diffraction limited resolution of optical systems: The hologram is the aperture of the optical system with side length $N\Delta x$. According to the theory of diffraction a diffraction pattern develops in a distance *d* behind the hologram. $\Delta\xi = \lambda d/N\Delta x$ is therefore the half diameter of the airy disk or speckle diameter in the plane of the reconstructed image, which limits the resolution. This can be regarded as "natural scaling" algorithm, setting the resolution of the image reconstructed by a discrete Fresnel transform always to the physical limit.

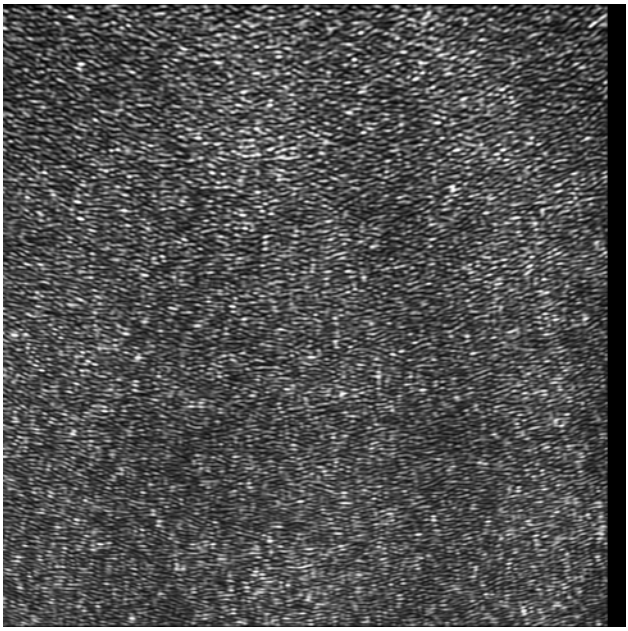

Fig. 3.4. Digital hologram

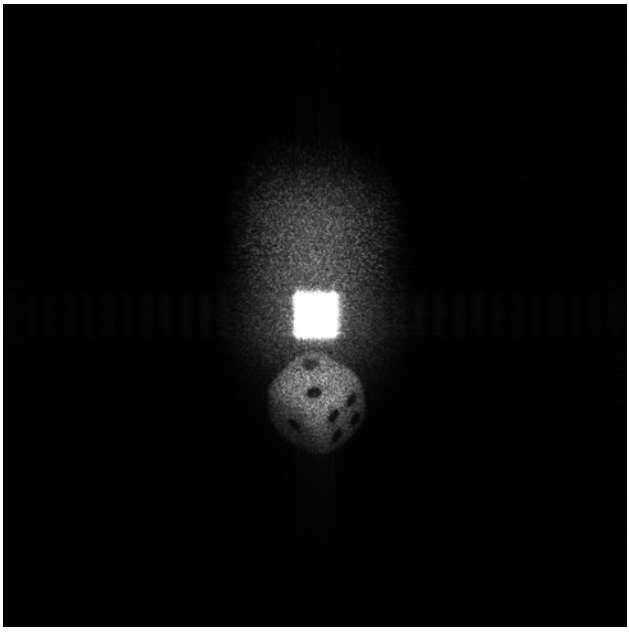

Fig. 3.5. Numerical reconstruction

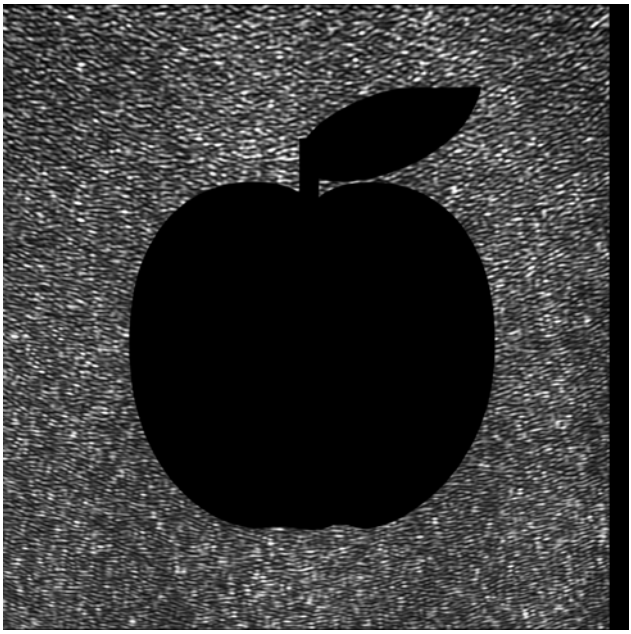

Fig. 3.6. Masked digital hologram

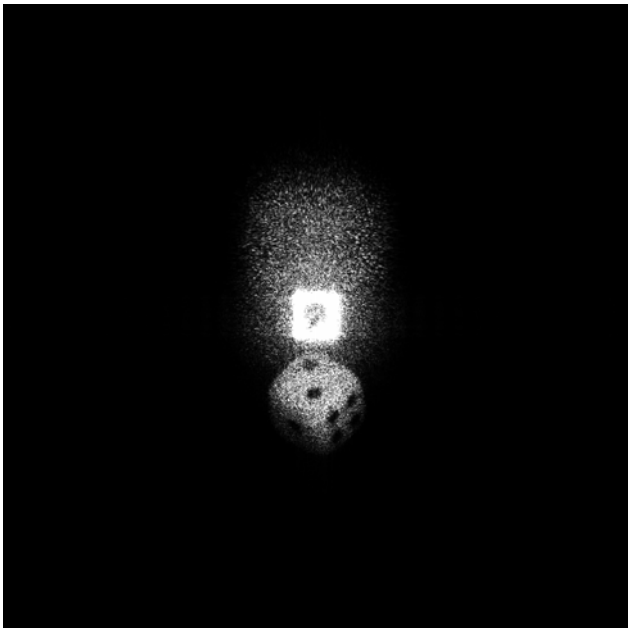

Fig. 3.7. Reconstruction

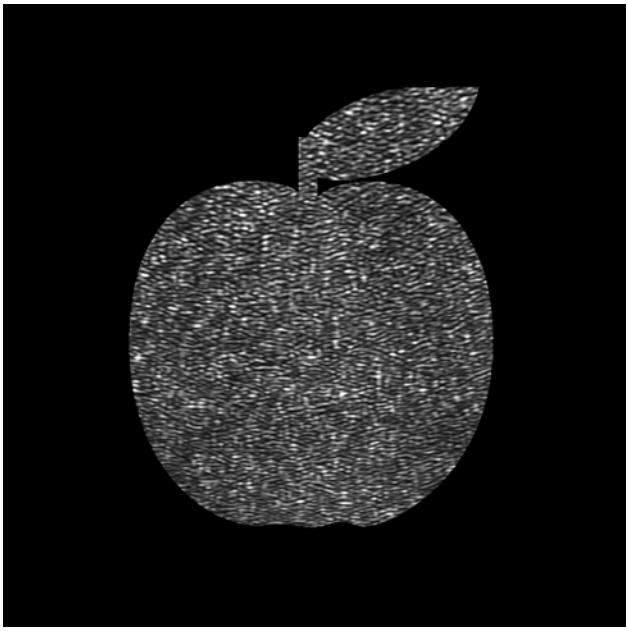

Fig. 3.8. Masked digital hologram

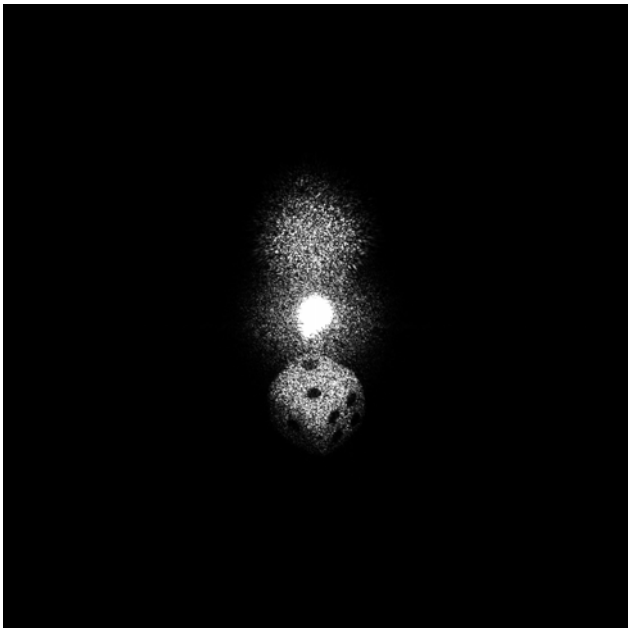

Fig. 3.9. Reconstruction

3.2.2 Reconstruction by the Convolution Approach

The numerical processing of the Fresnel-Kirchhoff integral Eq. (3.1) resp. (3.4) without approximation is time consuming. For numerical processing a different but equivalent formulation is much more suitable. This formulation makes use of the convolution theorem and is named "convolution approach" within the scope of this article. Demetrakopoulos and Mittra applied this way of processing for numerical reconstruction of sub optical holograms for the first time [25]. Later this approach was applied to optical holography by Kreis [74].

The reconstruction formula Eq. (3.4) can be interpreted as a superposition integral:

$$\Gamma(\xi,\eta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)E_R^{*}(x,y)g(\xi,\eta,x,y)dxdy \tag{3.26}$$

where the impulse response $g(x,y,\xi,\eta)$ is given by

$$g(\xi,\eta,x,y) = \frac{i}{\lambda}\frac{\exp\left[-i\frac{2\pi}{\lambda}\sqrt{d^2+(x-\xi)^2+(y-\eta)^2}\right]}{\sqrt{d^2+(x-\xi)^2+(y-\eta)^2}} \tag{3.27}$$

According to Eq. (3.26) the linear system characterized by $g(\xi,\eta,x,y) = g(\xi-x,\eta-y)$ is space-invariant. The superposition integral can be regarded therefore as a convolution and the convolution theorem (Annex A) can be applied. According to this theorem the Fourier transform of the convolution of $h \cdot E_R^{*}$ with g is the product of the individual transforms $\Im\{hE_R^{*}\}$ and $\Im\{g\}$. So $\Gamma(\xi,\eta)$ can be calculated by Fourier transforming $h \cdot E_R^{*}$ first. This is followed by multiplying with the Fourier transform of g, and taking an inverse Fourier transform of this product. In all three Fourier transforms are necessary for the whole process. The individual Fourier transforms are effectively carried out using the FFT-algorithm.

The numerical impulse response function is

$$g(k,l) = \frac{i}{\lambda}\frac{\exp\left[-i\frac{2\pi}{\lambda}\sqrt{d^2+\left(k-\frac{N}{2}\right)^2\Delta x^2+\left(l-\frac{N}{2}\right)^2\Delta y^2}\right]}{\sqrt{d^2+\left(k-\frac{N}{2}\right)^2\Delta x^2+\left(l-\frac{N}{2}\right)^2\Delta y^2}} \tag{3.28}$$

The shift of the coordinates by $N/2$ is on symmetry reasons.

In short notation the reconstruction into the real image plane is

$$\Gamma(\xi,\eta) = \Im^{-1}\left\{\Im\left(h \cdot E_R^{*}\right) \cdot \Im(g)\right\} \tag{3.29}$$

The Fourier transform of *g(k,l)* can be calculated and expressed analytically:

$$G(n,m)=\exp\left\{-i\frac{2\pi d}{\lambda}\sqrt{1-\frac{\lambda^2\left(n+\frac{N^2\Delta x^2}{2d\lambda}\right)^2}{N^2\Delta x^2}-\frac{\lambda^2\left(m+\frac{N^2\Delta y^2}{2d\lambda}\right)^2}{N^2\Delta y^2}}\right\} \tag{3.30}$$

This saves one Fourier transform for reconstruction:

$$\Gamma(\xi,\eta)=\Im^{-1}\left\{\Im\left(h\cdot E_R^{\,*}\right)\cdot G\right\} \tag{3.31}$$

For the reconstruction of the virtual image a lens transmission factor $L(x,y)$ according to Eq. (3.6) and a correction factor $P(\xi',\eta')$ according to Eq. (3.7) have to be considered:

$$\Gamma(\xi',\eta')=P(\xi',\eta')\Im^{-1}\left\{\Im\left(h\cdot E_R\cdot L\right)\cdot G\right\} \tag{3.32}$$

The pixel distances of the images reconstructed by the convolution approach are equal to that of the hologram:

$$\Delta\xi=\Delta x\,;\qquad \Delta\eta=\Delta y \tag{3.33}$$

The pixel distances of the reconstructed images differ from those of the Fresnel approximation (3.23). At first sight it seems to be possible to achieve a higher resolution with the convolution approach if the pixel distance is small enough. On closer examination one recognizes that the resolution calculated by Eq. (3.33) is only a numerical value. The physical image resolution is determined by diffraction, i. e. Eq. (3.23) describes the resolution limit also for the convolution approach.

As stated in chapter 3.1 the inclination factor in the diffraction integral can be neglected in the numerical reconstruction process. However, it should be mentioned that is also possible to consider the inclination factor in the convolution approach.

The area reconstructed with the impulse response function defined in Eq. (3.28) is symmetrical with respect to the optical axis. The area can be shifted by introducing integer numbers s_k, l_k:

$$g(k+s_k,l+s_l)=\frac{i}{\lambda}\frac{\exp\left[-i\frac{2\pi}{\lambda}\sqrt{d^2+\left(k-\frac{N}{2}+s_k\right)^2\Delta x^2+\left(l-\frac{N}{2}+s_l\right)^2\Delta y^2}\right]}{\sqrt{d^2+\left(k-\frac{N}{2}+s_k\right)^2\Delta x^2+\left(l-\frac{N}{2}+s_l\right)^2\Delta y^2}} \tag{3.34}$$

The convolution approach contains the possibility to introduce an image magnification in the reconstruction process. This is possible if the reconstruction distance is set to

$$d' = d \cdot m \tag{3.35}$$

where d is the recording distance (also used as reconstruction distance so far) and m stands for the magnification factor. A magnification of $m = 1$ corresponds to $\Delta\xi = \Delta x$, resp. $\Delta\eta = \Delta y$. The lens focal distance is given by the lens formula of geometrical optics:

$$f = \left(\frac{1}{d} + \frac{1}{d'} \right)^{-1} \tag{3.36}$$

Now Eq. (3.32) is applied for reconstruction with d' instead of d and

$$L(x,y) = \exp\left[i \frac{\pi}{\lambda f} \left(x^2 + y^2 \right) \right] = \exp\left[i \frac{\pi}{\lambda} \left(\frac{1}{d} + \frac{1}{d'} \right) \left(x^2 + y^2 \right) \right] \tag{3.37}$$

An example of a reconstruction with the convolution approach is shown in figure 3.10. The hologram of figure 3.4 is reconstructed with a magnification of $m = 1/7$. The corresponding pixel distance in the reconstructed image is $\Delta\xi = \Delta x / m = 48\mu m$. The comparison to $\Delta\xi = 96\mu m$ for the Fresnel reconstruction is shown in figure 3.5. That means twice as much pixels are available for the object field. However, it is emphasized again that the physical resolution is the same in figures 3.5 and 3.10.

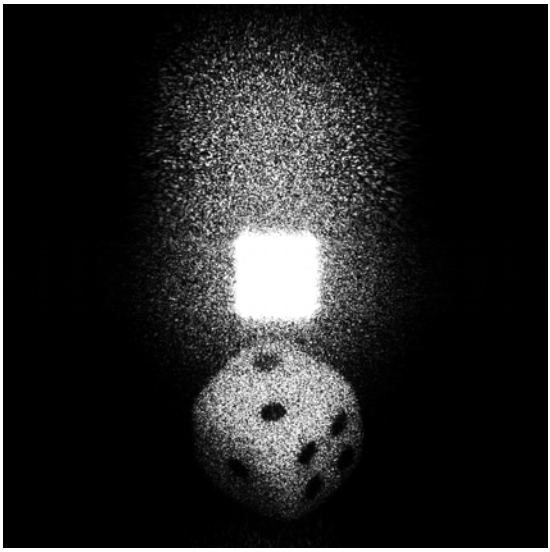

Fig. 3.10. Reconstruction with the convolution approach

3.2.3 Digital Fourier Holography

The special holographic recording geometry of figure 3.11 is named *lensless Fourier holography*. It also has been realized in Digital Holography [162]. The point source of the spherical reference wave is located in the plane of the object. The reference wave at the CCD plane is therefore described by:

$$E_R = \frac{\exp\left(-i\frac{2\pi}{\lambda}\sqrt{\left(d^2+x^2+y^2\right)}\right)}{\sqrt{\left(d^2+x^2+y^2\right)}} \approx \frac{1}{d}\exp\left(-i\frac{2\pi}{\lambda}d\right)\exp\left(-i\frac{\pi}{\lambda d}\left(x^2+y^2\right)\right) \tag{3.38}$$

The term $\sqrt{d^2+x^2+y^2}$ is the distance between the source point and the point with coordinates *(x,y)* in the CCD plane. The approximation in Eq. (3.38) is the same as used in chapter 3.2.1 to derive the Fresnel transform. Inserting this expression into the reconstruction formula for the virtual image (3.17) leads to following equation:

$$\Gamma(\xi,\eta) = C\exp\left[+i\frac{\pi}{\lambda d}\left(\xi^2+\eta^2\right)\right]\Im^{-1}\{h(x,y)\} \tag{3.39}$$

C is a complex constant. A lensless Fourier hologram is therefore reconstructed by a Fourier transform. It is not possible to focus on different areas within the object volume with lensless Fourier holography, because the reconstruction distance *d* does not appear in Eq. (3.39). The spherical phase factor $\exp\left(-i\pi/\lambda d\left(x^2+y^2\right)\right)$ associated with the Fresnel transform is eliminated by the use of a spherical reference wave with the same curvature.

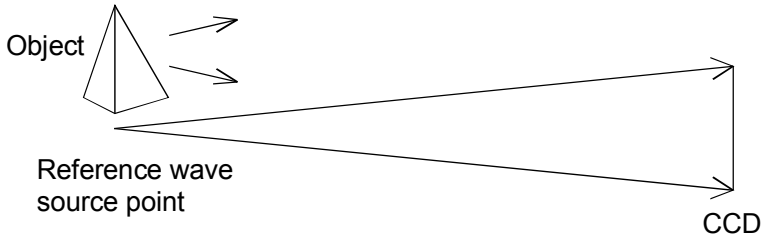

Fig. 3.11. Digital lensless Fourier holography

3.3 Separation of Virtual Image, Real Image and DC-term

3.3.1 Suppression of the DC term

The bright square in the centre of figure 3.5 is the undiffracted reconstruction wave. This zero order or DC term disturbs the image, because it covers all object parts lying behind. Methods have been developed to suppress this term e.g. by Kreis et. al. [75].

To understand the cause of this DC term hologram formation according Eq. (2.60) is considered. The equation is rewritten by inserting the definitions of E_R and E_O and multiplying the terms:

$$\begin{aligned} I(x,y) &= \left|E_O(x,y) + E_R(x,y)\right|^2 \\ &= a_R{}^2 + a_O{}^2 + 2a_R a_O \cos(\varphi_O - \varphi_R) \end{aligned} \tag{3.40}$$

The first two terms are leading to the DC term in the reconstruction process. The third term is statistically varying between $\pm 2a_R a_O$ from pixel to pixel at the CCD. The average intensity of all pixels of the hologram matrix is

$$I_m = \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} I(k\Delta x, l\Delta y) \tag{3.41}$$

The term $a_R{}^2 + a_O{}^2$ now can be suppressed by subtracting this average intensity I_m from the hologram:

$$I'(k\Delta x, l\Delta y) = I(k\Delta x, l\Delta y) - I_m(k\Delta x, l\Delta y) \tag{3.42}$$

for $k = 0,...,N-1; \quad l = 0,...,N-1$.

The reconstruction of I' creates an image free of the zero order.

An example is shown in figure 3.12. The upper left figure is a photograph of the object. The reconstruction without DC term suppression is depicted in the upper right figure. The object is covered by the DC term. The lower left figure shows the reconstruction with DC term suppression as described above. The object is clearly visible.

Instead of subtracting the average intensity it is also possible to filter the hologram matrix by a high-pass with low cut-off frequency. The lower right image of figure 3.12 shows the DC term suppression by high pass filtering of the hologram.

The subtraction of the average intensity from the hologram before reconstruction is the basic idea of the DC-term suppression. The same effect can be achieved, if two holograms with stochastically changed speckle structures are subtracted from each other [26]. The reconstruction of this subtraction hologram results in an image without zero order term.

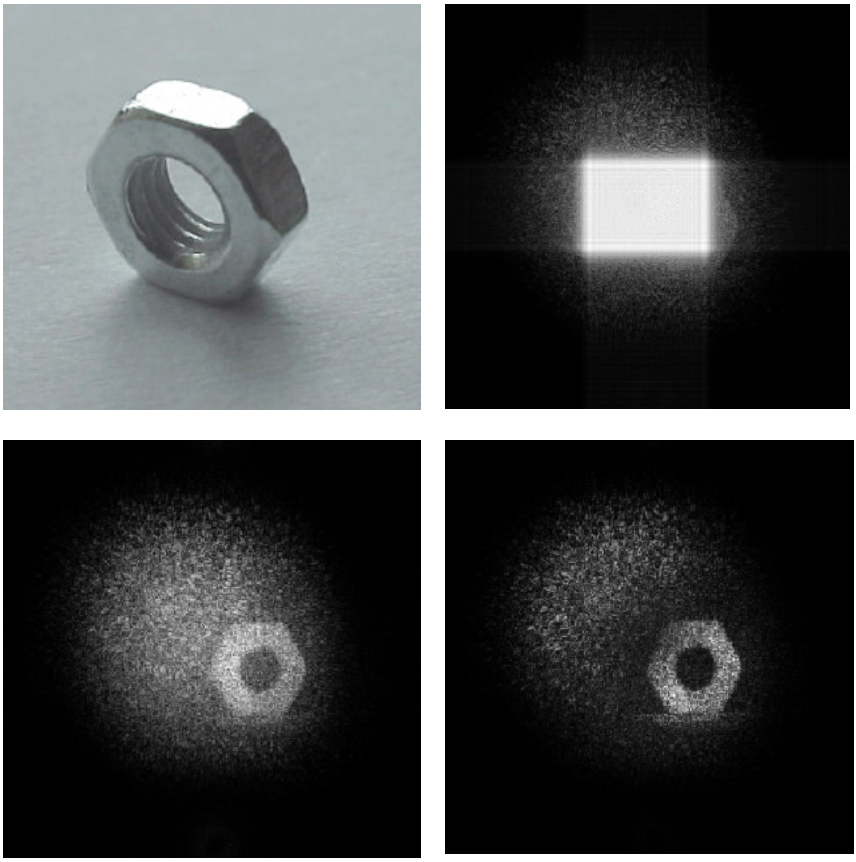

Fig. 3.12. Suppression of the DC term (courtesy of S. Seebacher)

Another method to suppress the DC term is to measure the intensities of reference wave a_R^2 and object wave a_O^2 separately. This can be done e. g. by blocking one of the waves while measuring the intensity of the other wave. Afterwards a DC term free image can be calculated by subtracting the intensities from the hologram before reconstruction. However, this requires higher experimental effort due to the additional measurements.

3.3.2 Spatial Separation of Images

With the recording geometry of figure 3.1 the real image and the virtual image are located in one line. In the numerical reconstruction one focuses either on the real or on the virtual image. The other image is usually out of focus due to the long

distance between object and CCD. Consequently only one image is visible in the reconstruction, see example in figure 3.5.

Still, there are some cases, in which the reconstruction might be disturbed by the out of focus twin image, e. g. if the recording distance is very short. For this special case it might be useful to record the holograms with a tilted reference, see set-up in figure 3.13. In this geometry both images are angularly separated in the reconstruction. The real image is deflected off the axis at an angle approximately twice as large as the one between reference wave and axis.

The tilted reference wave is described by

$$E_R = \exp\left(-i\frac{2\pi}{\lambda}x\sin\theta\right) \tag{3.43}$$

The disadvantage of this set-up are the much higher spatial frequencies on the CCD in comparison to the set-up of figure 3.1. This is critical with respect to the CCD resolution and limits the usable angle for the object wave, see discussion in chapter 3.4.

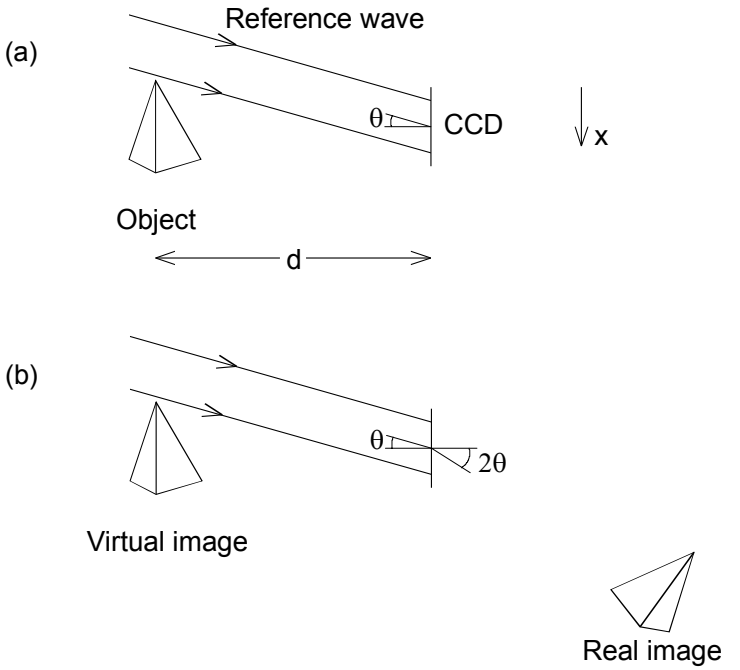

Fig. 3.13. Digital Holography with a tilted reference wave
(a) Recording
(b) Reconstruction

3.3.3 Phase Shifting Digital Holography

The amplitude and phase of a light wave can be reconstructed from a single hologram by the methods described in the preceding chapters. A completely different approach has been proposed by Skarman [146, 165]. He used a phase shifting algorithm to calculate the *initial* phase and thus the complex amplitude in any plane, e. g. the image plane: With the initial complex amplitude distribution in one plane the wave field in any other plane can be calculated by using the Fresnel-Kirchhoff formulation of diffraction. Later this Phase Shifting Digital Holography was improved and applied to opaque objects by Yamaguchi et al. [54, 171-174, 178, 179].

The principal arrangement for phase shifting Digital Holography is shown in figure 3.14. The object wave and the reference wave are interfering at the surface of a CCD. The reference wave is guided via a mirror mounted on a piezoelectric transducer (PZT). With this PZT the phase of the reference wave can be shifted stepwise. Several (at least three) interferograms with mutual phase shifts are recorded. Afterwards the object phase φ_O is calculated from these phase shifted interferograms, the procedure is similar to that of phase shifting in conventional HI (see chapter 2.7.5). The real amplitude $a_O(x,y)$ of the object wave can be measured from the intensity by blocking the reference wave.

As a result the complex amplitude

$$E_O(x,y) = a_O(x,y)\exp(+i\varphi_O(x,y)) \tag{3.44}$$

of the object wave is determined in the recording (x,y) plane.

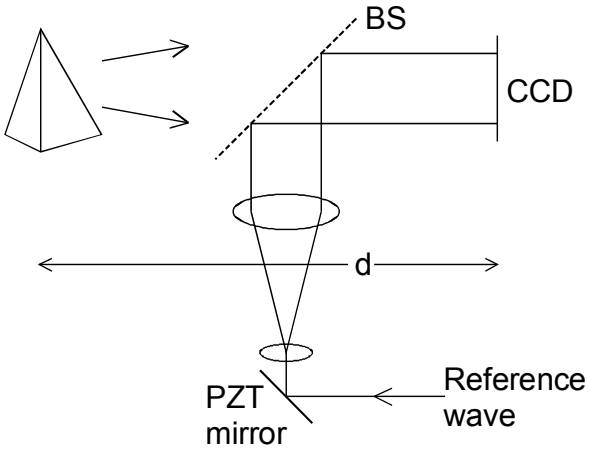

Fig. 3.14. Phase shifting Digital Holography, set-up

Now the Fresnel-Kirchhoff integral can be used to calculate the complex amplitude in any other plane. To calculate an image of the object an artificial lens with $f = d/2$ is introduced in the recording plane according to Eq. (3.6). By means of the Fresnel approximation Eq. (3.17) the complex amplitude in the image plane is then calculated by

$$\begin{aligned} E_O(\xi',\eta') &= C\exp\left[+i\frac{\pi}{\lambda d}\left(\xi'^2+\eta'^2\right)\right] \\ &\times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E_O(x,y)L(x,y)\exp\left[-i\frac{\pi}{\lambda d}\left(x^2+y^2\right)\right]\exp\left[i\frac{2\pi}{\lambda d}\left(x\xi'+y\eta'\right)\right]dxdy \\ &= C\exp\left[+\frac{i\pi}{\lambda d}\left(\xi'^2+\eta'^2\right)\right] \\ &\times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E_O(x,y)\exp\left[+i\frac{\pi}{\lambda d}\left(x^2+y^2\right)\right]\exp\left[i\frac{2\pi}{\lambda d}\left(x\xi'+y\eta'\right)\right]dxdy \end{aligned} \tag{3.45}$$

where again the coordinate system of figure 3.2 is applied. Since the complex amplitude is known in the hologram plane, it is also possible to reconstruct the object by inversion of the recording process [142]. Hologram recording is described by

$$\begin{aligned} E_O(x,y) &= \frac{i}{\lambda}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E_O(\xi,\eta)\frac{\exp\left(-i\frac{2\pi}{\lambda}\sqrt{d^2+(\xi-x)^2+(\eta-y)^2}\right)}{\sqrt{d^2+(\xi-x)^2+(\eta-y)^2}}d\xi d\eta \\ &= \Im^{-1}\left\{\Im\left(E_O(\xi,\eta)\right)\cdot\Im\left(g(\xi,\eta,x,y)\right)\right\} \end{aligned} \tag{3.46}$$

with

$$g(\xi,\eta,x,y) = \frac{i}{\lambda}\frac{\exp\left(-i\frac{2\pi}{\lambda}\sqrt{d^2+(\xi-x)^2+(\eta-y)^2}\right)}{\sqrt{d^2+(\xi-x)^2+(\eta-y)^2}} \tag{3.47}$$

$E_O(\xi,\eta)$ is the complex amplitude of the object wave at the emitting surface, see figure 3.2. Therefore it can be calculated directly by inverting Eq. (3.46):

$$E_O(\xi,\eta) = \Im^{-1}\left\{\frac{\Im\left(E_O(x,y)\right)}{\Im\left(g(\xi,\eta,x,y)\right)}\right\} \tag{3.48}$$

The numerical implementation of this reconstruction method is critical due to the division in Eq. (3.48).

The advantage of phase shifting Digital Holography is a reconstructed image of the object without the zero order term and the conjugate image. The price for this achievement is the higher technical effort: Phase shifted interferograms have to be

generated, restricting the method to slowly varying phenomena with constant phase during the recording cycle.

Phase shifting Digital Holography is illustrated by a holographic image of a nut, figure 3.15. The example demonstrates the improvement compared to conventional Digital Holography, figure. 3.12.

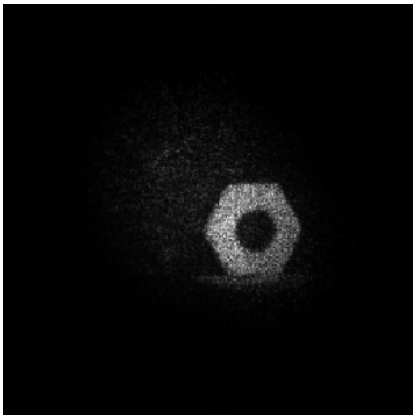

Fig. 3.15. Phase shifting Digital Holography, example

3.4 Recording of Digital Holograms

3.4.1 Charged-Coupled Devices

Charged-Coupled Devices (CCD's) were invented in the sixties of the last century by researchers at Bell Labs. A CCD is an electrical device that is used to create images of objects, store information or transfer electrical charge. The most popular application today is image recording. Therefore the following overview in this book is restricted to this application. CCD's are used as imaging devices in electronic cameras (video and still imaging) and scanners. They are available as line scanning devices, consisting of a single line of light detectors, and as area scanning devices, consisting of a rectangular matrix of light detectors. For Digital Holography only the latter architecture is of interest.

CCD imaging is performed in a three step process [14]:

1. Light exposure
   The incident light separates charges by the internal photo effect. This effect converts light into an electronic charge at the individual detectors called pixels.
2. Charge transfer
   The charge transfer function moves the packets of charge within the semiconductor (silicon) substrate to memory cells.
3. Charge to voltage conversion and output amplification

The capacitor matrix of the memory cells converts the transferred charge to a voltage. An amplifier adapts the voltage to the output requirements.

The three basic architectures of CCD's are called full-frame transfer devices, frame-transfer devices and interline transfer devices.

Interline transfer devices consist of a striped pattern of light sensitive detectors (photodiodes) and of separated non-sensitive or light shielded storage devices, see figure 3.16. The charge packages which are generated in the light sensitive pixels are shifted into the adjacent storage devices by a parallel clock. After that the charges in these storage devices are shifted line by line into the serial register. This serial register transfers the charge packages into a charge to voltage converter with amplifier, which forms the output signal. The major disadvantage of interline transfer CCD's is their complexity, which results from separating the photo-detecting and storage (readout) functions.

Frame-transfer CCD's also have different areas for light conversion and for storage. Yet, these elements are not arranged in stripes, but divided into two different blocks, see figure 3.17. The idea is to shift a captured scene from the photo-sensitive image array very fast to the storage array. The readout from the storage register is performed similar to the readout process of interline transfer devices.

Full-Frame CCD's have the simplest architecture, see figure 3.18. In contrast to interline transfer and frame-transfer devices they don't have a separated storage area. The entire sensor area is light sensitive. The photons of the illuminating light are converted into charge packages. The resulting rows of image information are then shifted in parallel to the serial register that subsequently shifts the row of information to the output as a serial stream of data. The process repeats until all rows are transferred off chip. Since the parallel register is used for both image detection and readout, a mechanical shutter is needed to preserve scene integrity. Yet, full-frame CCD's have highest resolution and the production costs are comparably cheap.

In principle all three types of CCD’s are suitable for Digital Holography. Full frame type CCD’s have the advantage that the exposure time can be adjusted according to the demands of a specific application. Even exposure times in the range of seconds are possible. However, the mechanical shutter limits the number of holograms, which can be recorded per second (frame rate). In addition the shutter may cause mechanical vibrations to the set-up, which deteriorate the hologram quality. An advantage of interline transfer type CCD’s is that such devices are equipped with an electronic shutter, allowing higher frame rates. The best suited camera type depends therefore on the specific holographic application.

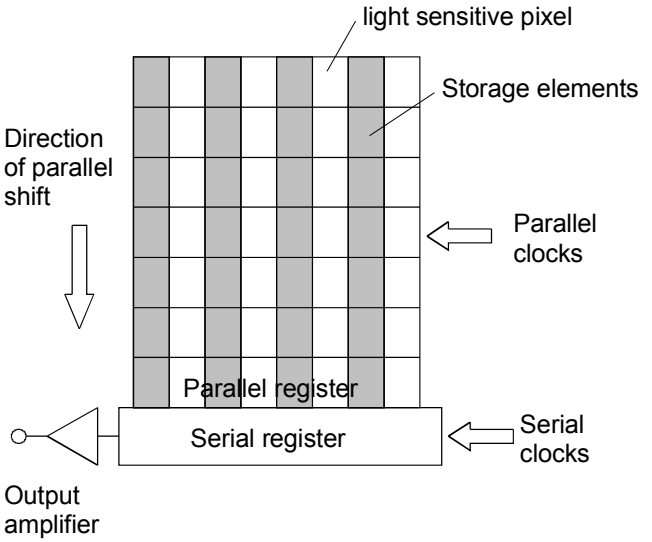

Fig. 3.16. Interline-transfer architecture

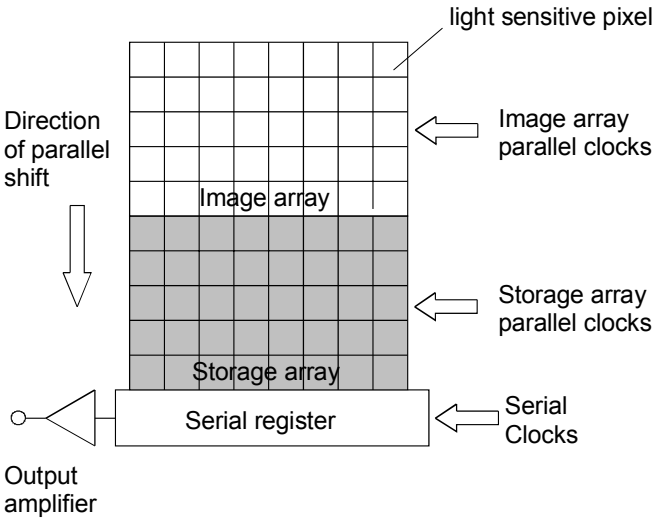

Fig. 3.17. Frame-transfer architecture

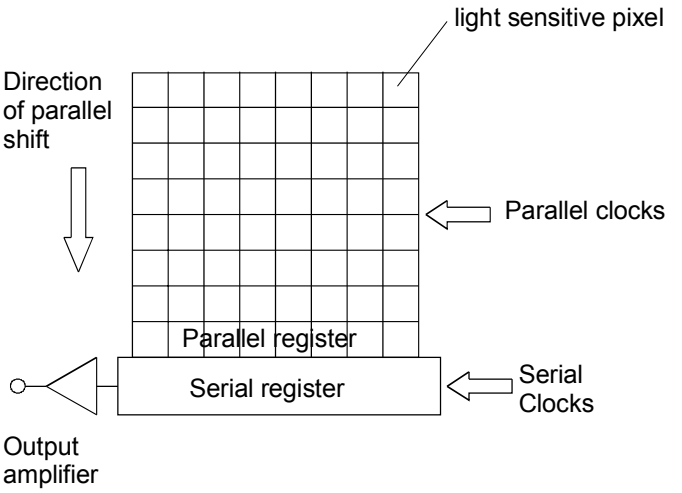

Fig. 3.18. Full-frame architecture

3.4.2 Spatial Frequency Requirements

A CCD used to record holograms must resolve the interference pattern resulting from superposition of the reference wave with the waves scattered from the different object points. The maximum spatial frequency to be resolved is determined by the maximum angle θ_{max} between these waves according to Eq. (2.30):

$$f_{\max} = \frac{2}{\lambda} \sin \frac{\theta_{\max}}{2} \tag{3.49}$$

Photographic emulsions used in optical holography have resolutions up to 5000 line pairs per millimetre (Lp/mm). Using these materials, holograms with angles of up to 180° between the reference and the object wave can be recorded. However, the distance between neighbouring pixels of a high resolution CCD is only in the order of $\Delta x \approx 5 \mu m$. The corresponding maximum resolvable spatial frequency calculated by

$$f_{\max} = \frac{1}{2\Delta x} \tag{3.50}$$

is therefore in the range of 100 Lp/mm.
Combining Eq. (3.49) and (3.50) leads to

$$\theta_{\max} = 2 \arcsin\left(\frac{\lambda}{4\Delta x}\right) \approx \frac{\lambda}{2\Delta x} \tag{3.51}$$

where the approximation is valid for small angles. The distance between neighboring pixels is therefore the quantity, which limits the maximum angle between reference- and object wave.

3.4.3 CCD's for Digital Hologram Recording

Main data of selected CCD cameras suitable for Digital Holography are listed in Table 3.1.

Table 3.1. CCD cameras suitable for Digital Holography

Camera	Chip type	Number of pixels	Pixel size [μm^2]	Frames per second	Dynamic range	θ_{max} for λ=633nm
Roper Sci. MegaPlus 1.4i	FT	1317 x 1035	6.8 x 6.8	6.9	8 bit	2.7°
Roper Sci. MegaPlus 16.8i	FT	4096 x 4096	9 x 9	0.47	8 bit	2.0°
Roper Sci. MegaPlus ES1.0	IT	1008 x 1018	9 x 9	30	8 or 10 bit	2.0°
Roper Sci. MegaPlus 4.0	IT	2048 x 2048	7.4 x 7.4	30	8 or 12 bit	2.45°
Hamamatsu C8484-01	PSI	1344 x 1024	6.45 x 6.45	8.3	12 bit	2.81°
Duncan DT1100	PS	1392 x 1040	4.65 x 4.65	12	8 or 10 bit	3.9°

FT: full frame, IT: interline transfer, PSI: progressive scan interline, PS: progressive scan

The sensitivity of CCD cameras is typically in the range of $10^{-4} J/m^2$ to $10^{-3} J/m^2$, which is better than the sensitivity of photographic emulsions used for classical holography. The spectral range reaches approximately from 400*nm* to 1000*nm*, based on silicon as chip material. Yet, most commercial CCD cameras are equipped with a filter, restricting the sensitivity to the visible range.

In conventional holography with photographic plates the intensity ratio between reference and object wave should be in the range of 5:1 to 10:1 in order to avoid nonlinear effects due to the recording medium. However, from interference theory it is known that the maximum contrast in an interference pattern is achieved if the intensity ratio is 1:1. CCD's have a much better linearity in the exposure curve than photographic emulsions. Consequently, the intensity ratio between reference and object wave should be adjusted to the optimum of 1:1. This can be controlled by covering one half of the expanded reference wave by an aperture. The brightness (measured in grey values) in that half of the CCD being illuminated by reference and object wave together should be twice as high of the brightness in the other half, which is illuminated only by the object wave.

As for classical holography using photographic material, the total light energy impinging on the CCD can be controlled by varying the exposure time. This is usually done with a mechanical or the electronic camera shutter.

CCD-cameras typically have a dynamic range of 8 bit (256 grey values) or higher. This is comparable to the dynamics of photographic materials and fully sufficient for hologram recording. Even objects with brightness variations exceeding the dynamic range of the recording medium by far can be stored and reconstructed, because the object information is coded as interference pattern (hologram).

Numerical reconstruction of digital holograms requires a pixel number, which is a power of 2 (e. g. 1024 x 1024). The pixel numbers of some of the cameras listed in table 3.1 differ from that rule. If the pixel number is e. g. 1317 x 1035 (MegaPlus 1.4i) only 1024 x 1024 pixels are used for reconstruction. In case of a pixel number slightly lower than a power of 2 (e. g. 1008 x 1018 for the MegaPlus ES 1.0) it is advisable to add artificial pixels with grey value zero (black) to the recorded hologram until a pixel number of $2^n \times 2^n$ is reached. This *zero padding* does not distort the computation, it only causes a smoothing or interpolation of the reconstructed image.

3.4.4 Recording Set-ups

In this chapter typical set-ups used in Digital Holography are discussed and analysed with respect to their spatial frequency limitations. In figure 3.19(a) a plane reference wave according to Eq. (3.3) is used, which propagates perpendicularly to the CCD. The object is located unsymmetrical with respect to the centre line. This set-up is very simple, but the space occupied by the object is not used effectively. The maximum angle between a ray emitted from the edge of a quadratic object with side length L to the opposite edge of the CCD with side length $N\Delta x$ is (distance x explained in figure 3.19):

$$\theta_{\max} \approx \frac{x}{d_{\min}} = \frac{\sqrt{\frac{5}{4}}\left(L + N\Delta x\right)}{d_{\min}} \tag{3.52}$$

The corresponding minimum object distance d_{min} is calculated by equating this expression with the approximation for $\theta_{\max}$ in Eq. (3.51):

$$d_{\min} = \sqrt{\frac{5}{4}}\frac{2\Delta x}{\lambda}\left(L + N\Delta x\right) = \sqrt{5}\frac{\Delta x}{\lambda}\left(L + N\Delta x\right) \tag{3.53}$$

In figure 3.19(b) the plane reference wave is coupled into the set-up via a beam splitter. This allows positioning the object symmetrically, i.e. objects with larger dimensions can be recorded at a given distance d. The minimum object distance is:

$$d_{\min} \approx \frac{x}{\theta_{\max}} = \sqrt{2}\,\frac{\Delta x}{\lambda}\left(L + N\Delta x\right) \tag{3.54}$$

However, the DC term is in the centre of the reconstructed image and has to be suppressed by the procedures described in chapter 3.3.1.

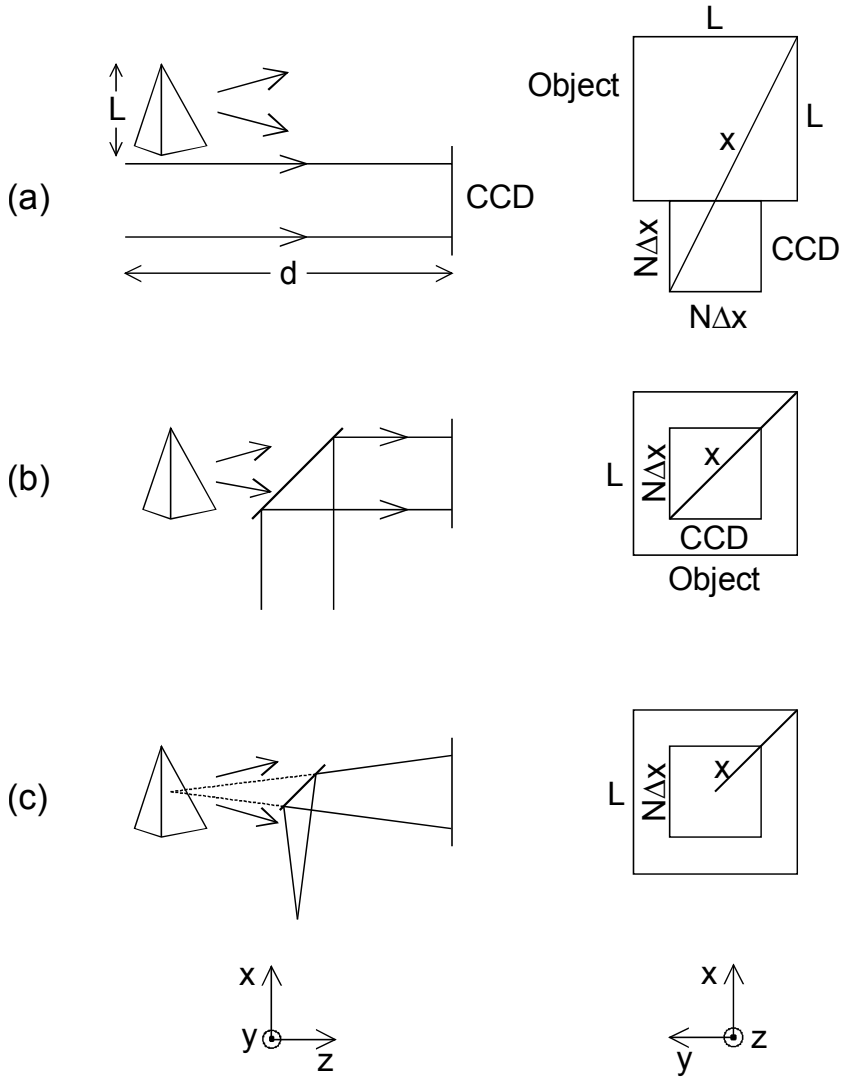

Fig. 3.19. Recording set-ups. Left: side views; Right: top views seen from CCD

Figure 3.19(c) shows a set-up for lensless Fourier holography. The spherical reference wave is coupled into the set-up via a beam splitter in order to have the source point in the object plane. The minimum object distance is:

$$d_{\min} = \sqrt{2}\frac{\Delta x}{\lambda}L \tag{3.55}$$

With the lensless Fourier set-up the shortest object distance can be chosen. On the other hand numerical refocusing is not possible with this geometry (see chapter 3.2.3).

For all set-ups the maximum spatial frequency has to be adapted very carefully to the resolution of the CCD. If too high spatial frequencies occur, the contrast of the entire hologram decreases or, in the extreme case, it vanishes totally. Therefore it has to be ensured that the spatial frequency requirements are kept in any case. This can be achieved in practice by apertures, which restrict the light propagation laterally.

If objects with dimensions larger than a few centimetres shall be recorded by CCD's, the recording distance d increases up to several meters. This is not feasible in practice. That is why set-ups have been developed to reduce the object angle to values with a resolvable spatial frequency spectrum [118, 139]. An example of such a set-up is shown in figure 3.20. A divergent lens is arranged between the object and the target. This lens generates a reduced virtual image of the object in a distance d'. The wave field emerging from this virtual image is superimposed with the reference wave and the resulting hologram is recorded. The maximum spatial frequency is lower compared to a hologram, which is recorded without object reduction.

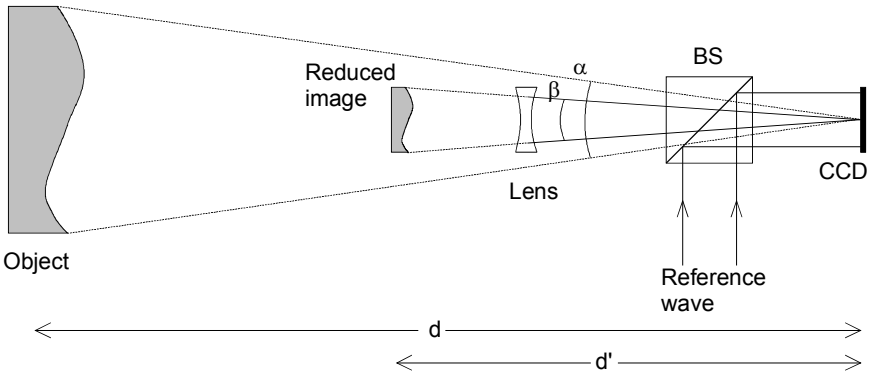

Fig. 3.20. Recording geometry for large objects

3.4.5 Stability Requirements

A stable optical set-up is necessary for digital as well as for conventional holography. Any change in the optical path difference between the interfering beams will result in a movement of the fringes and reduced contrast in the hologram. In practice the path variations should not exceed 1/4, better 1/10 of a wavelength during hologram exposure. It is therefore recommended to mount the optical set-up on a vibration isolated table, or to use a pulsed laser as light source. In contrast to conventional holography disturbances due to vibrations are visible in Digital Holography even in the recording process: The hologram visible on the monitor of the recording system appears flat and has a low modulation. This is an easy way to monitor the stability of the set-up against vibrations.