Lect. 5. Methods for Encoding CGH for Recording on Physical Media

5.1. Methods for hologram encoding for amplitude media.

Detour phase method for recording holograms on binary media.

Historically, the first method for recording computer-generated holograms was proposed by A.Lohmann and his collaborators for amplitude-only binary media ([1,2]). In this method, an elementary cell of the medium is allocated for reproducing the amplitude and phase of each sample of a mathematical hologram. The modulus of the complex number is represented by the size of the opening (aperture) in the cell and the phase - by the position of the opening within the cell.

All the cells corresponding to mathematical hologram samples are arranged over a regular (usually, rectangular) raster. A shift of the aperture by $\Delta\tilde{\xi}$ in a given cell with respect to its raster node corresponds to a phase detour for this cell equal to $\left(2\pi\Delta\tilde{\xi}\cos\theta\right)/\lambda$ for hologram reconstruction at an angle θ to the system's optical axis perpendicular to the hologram plane (Fig.5.1), where λ is wavelength of the illumination light used for hologram reconstruction. The use of the aperture spatial shift for representing complex number phase is known as the ***detour phase method***.

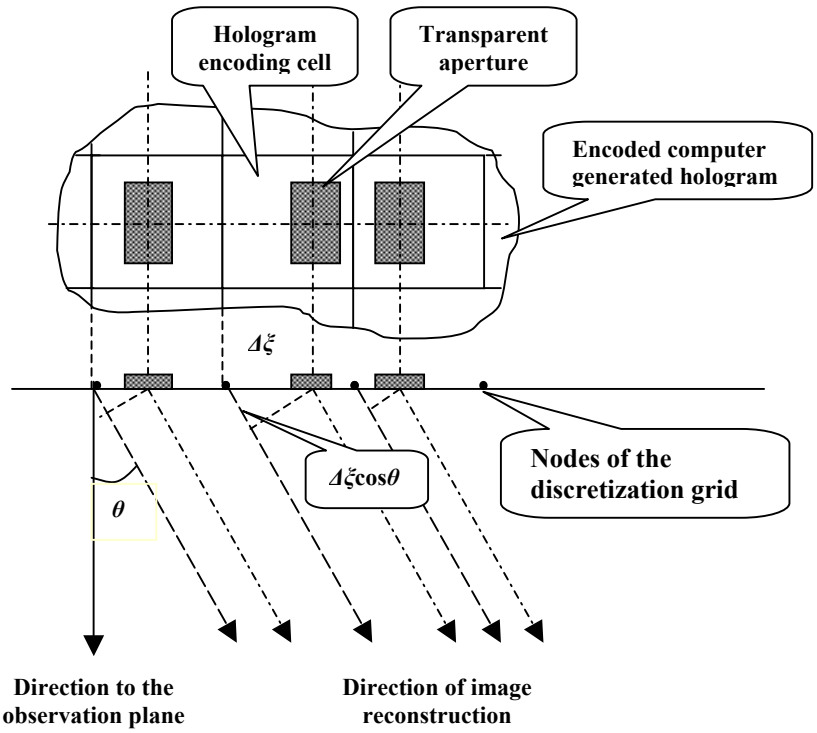


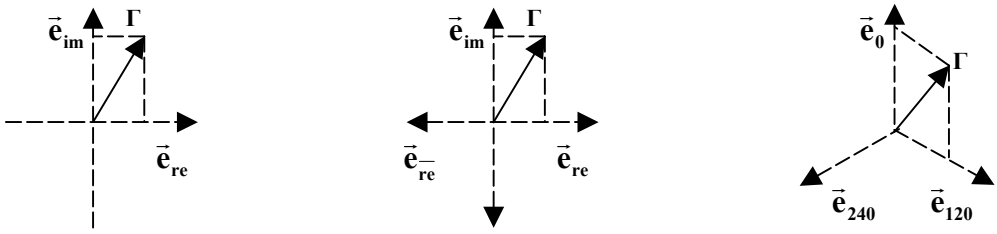
Figure 5.1. Detour phase method for coding phase shift by spatial shift of the transparent aperture

As was already noted, only spatial degrees of freedom are used for recording in binary media; therefore, the number of binary medium cells should exceed the number of hologram samples by a factor equal to the product of the amplitude and phase quantization levels. This product may run into several tens or even hundreds.

The low effectiveness of using the degrees of freedom of the hologram carrier is the major drawback of the binary hologram method. However many modern technologies such as lithography and laser and electron beam pattern generators provide much higher number of spatial degrees of freedom that the required number of hologram samples. Therefore this drawback is usually disregarded and such merits of binary recording as simpler recording and copying technology of synthesized holograms and the possibility of using commercially available devices made it the most widespread. Numerous modifications of the method are known, oriented to different types of recording devices.

Orthogonal, bi-orthogonal and 2-D simplex encoding methods

In an additive representation, the complex numbers re regarded as vectors in the complex plane in orthogonal coordinate representing real and imaginary parts of the numbers. These vectors can be represented as a sum of several components. For recording on amplitude-only media these components should be assigned standard directions on a complex plane (phase angle) and controllable length (amplitude).



a) b c)

Figure 5.2. Additive representation of complex numbers over orthogonal basis (a), over biorthogonal basis (b), over 2-D simplex (c) and as a sum of two equal length vectors (d).

The simplest case is the representation of a vector $\mathbf{\Gamma}$ by its orthogonal components, say, real Γ_{re} and imaginary Γ_{im} parts:

$$\mathbf{\Gamma} = \Gamma_{re}\mathbf{e}_{re} + \Gamma_{im}\mathbf{e}_{im}, \qquad (5.1.1)$$

where $\mathbf{e}_{re}$ and $\mathbf{e}_{im}$ are orthogonal unit vectors. When recording a hologram, the phase angle between the orthogonal components may be encoded by the detour phase method, Γ_{re} and Γ_{im} being recorded into neighboring hologram resolution cells neighboring in a raster rows as shown in Fig.5.2, a). The reconstructed image will be then observed at an angle θ_ξ to the axis defined as follows:

$$\Delta\xi \cos\theta_\xi = \lambda/4. \qquad (5.1.2)$$

For recording negative values of Γ_{re} and Γ_{im} a constant positive bias may be added to recorded values. We will refer to this method as to ***orthogonal encoding*** method.

Since two resolution elements are used here for recording one hologram sample, such a hologram has double redundancy. With such encoding and recording of a hologram, one must take into account that the optical path differences $\lambda/2$ and $3\lambda/4$

will correspond to the next pair of hologram resolution cells (see Fig.5.3, a); that is, values of Γ_{re} and Γ_{im} for each odd sample of the mathematical hologram should be recorded with opposite signs.

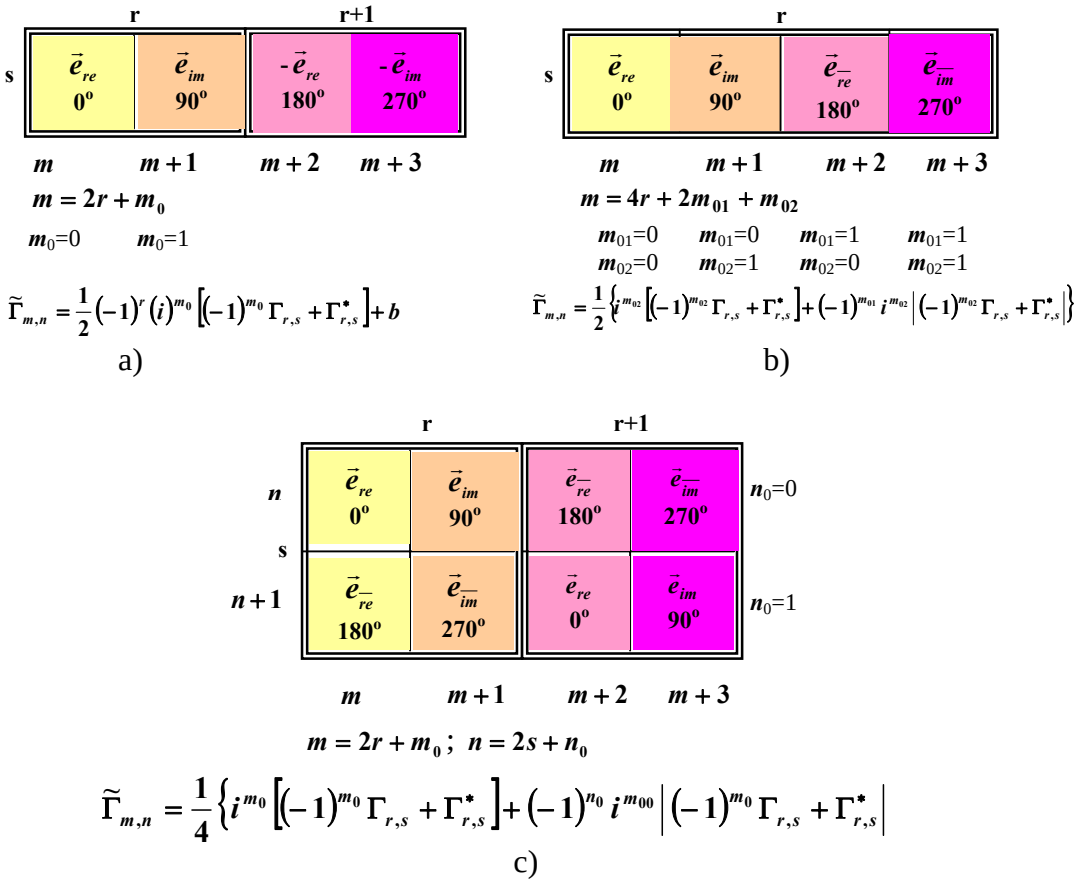


Figure 5.3. Hologram encoding by decomposition of complex numbers in orthogonal (a) and biorthogonal (b, c) bases. Hologram encoding cells corresponding to one hologram sample is outlined with a double line.

This encoding technique may be described formally as follows. Let $\{r,s\}$ be indices of mathematical hologram samples $\{\Gamma_{r,s}\}$, $r = 0,1,...,M-1$, $\{m,n\}$ be indices of the recording medium resolution cells, $\{\tilde{\Gamma}_{r,s}\}$ be samples of the encoded hologram ready for recording and let index m is counted as a two digit number:

$$m = 2r + m_0,\ m_0 = 0,1, \tag{5.1.3}$$

and index n is counted as $n = s$. The encoded hologram can then be written as

$$\tilde{\Gamma}_{m,n} = \tilde{\Gamma}_{2r+m_0,s} = \frac{1}{2}(-1)^r (i)^{m_0} \left[(-1)^{m_0} \Gamma_{r,s} + \Gamma^*_{r,s}\right] + b, \tag{5.1.4}$$

where b is a positive bias constant needed to secure that recording values are non-negative.

In reconstruction of computer generated holograms recorded with a constant bias, substantial part of energy of the reconstructing light beam is not used for image reconstruction and goes to the zero-order diffraction spot. One can avoid the constant biasing for recording Γ_{re} and Γ_{im} by allocating, for recording one hologram sample, four neighboring in sampling grid medium resolution cells rather then two. This arrangement is shown in Fig.5.3, b). With a reconstruction angle as defined by Eq.(5.1.2), phase detours $\mathbf{0}$, $\pi/2$, π and $3\pi/2$ correspond to them. Therefore, cells in each group of four cells allocated for recording of one sample of the mathematical hologram should be written in the following order: $\left(\Gamma_{re}+|\Gamma_{re}|\right)/2$; $\left(\Gamma_{im}+|\Gamma_{im}|\right)/2$; $\left(|\Gamma_{re}|-\Gamma_{re}\right)/2$; $\left(|\Gamma_{im}|-\Gamma_{im}\right)/2$. One can see that in this case all the recorded values are nonnegative. This method represents a vector in the complex plane in a ***biorthogonal basis*** ($\vec{\mathbf{e}}_{re}, \vec{\mathbf{e}}_{\overline{re}}, \vec{\mathbf{e}}_{im}, \vec{\mathbf{e}}_{\overline{im}}$):

$$\Gamma=\frac{1}{2}\left\{\left(\Gamma_{re}+|\Gamma_{re}|\right)\vec{\mathbf{e}}_{re}+\left(\Gamma_{re}-|\Gamma_{re}|\right)\vec{\mathbf{e}}_{\overline{re}}+\left(\Gamma_{im}+|\Gamma_{im}|\right)\vec{\mathbf{e}}_{im}+\left(\Gamma_{im}-|\Gamma_{im}|\right)\vec{\mathbf{e}}_{\overline{im}}\right\}. \qquad (5.1.5)$$

In terms of indices (m,n) of recording medium resolution cells counted as

$$m=4r+2m_{01}+m_{02};\ r=0,1,...,M-1;\ m_{01}=0,1;\ m_{02}=0,1;\ n=s, \qquad (5.1.6)$$

this encoding method may be described as follows:

$$\tilde{\Gamma}_{m,n}=\tilde{\Gamma}_{4r+2m_{01}+m_{02},s}=\frac{1}{2}\left\{i^{m_{02}}\left[(-1)^{m_{02}}\Gamma_{r,s}+\Gamma^{*}_{r,s}\right]+(-1)^{m_{01}}i^{m_{02}}\left|(-1)^{m_{02}}\Gamma_{r,s}+\Gamma^{*}_{r,s}\right|\right\}. \qquad (5.1.7)$$

In hologram encoding by this method, it is required, that the size of the medium resolution cell in one direction be four times smaller than that in the perpendicular one. In this way the proportions of the reconstructed picture can be preserved. A natural way to avoid this anisotropy is to allocate pairs of resolution cells in two neighboring raster rows for each sample of the mathematical hologram as it is shown in Fig.5.3, c); that is, to record the hologram samples according to the following relationship:

$$\tilde{\Gamma}_{m,n}=\tilde{\Gamma}_{2r+m_0,2s+n_0}=\frac{1}{4}\left\{i^{m_0}\left[(-1)^{m_0}\Gamma_{r,s}+\Gamma^{*}_{r,s}\right]+(-1)^{n_0}i^{m_{00}}\left|(-1)^{m_0}\Gamma_{r,s}+\Gamma^{*}_{r,s}\right|\right\}, \qquad (5.1.8)$$

where indices (m,n) are counted as two digit numbers:

$$m=2r+m_0;\ n=2s+n_0;\ m_0=0,1;\ n_0=0,1. \qquad (5.1.9)$$

With this encoding method, images are reconstructed along a direction making angles θ_ξ and θ_η with axes (ξ,η) in the hologram plane defined by the equations:

$$\Delta\xi\cos\theta_\xi=\lambda/2;\ \Delta\eta\cos\theta_\eta=\lambda/2. \qquad (5.1.10)$$

Representation of complex numbers in the bi-orthogonal basis of Eq.(13.2.6) is redundant because two of the four components are always zero. This redundancy is

reduced in the representation of complex numbers with respect to a ***two-dimensional simplex*** ($\vec{\mathbf{e}}_0, \vec{\mathbf{e}}_{120}, \vec{\mathbf{e}}_{240}$) (Fig.5.2, c):

$$\mathbf{\Gamma} = \Gamma_0 \vec{\mathbf{e}}_0 + \Gamma_1 \vec{\mathbf{e}}_{120} + \Gamma_2 \vec{\mathbf{e}}_{240}. \qquad (5.1.11)$$

Similarly to the bi-orthogonal basis, this basis is also not linearly independent because

$$\vec{\mathbf{e}}_0 + \vec{\mathbf{e}}_{120} + \vec{\mathbf{e}}_{240} = \mathbf{0}, \qquad (5.1.12)$$

and its redundancy is exploited to insure that components $\{\Gamma_0, \Gamma_1, \Gamma_2\}$ of complex vectors be non-negative.

Two versions of hologram coding by the two-dimensional simplex are known. In the first version, complex vectors in the complex plane are represented as the sum of two components directed along those of three vectors ($\vec{\mathbf{e}}_0, \vec{\mathbf{e}}_{120}, \vec{\mathbf{e}}_{240}$), which confine the third part of the plane where this vector is situated ([3]). It follows that of the three numbers $\{\Gamma_0, \Gamma_1, \Gamma_2\}$ defining vector $\mathbf{\Gamma}$ through Eq.(5.1.11), two are always non-negative and are projections of the vector $\mathbf{\Gamma}$ on the corresponding basis vectors, and the third one is zero. The following relations may be obtained for components $\{\Gamma_0, \Gamma_1, \Gamma_2\}$ from this condition

$$\begin{aligned}
\Gamma_0 &= \left[(1 + signA)(B - |B|) + (1 - signA)(C + |C|)\right] / 2\sqrt{3}; \\
\Gamma_1 &= \left[(1 + signC)(A - |A|) + (1 - signC)(B + |B|)\right] / 2\sqrt{3}; \\
\Gamma_0 &= \left[(1 + signB)(C - |C|) + (1 - signB)(A + |A|)\right] / 2\sqrt{3}, \qquad (5.1.13)
\end{aligned}$$

where

$$\begin{aligned}
signX &= X / |X| \\
A &= (\Gamma + \Gamma^*)/2; \\
B &= \left[\Gamma \exp(i2\pi/3) + \Gamma^* \exp(-i2\pi/3)\right]/2 = \mathrm{Re}\{\Gamma \exp(i2\pi/3)\};; \\
C &= \left[\Gamma \exp(-i2\pi/3) + \Gamma^* \exp(i2\pi/3)\right]/2 = \mathrm{Re}\{\Gamma \exp(-i2\pi/3)\},, \qquad (5.1.14)
\end{aligned}$$

and is the real part of the complex value.

The second version of the method makes use of the fact that, by virtue of Eq. 5.1.12, the addition of an arbitrary constant to $\{\Gamma_0, \Gamma_1, \Gamma_2\}$ leaves Eq. 5.1.11 unchanged. In this version vector components $\{\Gamma_0, \Gamma_1, \Gamma_2\}$ are represented as

$$\{\Gamma_t = \Gamma_t^{(0)} + V\}; \quad t = 0,1,2. \qquad (5.1.15)$$

The presence of an arbitrary constant V allows to impose on $\{\Gamma_t^{(0)}\}; t = 0,1,2$ an additional constraint:

$$\Gamma_0^{(0)} + \Gamma_1^{(0)} + \Gamma_2^{(0)} = 0. \qquad (5.1.16)$$

In this case $\{\Gamma_t^{(0)}\}$ are computed as

$$\Gamma_0^{(0)} = (\Gamma + \Gamma^*)/3 = 2A/3;$$

$$\Gamma_1^{(0)} = \left[\Gamma \exp\left(-i2\pi/3\right) + \Gamma^* \exp\left(i2\pi/3\right)\right]/3 = 2C/3;$$
$$\Gamma_2^{(0)} = \left[\Gamma \exp\left(i2\pi/3\right) + \Gamma^* \exp\left(-i2\pi/3\right)\right]/3 = 2B/3 \qquad (5.1.17)$$

and constant V is chosen so as to make $\{\Gamma_0, \Gamma_1, \Gamma_2\}$ nonnegative. The bias V, evidently, may differ between different hologram samples. The set of these biases for hologram samples may be optimized so as to improve the quality of reconstructed images ([4]).

When recording holograms with this method, one may encode the phase angle corresponding to unit vectors ($\vec{e}_0, \vec{e}_{120}, \vec{e}_{240}$) by means of the mentioned above detour phase method by writing the components $\{\Gamma_0, \Gamma_1, \Gamma_2\}$ of the simplex-decomposed vector into groups of three neighboring hologram resolution cells in the raster as it is shown in Fig.5.4 , a), b) and c)). The latter two arrangements are more isotropic than the first one. Coordinate scale ratio for arrangement b) is 2:1.5 and for arrangement c) is $3\sqrt{3}:4$ whereas for arrangement a) it is 3:1.

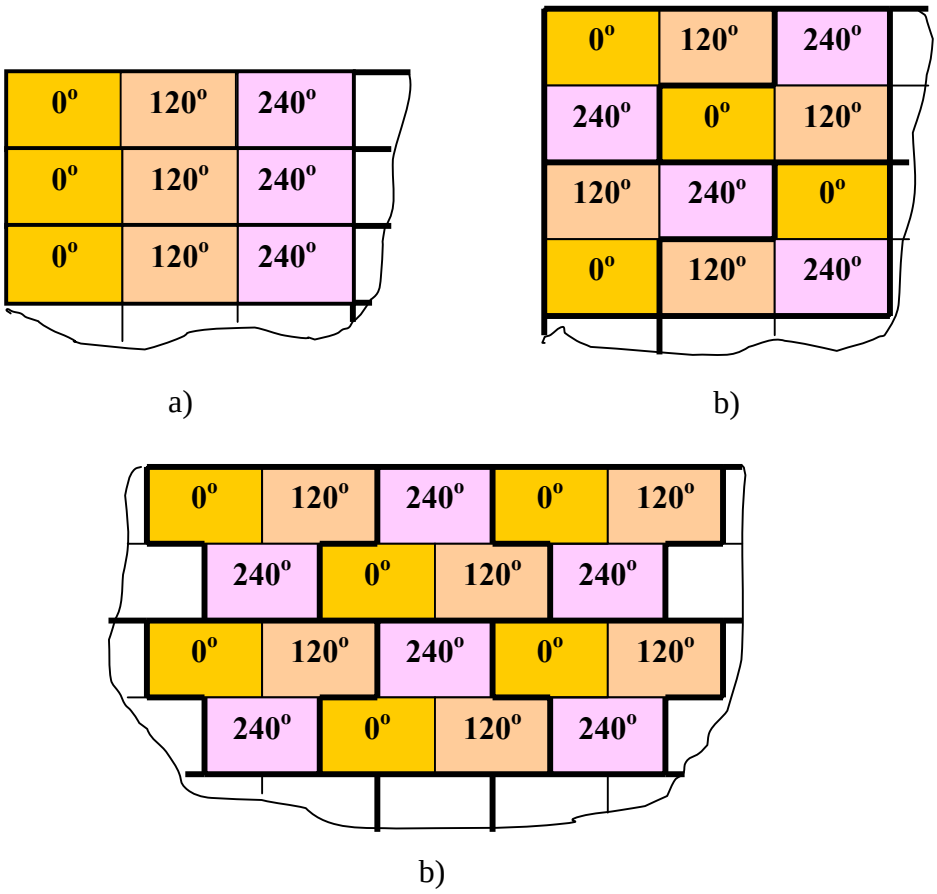


Figure 5.4. Three methods of arrangement of recording medium resolution cells for hologram encoding by decomposition of complex numbers with respect to a 2-D simplex. Triples of medium resolution cells used for recording one hologram sample are outlined with bold lines.

For holograms recorded by this method according to the arrangement of Fig. 13-8, a), images are reconstructed at angles $\theta_\xi = \arccos\left(\lambda/3\Delta\xi\right)$ to the axis ξ coinciding with the direction of the hologram raster rows and $\theta_\eta = 0$ to the perpendicular axis. For the

arrangements of Fig. 5.4, b) and c) the reconstruction angles are, respectively, $\theta_\xi = \arccos(\lambda / 3\Delta\xi)$, $\theta_\eta = \arccos(2\lambda / 3)$ and $\theta_\xi = \arccos(4\lambda / 3\Delta\xi)$, $\theta_\eta = \arccos(2\lambda / 3)$. One can show by using Eq. 5.1.13 that the following formula, which relates values written in recording medium resolution cells with indices (m,n) with samples $\{\Gamma(r,s)\}$ of the mathematical hologram, corresponds to the first version of the coding method:

$$\widetilde{\Gamma}_{m,n} = \widetilde{\Gamma}_{3r+m_0,s} = \frac{1}{2\sqrt{3}}\left[\left(1 + sign\left\{\mathrm{Re}\left[\Gamma_{r,s}\exp(-i2\pi m_0/3)\right]\right\}\right)\times\right.$$
$$\left\{\mathrm{Re}\left[\Gamma_{r,s}\exp(-i2\pi m_0/3)\right] - \left|\mathrm{Re}\left[\Gamma_{r,s}\exp(-i2\pi m_0/3)\right]\right|\right\} +$$
$$\left(1 - sign\left\{\mathrm{Re}\left[\Gamma_{r,s}\exp(-i2\pi m_0/3)\right]\right\}\right)\times$$
$$\left.\left\{\mathrm{Re}\left\{\Gamma_{r,s}\exp\left[i2\pi(m_0+1)/3\right]\right\} + \left|\mathrm{Re}\left\{\Gamma_{r,s}\exp\left[i2\pi(m_0+1)/3\right]\right\}\right|\right\}\right]. \qquad (5.1.18)$$

where horizontal and vertical indices $\{m,n\}$ (Fig. 5.4, a)) are counted as: $m = 3r + m_0$, $m_0 = 0,1,2$, $n = s$ and $\mathrm{Re}(X)$ is real part of X.

The above described encoding methods based on additive representation of complex vectors are applicable for recording on amplitude-only media, both continuous tone and binary. In the latter case, projections of the complex number on basic vectors are represented by varying the size of the transparent aperture in each of the appropriate resolution cells.

"Symmetrization" method for continuous tone amplitude SLM.

In hologram recording on amplitude media, the main problem is recording the phase component of hologram samples. The ***symmetrization method*** ([5]) offers a straightforward solution of this problem for recording Fourier holograms. The method assumes that prior to hologram synthesis, the object be symmetrized, so that, according to properties of the Discrete Fourier transform, its Fourier hologram contains only real valued samples and, thus, may be recorded on amplitude-only media. As real numbers may be both positive and negative, holograms should be recorded in the amplitude-only medium with a constant positive bias, making all the recorded values positive.

For SDFT as a discrete representation of the integral Fourier transform, symmetry properties require symmetrization through a rule that depends on the shift parameters. For example, with integer $2u$ and $2v$ the symmetrization rule for the object wave front specified by the array of its samples $\{\overline{A}_{k,l}\}$, $k = 0,1,...,N_1 - 1$, $l = 0,1,...,N_2 - 1$ becomes

$$\widetilde{\overline{A}}_{k,l} = \begin{cases} \overline{A}_{k,l}, 0 \le k \le N_1 - 1;\ 0 \le l \le N_2 - 1 \\ \overline{A}_{2N-k,l},\ N_1 \le k \le 2N_1 - 1;\ 0 \le l \le N_2 - 1 \end{cases}, \qquad (5.1.19)$$

which implies symmetrization by object duplication. In doing so, the number of samples of the object and, correspondingly, its Fourier hologram, is twice that of the original object. It is this two-fold redundancy that enables one to avoid recording the phase component. Symmetrization by quadruplicating is possible as well. This

consists in symmetrizing the object according to the rule of Eq. 5.19 with respect to both indices k and l. Hologram redundancy becomes in this case four-fold. Holograms of symmetrized objects are also symmetrical and reconstruct duplicated or quadruplicated objects depending on the particular symmetrization.

Note that duplication and quadruplication do not imply a corresponding increase in computer time for execution of SDFT at the hologram calculation step because, for computation of SDFT, one may use combined algorithms making use of signal redundancy for accelerating computations ([5, 6]).

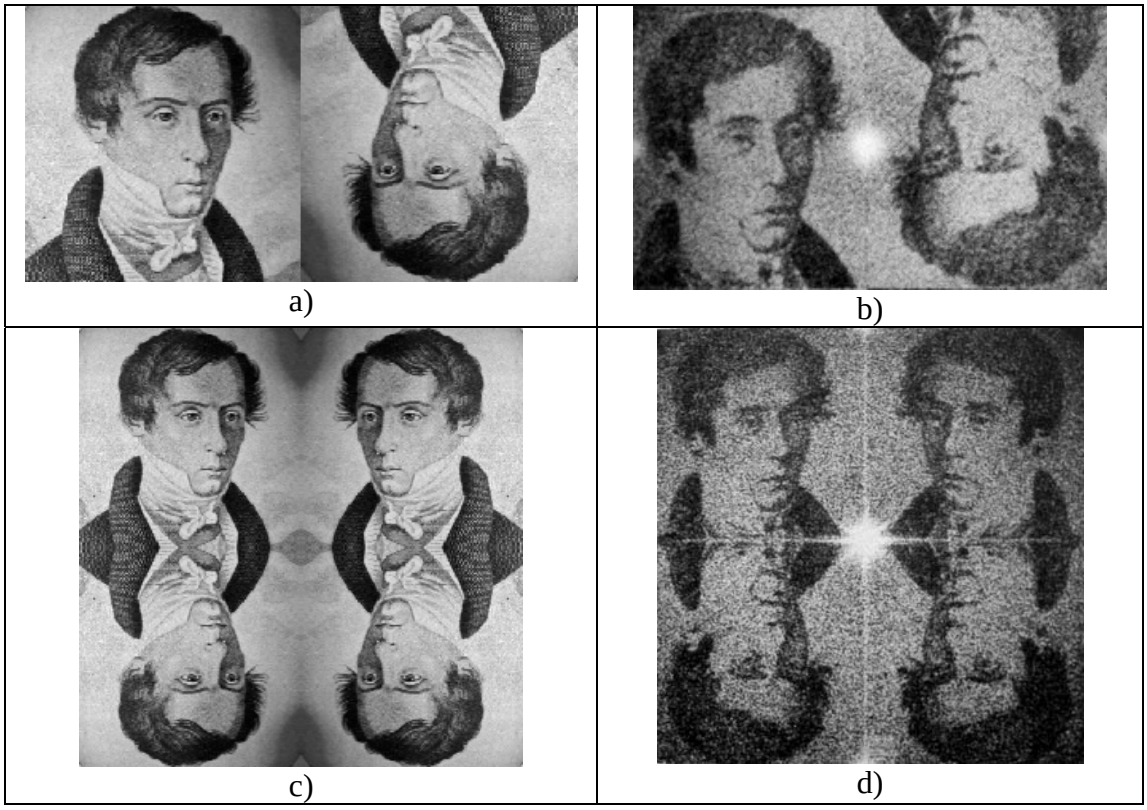


Fig. 5.3. Symmetrization method for recording continuous tone computer-generated holograms: a), b) – symmetrization by duplication: symmetrized images and an example of image optically reconstructed from computer-generated hologram; c), d) same for symmetrization by quadruplicating.

5.2. Methods for recording computer-generated holograms on phase media.

Double-phase and multiple-phase methods.

For recording on phase media, additive representation of complex vectors assumes complex vector representation as a sum of complex vectors of a standard length. Two versions of such encoding are known: double phase method and multiple phase method.

In the ***double-phase method*** ([7]), hologram samples are encoded as

$$\Gamma = |\Gamma| \exp(i\varphi) = A_0 \exp(i\varphi_1) + A_0 \exp(i\varphi_2), \qquad (5.2.1)$$

where (φ_1, φ_2) are component phase angles defined by the following equations:

$$\varphi_1 = \varphi + \arccos\left(|\Gamma|/2A_0\right) ; \ \varphi_2 = \varphi - \arccos\left(|\Gamma|/2A_0\right), \qquad (5.2.2)$$

that can be easily derived from the geometry of the method illustrated in Fig. 5. 4.

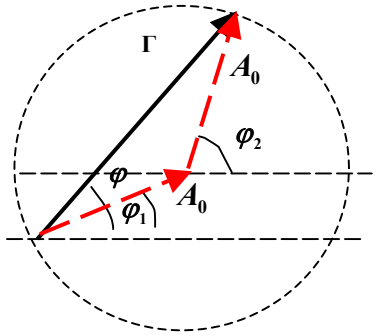


Fig. 5. 4. Complex vector represented as a sum of two vectors of equal length

Two neighboring medium resolution cells should be allocated for representing two vector components. Formally, the encoded in this way hologram may be written as

$$\tilde{\Gamma}_{m,n} = \tilde{\Gamma}_{2r+m_0,s} = A_0 \exp\left\{i\left[\varphi_{r,s} + (-1)^{m_0} \arccos\left(|\Gamma_{r,s}|/2A_0\right)\right]\right\}, \qquad (5.2.3)$$

where $\left\{\Gamma_{r,s} = |\Gamma_{r,s}| \exp\left(i\varphi_{r,s}\right)\right\}$ are samples of the mathematical hologram, indices of the recording medium resolution cells (m,n) are counted as $m = 2r + m_0$, $m_0 = 0,1$, $n = s$ and A_0 is found as half of maximal value of $\left\{|\Gamma_{r,s}|\right\}$.

For hologram encoding with the double-phase method, images are reconstructed in a direction normal to the hologram plane, because the optical path difference of rays passing along this direction through the neighboring resolution cells is zero. This holds, however, only for the central area of the image through which the optical axis of the reconstruction system passes. In the peripheral areas of the image, some phase shift between the rays appears, thus leading to distortions in the peripheral image. This will be discussed in more details in Lect. 6.

The method can also be used for recording on amplitude binary media with phase-detour encoding of phases. Two versions of such an implementation of the double-phase method were suggested ([7]). In the first version, two elementary cells of the binary medium are allocated to each of the two component vectors, their phases being coded by a shift of the transparent (or completely reflecting) aperture along a direction perpendicular to the line connecting cell centers as it is illustrated in Fig.5.5, a). This technique exhibits above mentioned distortions due to mutual spatial shift of elementary cells. The second version reduces these distortions by means of decomposing each elementary cell into sub-cells alternating as shown in Fig.5.5, b).

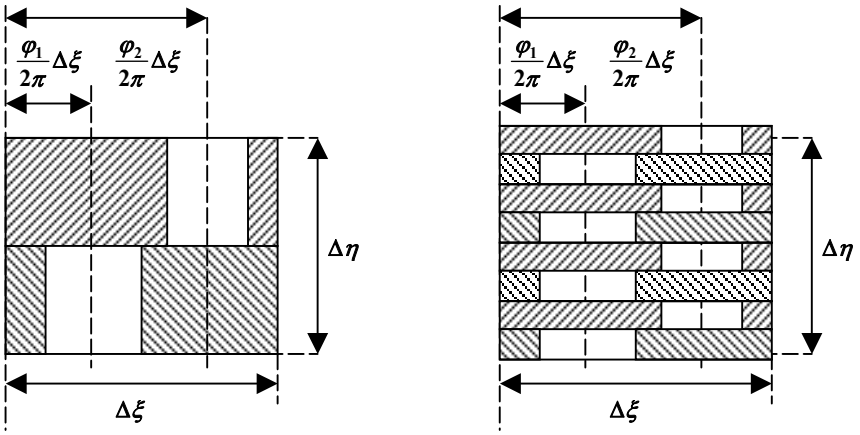


Figure 5.5. Double-phase encoding method for hologram recording on a binary phase medium: two separate cells (a); decomposition of cells into alternating sub-cells

The double-phase method is readily generalized to ***multi-phase encoding*** by vector decomposition into an arbitrary number Q of equal-length vector components:

$$\Gamma = \sum_{q=1}^{Q} A_0 \exp\left(i\varphi_q\right). \tag{5.2.4}$$

As Γ is a complex number, Eq.(5.2.4) represents two equations for Q unknown values of $\{\varphi_q\}$. They have a unique solution defined by Eq. 5.2.2 only for $Q = 2$. When $Q > 2$, $\{\varphi_q\}$ may be chosen in a rather arbitrary manner. For example, for odd Q one may choose phases $\{\varphi_q\}$ so as to make of them an arithmetic progression

$$\varphi_{q+1} - \varphi_q = \varphi_q - \varphi_{q-1} = \Delta\varphi. \tag{5.2.5}$$

In this case, the following equations can be derived for the increment $\Delta\varphi$:

$$\frac{\sin\left(Q\Delta\varphi/2\right)}{\sin\left(\Delta\varphi/2\right)} = \frac{|\Gamma|}{A_0}. \tag{5.2.6}$$

and for phase angles $\{\varphi_q\}$:

$$\varphi_q = \varphi + \left[q - \left(Q+1\right)/2\right]\Delta\varphi, \quad q = 1,2,...,Q. \tag{5.2.7}$$

For odd Q, Eq. 5.2.6 reduces to an algebraic equation of power $\left(Q-1\right)/2$ with respect to $\sin^2\left(\Delta\varphi/2\right)$. Thus, for $Q = 3$,

$$\Delta\varphi = 2\arcsin\left(\frac{\sqrt{3}}{2}\sqrt{1-|\Gamma|/3}\right). \tag{5.2.8}$$

For even Q, it is more expedient to separate all the component vectors into two groups having the same phase angles φ_1 and φ_2 that are defined by analogy with Eq.(5.2.2) as

$$\varphi_+ = \varphi + \arccos\left(|\Gamma|/QA_0\right);\ \varphi_- = \varphi - \arccos\left(|\Gamma|/QA_0\right). \qquad (5.2.9)$$

With multi-phase encoding, the dynamic range of possible hologram values may be extended because the maximal reproducible amplitude in this case is QA_0. Most interesting of the $Q > 2$ cases are those of $Q = 3$ and $Q = 4$ because the two-dimensional spatial degrees of freedom of the medium and hologram recorder may be used more efficiently through allocation of the component vectors according to Figs.5.3, c) and 5.4, b) and c).

Kinoform.

For recording on phase-only or amplitude-only media, special hologram encoding methods are required. One of the most straightforward encoding methods oriented on using phase-only media is that of ***kinoform*** ([8, 9]). Kinoform is a computer-generated hologram in which amplitude data of the mathematical hologram intended for recording are disregarded and only the hologram sample phases are recorded on a phase-only medium. Although disregarding hologram sample amplitudes results in substantial distortions such as appearance of speckle noise in the reconstructed wave front, kinoforms are advantageous in terms of the use of energy of reconstruction light because the total energy of the reconstruction light is transformed into the energy of the reconstructed wave field without being absorbed in the hologram. Moreover, the distortions may to some degree be reduced by an appropriate choice of the diffusion component of the wave field phase on the object using an iterative optimization procedure ([10]).

In synthesis of kinoform, a specially prepared array of pseudo-random numbers in the range $(-\pi, \pi)$ is generated and recorded on a phase media as a hologram phase profile. The array should be generated in such a way as to reconstruct, from the kinoform, in Fourier plane of an optical reconstruction setup a certain given spatial distribution of light energy.

Let $\{\vartheta_{k,l}\}$ be an array of numbers in the range $(-\pi, \pi)$ representing the required phase mask and $\{P(r,s)\}$ is a discrete representation of the required distribution of light intensity in the Fourier domain:

$$P(r,s) = C \cdot \left|\mathbf{DFT}\left(\{\exp(i\vartheta_{k,l})\}\right)\right|^2 \qquad (5.2.10)$$

where C is an appropriate normalizing constant. Given $\{P(r,s)\}$, $\{\vartheta_{k,l}\}$ should be found as a solution of Eq. 5.2.10. In general, the exact solution of the equation may not exist. However, one can always replace it by a solution $\{\hat{\vartheta}_{k,l}\}$, taken from a certain class of phase distributions, that minimizes an appropriate measure $D(\cdot,\cdot)$ of deviation of $C \cdot \left|\mathbf{DFT}\left\{\exp\left(i2\pi\hat{\vartheta}_{k,l}\right)\right\}\right|^2$ from $\{P(r,s)\}$:

$$\left\{\hat{\vartheta}_{k,l}\right\} = \arg\min_{\{\vartheta_{k,l}\}}\left[D\left(\{P(r,s)\}, C \cdot \left|\mathrm{DFT}\left[\left\{\exp\left(i2\pi\hat{\vartheta}_{k,l}\right)\right\}\right]\right|^2 \right)\right]. \qquad (5.2.11)$$

Pseudo-random number generator may serve as a source for realizations of the phase distributions and a solution of Eq. 5.2.11 may be searched through an iterative procedure according to a flow diagram shown in Fig. 5.6. One can also include in this process, when it is necessary, quantization of the phase distribution to a certain specified number of quantization levels to take into account properties of the phase spatial light modulator intended for recording of the kinoform.

Flow diagram of Fig. 5.6 assumes that iterations are carried out over one realization of the array of primary pseudo-random numbers. In principle, one can repeat the iterations for different realizations and then select of all obtained phase masks the one that provides the least deviation from the required distribution $\{P(r,s)\}$.

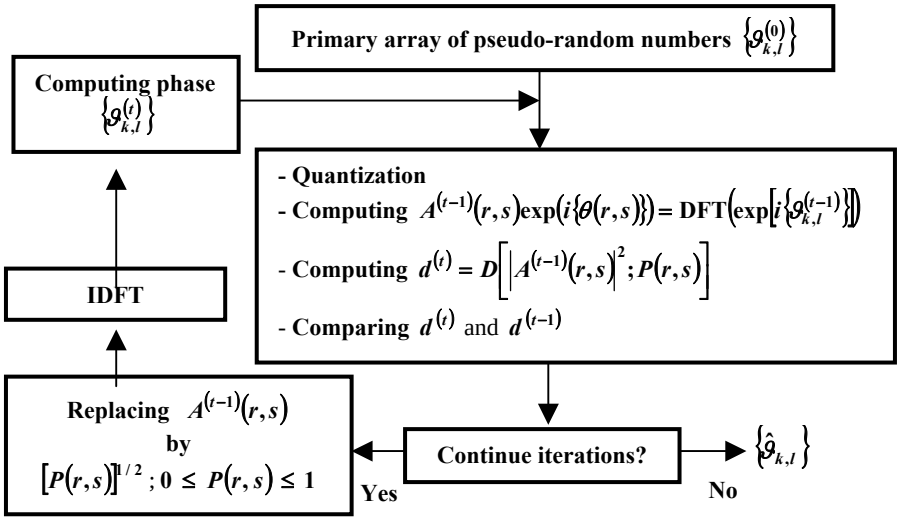


Figure 5.6. Flow diagram of an iterative algorithm for generating pseudo-random phase mask with a specified power spectrum.

Illustrative examples of images reconstructed, in computer simulation, from kinoforms synthesized by the described method with 255 and 5 phase quantization levels are shown in Fig. 5.7, b) and c) along with original image (Fig. 5.2, a) used as an amplitude mask $[P(r,s)]^{1/2}$. In this simulation, mean squared error criterion was used as a measure of deviation of the reconstructed image from the original one. On the graph of Fig. 5.7, d), one can also evaluate quantitatively the reconstruction accuracy and the speed of convergence of iterations. The graphs show that iterations converge quite rapidly, especially for coarser quantization and that the reconstruction accuracy for 255 phase quantization levels can be practically the same as that for kinoform without phase quantization. Coarse quantization of the kinoform phase worsens the reconstruction quality substantially.

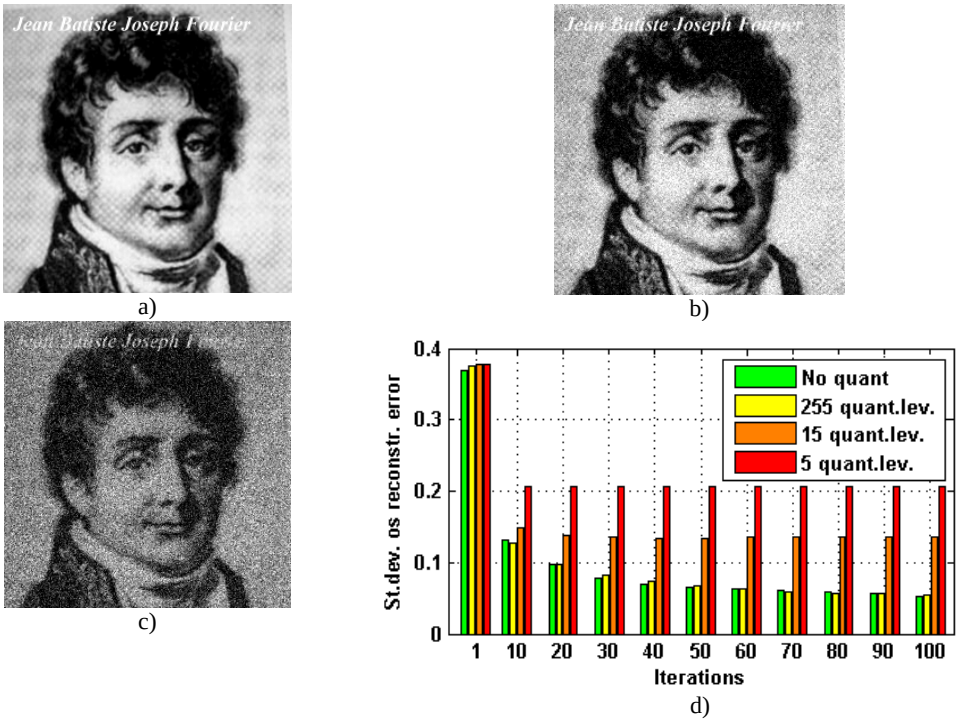


Figure 5.7. Examples of images reconstructed from kinoforms: a) original image used for the synthesis of kinoforms; b) and c) - reconstructed images for kinoforms with 255 and 5 phase quantization levels, respectively; d) plots of standard deviation of the reconstruction error as a function of the number of iterations for no quantization, 255, 15 and 5 phase quantization levels normalized to the dynamic range [0,1].

5.3. Encoding methods and introducing spatial carrier

All hologram encoding methods can be treated in a unified way as methods that explicitly or implicitly introduce to recorded holograms some form of a spatial carrier similarly to how optical holograms are recorded. This can be shown by writing Eqs.5.1.4, 5.1.7, 5.1.8, 5.1.18 and 5.2.3 in the following equivalent form explicitly containing hologram samples multiplied by those of the spatial carrier with respect to one or both coordinates:

$$\widetilde{\Gamma}_{m,n} = \widetilde{\Gamma}_{2r+m_0,s} = \mathrm{Re}\left\{\Gamma_{r,s}\exp\left(-i2\pi\frac{m}{2}\right)\right\}\Big/4 + b =$$

$$\widetilde{\Gamma}_{2r+m_0,s} = \mathrm{Re}\left\{\Gamma_{r,s}\exp\left(-i2\pi\frac{2r+m_0}{2}\right)\right\}\Big/4 + b$$

$$m = 2r + m_0;\ m_0 = 0,1;\ n = s; \qquad (5.1.4')$$

$$\widetilde{\Gamma}_{m,n} = \frac{1}{2}\mathrm{rctf}\left\{\mathrm{Re}\left[\Gamma_{r,s}\exp\left(-i2\pi\frac{m}{2}\right)\right]\right\} =$$

$$\widetilde{\Gamma}_{4r+2m_{01}+m_{02},s} = \frac{1}{2}\mathrm{rctf}\left\{\mathrm{Re}\left[\Gamma_{r,s}\exp\left(-i2\pi\frac{4r+2m_{01}+m_{02}}{2}\right)\right]\right\};$$

$$m = 4r + 2m_{01} + m_{02};\ m_{01}, m_{02} = 0,1;\ n = s; \qquad (5.1.7')$$

$$\widetilde{\Gamma}_{m,n} = \frac{1}{2}\mathbf{rctf}\left\{\mathbf{Re}\left[\Gamma_{r,s}\exp\left(-i2\pi(m+2n)/4\right)\right]\right\} =$$

$$\widetilde{\Gamma}_{2r+m_0,2s+n_0} = \frac{1}{2}\mathbf{rctf}\left\{\mathbf{Re}\left[\Gamma_{r,s}\exp\left(-i2\pi\frac{2r+m_0}{4}\right)\exp\left(-i2\pi\frac{4s+n_0}{4}\right)\right]\right\}$$

$$m = 2r + m_0;\ n = 2s + n_0, m_0, n_0 = 0,1; \qquad (5.1.8')$$

$$\widetilde{\Gamma}_{m,n} = \frac{1}{\sqrt{3}}\sum_{p=0}^{1}\left(\mathbf{hlim}\left\{\mathbf{Re}\left[\Gamma_{r,s}\exp\left(-i2\pi\frac{m+n}{3}\right)\right]\right\}\right)\times$$

$$\mathbf{rctf}\left\{\mathbf{Re}\left[\Gamma_{r,s}\exp\left(-i2\pi\frac{m+s+1/2}{3}\right)\right]\right\} =$$

$$\widetilde{\Gamma}_{3r+m_0,s} = \frac{1}{\sqrt{3}}\sum_{p=0}^{1}\left(\mathbf{hlim}\left\{\mathbf{Re}\left[\Gamma_{r,s}\exp\left(-i2\pi\frac{3r+m_0+s}{3}\right)\right]\right\}\right)\times$$

$$\mathbf{rctf}\left\{\mathbf{Re}\left[\Gamma_{r,s}\exp\left(-i2\pi\frac{3r+m_0+s+1/2}{3}\right)\right]\right\};$$

$$m = 3r + m_0; m_0 = 0,1,3;\ n = s; \qquad (5.1.18')$$

$$\widetilde{\Gamma}_{m,n} = A_0\exp\left\{i\left[\varphi_{r,s} - \cos\left(2\pi\frac{m}{2}\right)\arccos\left(|\Gamma_{r,s}|/2A_0\right)\right]\right\} =$$

$$\widetilde{\Gamma}_{2r+m_0,s} = A_0\exp\left\{i\left[\varphi_{r,s} - \cos\left(2\pi\frac{2r+m_0}{2}\right)\arccos\left(|\Gamma_{r,s}|/2A_0\right)\right]\right\}$$

$$m = 2r + m_0; m_0 = 0,1;\ n = s, \qquad (5.2.3')$$

where $\mathbf{rctf}(X)$ is the "rectifier" function

$$\mathbf{rctf}(X) = \begin{cases} X, & X \geq 0 \\ 0, & X < 0 \end{cases}, \qquad (5.3.1)$$

and $\mathbf{hlim}(X)$ is the "hard-limiter" function:

$$\mathbf{hlim}(X) = \begin{cases} 1, & X \geq 0 \\ 0, & X < 0 \end{cases}. \qquad (5.3.2)$$

As one can see from these expressions, the spatial carrier has a period no greater than one half that of hologram sampling. At least two samples of the spatial carrier have to correspond to one hologram sample in order to enable reconstruction of amplitude and phase of each hologram sample through the modulated signal of the

spatial carrier. This redundancy implies that, in order to modulate the spatial carrier by a hologram, one or more intermediate samples are required between the basic ones. They may be determined by any of variety of interpolation methods.

The simplest interpolation method, the zero order, or nearest neighbor, interpolation, simply repeats samples. It is this interpolation that is implied in the recording methods described above. For instance, according to Eq.(5.1.4') each hologram sample is repeated twice for two samples of the spatial carrier, in Eq.(5.1.7') it is repeated four times for four samples, etc. Of course such an interpolation found in all modifications of the detour phase method yields a very rough approximation of hologram intermediate samples. In Lect. 6 it will be shown that the distortions of the reconstructed image caused by improper interpolation of hologram samples manifest themselves in aliasing effects.

Aliasing free interpolation can be achieved with discrete sinc-interpolation ([6]). Discrete sinc-interpolated values of the desired intermediate samples can be obtained by using, at the hologram synthesis, Shifted DFT with appropriately found shift parameters that correspond to the position of required intermediate samples of the hologram. It should be also noted that the symmetrization method may be regarded as an analog of the method of Eq.(5.1.4) with discrete sinc-interpolation of intermediate samples. In the symmetrization method, such an interpolation is secured automatically and the restored images are free of aliasing images.

Along with methods that introduce the spatial carriers implicitly, explicit introduction of spatial carriers is also possible that directly simulates optical recording of holograms and interferograms. Of the methods oriented to amplitude-only media, one can mention those of Burch ([11]):

$$\widetilde{\Gamma}_{m,n} = 1 + \left|\Gamma_{m,n}\right| \cos\left(2\pi f m + \varphi_{m,n}\right) \tag{5.3.3}$$

and of Huang and Prasada ([12])

$$\widetilde{\Gamma}_{m,n} = \left|\Gamma_{m,n}\right| \left[1 + \cos\left(2\pi f m + \varphi_{m,n}\right)\right] \tag{5.3.4}$$

where f is frequency of the spatial carrier.

Of binary-media-oriented methods, one may mention the method described in the overview [13]:

$$\widetilde{\Gamma}_{m,n} = \mathrm{hlim}\left\{\cos\left[\arcsin\left(\left|\Gamma_{m,n}\right| / A_0\right)\right] + \cos\left(2\pi f m + \varphi_{m,n}\right)\right\}, \tag{5.3.5}$$

where A_0 is maximal value of $\left|\Gamma_{m,n}\right|$.

Of the phase-media-oriented methods, we may mention that of Kirk and Jones ([14]):

$$\widetilde{\Gamma}_{m,n} = A_o \exp\left\{i\left[\varphi_{m,n} - h_{m,n}\cos\left(2\pi f m\right)\right]\right\} \tag{5.3.6}$$

where $h_{m,n}$ depends in a certain way on $\left|\Gamma_{m,n}\right|$ and on the diffraction order where the reconstructed image should be obtained. In a sense, this method is equivalent to the multi phase encoding method. When $f = 1/2$ and

$$h_{m,n} = \arccos\left(\left|\Gamma_{m,n}\right| / 2A_0\right) \qquad (5.3.7)$$

it coincides with the double phase encoding method according to Eq. 5.2.3.

5.4. Artificial diffusers in synthesis of display holograms

One problem common to all hologram recording methods is extremely high dynamic range of mathematical Fourier holograms of regular images because intensity of image low frequency components is several order or magnitude larger than that of middle frequencies and this discrepancy grows for high frequencies (Fig. 5, 8, a) and b)). At the same time, dynamic range of best spatial light modulators is only of the order $10^2 \div 10^3$. One can fit the dynamic range of the hologram and hologram recording device by a non-linear compressing signal transformation such as P-th law nonlinearity

$$OUTPUT = [abs(INPUT)]^P \, sign(INPUT) , \qquad (5.4.1)$$

where $P \le 1$ is a dynamic range compressing parameter, $abs(.)$ and $sign(.)$ are absolute value and sign of the variable. However the compressive nonlinear transformation will redistribute signal energy in favor of its high frequency components, which will result in restoration of only image contours as one can see on Fig. 5.8, d).

An alternative solution is assigning to images artificial phase component to make image spectrum almost uniform. In synthesis of holograms, one has to specify the object wave front amplitude and phase. The amplitude component is defined by the image to be displayed. The object wave front phase component is irrelevant to visual observation, although it affects very substantially object wave front spectrum spread. Therefore one can select object wave front phase distribution in such a way as to secure least possible distortions of the object's hologram due to limitation of the hologram dynamic range and quantization in the process of hologram recording. The simplest way to do this is use arrays of pseudo-random non-correlated numbers uniformly distributed in the range $[-\pi, \pi]$ as a phase component, which makes image Fourier spectrum statistically uniform over all range of spatial frequencies (see an example in Fig. 5. 8, c)). Mathematical holograms of such images are less vulnerable to quantization in the process of recording, and recorded holograms reconstruct images with fewer distortions than holograms synthesized without assigning to images pseudo-random phase distribution as one can see comparing images 5.8 d) and e).

Characteristic distortion of images reconstructed from such holograms is speckle noise caused by distortions of the hologram in the process of recording ([6]). In principle, this distortion can be minimized using above described iterative algorithm for synthesis of kinoform.

Assigning to the object wave front a pseudo-random phase component imitates diffuse properties of real object to scatter light. Non-correlated pseudo-random phase component corresponds to uniform scattering of light in all directions. One can also imitate different other types non-uniform scattering. This option is used in synthesis of holograms with programmed diffuser discussed in Lect. 8.

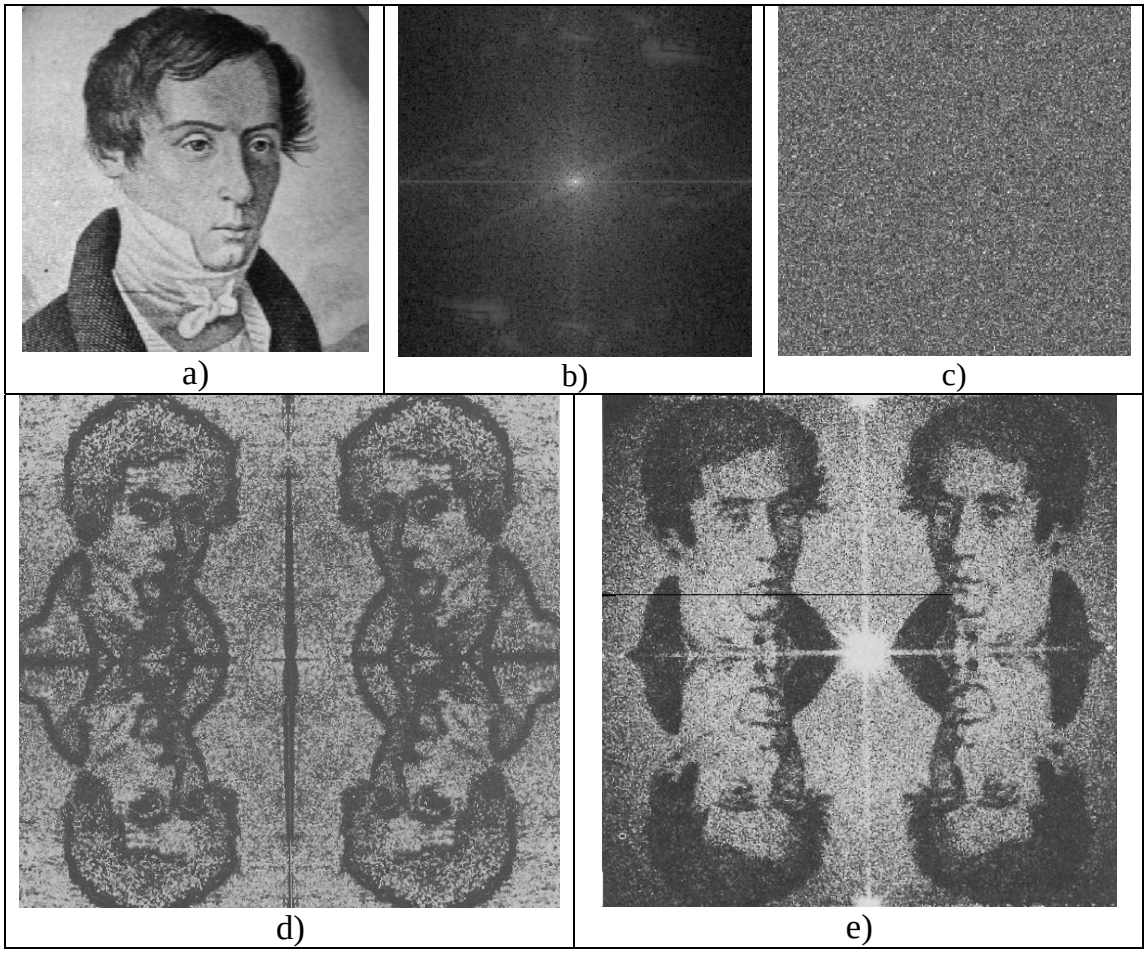


Fig. 5.8. Dynamic range of Fourier spectra of images and synthesis of hologram with artificial diffuser: a) – test image; b) – module of Fourier spectrum of the test image raised to a power 0.1 to enable its display in the compressed dynamic range; c) – module of Fourier spectrum of the same image with assigned to it a pseudo-random non-correlated phase component; d) image reconstructed from a Fourier hologram of the test image without pseudo-random phase component; e) – result of the reconstruction of a Fourier hologram of the test image with pseudo-random phase component.

References

1. B. R. Brown, A. Lohmann, Complex spatial filtering woth binary masks, *Appl. Optics*, **5**, No. 6, 967-969 (1966)
2. B. R. Brown, A. Lohmann, Computer generated binary holograms, IMB J. Res. Dev., v. 13, No. 2, 160-168 (1969)
3. C. B. Burckhardt, A simplification of Lee's method of generating holograms by computer, Appl. Opt., v. 9, No.8, p. 1949 (1970)
4. P. Chavel, J. P. Hugonin, High quality computer generated holograms: the problem of phase representation, JOSA, v. 66, p. 989 (1976)
5. L. Yaroslavskii, N. Merzlyakov, Methods of Digital Holography, Consultance Bureau, N.Y., 1980
6. L. Yaroslavsky, Digital Holography and Digital Image Processing, Kluwer Academic Publishers, Boston, 2004.
7. C. K. Hsueh, A. A. Sawchuk, Computer generated double phase holograms,, Appl. Opt., v. 17, No. 24, pp. 3874-3883, 1978
8. L. B. Lesem, P. M. Hirsch, J. A. Jordan, Kinoform, IBM Journ. Res. Dev., 13, 150, 1969
9. L. B. Lesem, P. M. Hirsch, J. A. Jordan, Computer synthesis of holograms for 3-D display, Commun. Of Associat. For Computing Machinery, v. 11, 1968, pp. 661-674
10. N. C. Gallagher, B. Liu, Method for Computing Kinoforms that Reduces Image Reconstruction Error, Applied Optics, v. 12, No. 10, Oct. 1973, p. 2328
11. J. J. Burch, Algorithms for computer synthesis of spatial filters, Proc. IEEE, v. 55, 999, 1967
12. T. S. Huang, E. Prasada, Considerations on the generation and processing of holograms by digital computers, MIT/RLE Quat. Progr. Rept., 1966, v. 81
13. W. I. Dallas, Computer generated holograms, in*: The Computer in Optical Research*, Methods and Applications, B. R. Friede, Ed., Topics in Applied Physics, v. 41, Springer, Berlin, 1980, pp. 297-367
14. J. P. Kirk, A. L. Jones, Phase-only complex valued spatial filter, JOSA, v. 61, No. 8, pp. 1023-1028, 1971