CHAPTER 5

5 Hologram Encoding for Recording on Physical Media

Abstract: The purpose of hologram encoding is converting complex valued samples of mathematical holograms obtained in the result of synthesis of computer-generated holograms into data that can be used to directly control physical parameters of recording media. This chapter overviews different hologram encoding methods oriented on different recording media that can be used for this purpose.

5.1 Methods for hologram encoding for amplitude media.

5.1.1 Detour phase method

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, illustrated in Fig. 5.1, individually controlled elementary cells of the medium are allocated, one cell per sample, for recording amplitudes and phases of samples of mathematical holograms. The modulus of the complex number representing the sample defines the size of the opening (aperture) in the cell and the phase - the position of the opening within the cell.

Fig. 5.1. Detour phase method for encoding phase shift by a spatial shift of the transparent aperture

All cells are arranged in a rectangular lattice, or sampling raster. A shift of the aperture by $\Delta\xi$ in a given cell with respect to its raster node corresponds to a phase detour for this cell equal to $\left(\frac{2\pi\Delta\xi\cos\theta}{\lambda}\right)$ for hologram reconstruction at an angle $\theta$ to the system's optical axis perpendicular to the hologram plane, where $\lambda$ is wavelength of light used for hologram reconstruction. The use of the aperture spatial shifts for representing phases of complex numbers is known as the detour phase method.
By default, samples of the mathematical hologram are defined for nodes of the hologram sampling raster. In order to determine, for each sample of the mathematical hologram, exact position and size of the opening for encoding the phase and amplitude of this sample, one should, in principle, have values of phase and amplitude of the mathematical hologram in all positions between given samples. This requires generating a continuous model of the mathematical hologram by means of interpolation of its available samples. Interpolated amplitude and phase profiles can be re-sampled in positions, in which phase detour of the opening for each hologram recording cell coincides with the phase profile of the mathematical hologram as it is illustrated in Fig. 5.2.

![Fig. 5.2. Illustration of hologram re-sampling in phase-detour method of hologram encoding. Black rectangles are openings placed in positions, where their phase detour coincides with the interpolated phase profile of the mathematical hologram.](image)

For interpolation, different methods can be used ([3]). The simplest one, the nearest neighbor interpolation, is the fastest and the least accurate. It results in considerable distortions in reconstructed images. Some of them will be illustrated in Chapt. 7, Sect 7.4. Brown and Lohmann ([1]) tried in their experiments the Newton’s interpolation method using four neighboring sampling points, which did show certain improvement of reconstructed images, though the need for further improvement remained. The most accurate interpolation method for sampled data is the discrete sinc-interpolation ([3]). More accurate resampling requires more computations. Specifically, with perfect discrete sinc-interpolation, the number of operations for computing interpolated mathematical hologram is equal to the number of operations for computing non-interpolated sampled hologram times the number of intermediate interpolated samples per each initial one. This is a computational drawback of the method.

The method also quite inefficiently utilizes spatial resolution of the recording media. As was already noted, only spatial degrees of freedom are used for recording on 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 number the amplitude and...
phase quantization levels. This product may run into several tens or even hundreds. Such low efficiency in using degrees of freedom of the recording media is probably, a major drawback of binary holograms. 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 frequently 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. A number of modifications of the method are known, oriented to different types of recording devices ([4]).

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

In an additive representation, complex numbers are regarded as vectors on the complex plane in the orthogonal coordinate system representing real and imaginary parts of the numbers as it is shown in Fig. 5.3, a).

![Fig. 5.3. Additive representation of complex numbers in orthogonal basis (a), in biorthogonal basis (b) and in 2-D simplex (c).](image)

The representation of a vector \( \vec{A} \) by its real \( \vec{A}_{re} \) and imaginary \( \vec{A}_{im} \) parts:

\[
\vec{A} = \vec{A}_{re} \, e_{re} + \vec{A}_{im} \, e_{im},
\]

(5.1.1)

where \( e_{re} \) and \( e_{im} \) are orthogonal unit vectors, is the base of the hologram encoding method, which we will refer to as the **orthogonal encoding method**. When recording a hologram, the phase angle between the orthogonal components, which is equal to \( \pi / 2 \), can be encoded by the detour phase method by means of recording \( \vec{A}_{re} \) and \( \vec{A}_{im} \) into neighboring hologram resolution cells in sampling raster rows as it is shown in Fig.5.4, a). The reconstructed image will be in this case observed at an angle \( \theta_\xi \) to this axis defined by the equation

\[
\Delta \xi \cos \theta_\xi = \lambda / 4.
\]

(5.1.2)

For recording negative values of \( \vec{A}_{re} \) and \( \vec{A}_{im} \), a constant positive bias can be added to recorded values.
With such encoding and recording of holograms, one must take into account that the optical path differences \(\lambda/2\) and \(3\lambda/4\) in the same direction under angle \(\theta_z\) will correspond to the next pair of hologram resolution cells (see Fig.5.3, a). This implies that values of \(\hat{A}_r\) and \(\hat{A}_{im}\) for each second sample of the mathematical hologram should be recorded with opposite signs.

Since two resolution elements are used here for recording one hologram sample, recorded holograms have a double redundancy with respect to the hologram recording cells.

![Diagrams](image)

**Fig. 5.4.** Hologram encoding by decomposition of complex numbers in orthogonal (a) and biorthogonal (b, c) bases. Groups of hologram encoding cells corresponding to one hologram sample are outlined with a double line. Numbers in the cells indicate phase detour angle that corresponds to the cell.

The orthogonal encoding technique can be described formally as follows. Let \( \{r,s\} \) be indices of samples \( \{\Gamma_{r,s}\} \) of the mathematical hologram, \( r = 0,1,\ldots, M - 1 \), \( \{m,n\} \) be indices of the recording medium resolution cells, \( \{\hat{\Gamma}_{r,s}\} \) be samples of the encoded hologram ready for recording. Let also index \( m \) be counted as a two digits number:

\[
m = 2r + m_0, \quad m_0 = 0,1,
\]

and index \( n \) be counted as \( n = s \). The encoded hologram can then be written as
\[
\hat{\Gamma}_{m,n} = \hat{\Gamma}_{2r+m_0, s} = \frac{1}{2} \left[ (-1)^{m_0} \Gamma_{r,s} + \Gamma^*_{r,s} \right] + b, \quad (5.1.4)
\]

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 \( \Gamma_{re} \) and \( \Gamma_{im} \) by allocating, for recording one hologram sample, four neighboring in sampling raster medium resolution cells rather than two. This arrangement is shown in Fig.5.4, b). With a reconstruction angle defined by Eq. (5.1.2), phase detours \( 0, \pi/2, \pi \) and \( 3\pi/2 \) will 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:

\[
\begin{align*}
\Gamma_{re} + \Gamma_{re} &; \\
\Gamma_{im} + \Gamma_{im} &; \\
\Gamma_{re} &; \\
\Gamma_{im} &; 
\end{align*}
\]

This encoding method can be described as follows:

\[
\hat{\Gamma}_{m,n} = \hat{\Gamma}_{4r+2m_0+m_02, s} = \frac{1}{2} \left[ (-1)^{m_0} \Gamma_{r,s} + \Gamma^*_{r,s} \right] + \frac{1}{4} \left[ (-1)^{m_0} \Gamma_{r,s} + \Gamma^*_{r,s} \right], \quad (5.1.7)
\]

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 order to preserve proportions of the reconstructed picture. 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.5.4, c). In this case encoded hologram samples \( \{\hat{\Gamma}_{m,n}\} \) are recorded according to the following relationship:

\[
\hat{\Gamma}_{m,n} = \hat{\Gamma}_{2r+m_0, 2s+n_0} = \frac{1}{4} \left[ (-1)^{m_0} \Gamma_{r,s} + \Gamma^*_{r,s} \right] + \frac{1}{4} \left[ (-1)^{m_0} \Gamma_{r,s} + \Gamma^*_{r,s} \right], \quad (5.1.8)
\]

where indices \( (m,n) \) are counted as two digits numbers:
\[ m = 2r + m_0; \quad n = 2s + n_0; \quad m_0 = 0,1; \quad n_0 = 0,1. \] (5.1.9)

With this encoding method, images are reconstructed along a direction making angles \( \theta_\xi \) and \( \theta_\eta \) with axes \((\xi, \eta)\) in the hologram plane defined by equations:

\[ \Delta \xi \cos \theta_\xi = \lambda / 2; \quad \Delta \eta \cos \theta_\eta = \lambda / 2. \] (5.1.10)

Representation of complex numbers in the bi-orthogonal basis is redundant because two of four components are always zero. This redundancy is reduced in the representation of complex numbers with respect to a two-dimensional simplex \((\mathbf{e}_0, \mathbf{e}_{120}, \mathbf{e}_{240})\):

\[ \tilde{\mathbf{A}} = \Gamma_0 \mathbf{e}_0 + \Gamma_1 \mathbf{e}_{120} + \Gamma_2 \mathbf{e}_{240}. \] (5.1.11)

as it is illustrated in Fig.5.3, c). Similarly to the bi-orthogonal basis, this basis is also redundant because

\[ \mathbf{e}_0 + \mathbf{e}_{120} + \mathbf{e}_{240} = 0. \] (5.1.12)

This redundancy is exploited to insure that components \( \{\Gamma_0, \Gamma_1, \Gamma_2\} \) of complex vectors be non-negative. Formulas for computing these components are presented, along with diagrams illustrating their derivation, in Table 5.1.

**Table 5.1.**

<table>
<thead>
<tr>
<th>( \Gamma_0 )</th>
<th>( \Gamma_1 )</th>
<th>( \Gamma_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Gamma_0 =</td>
<td>\Gamma_{im}</td>
<td>+ \frac{\sqrt{3}}{4}</td>
</tr>
<tr>
<td>( \Gamma_1 = 0 )</td>
<td>( \Gamma_0 =</td>
<td>\Gamma_{im}</td>
</tr>
<tr>
<td>( \Gamma_2 = 0 )</td>
<td>( \Gamma_0 = \frac{\sqrt{3}}{2}</td>
<td>\Gamma_{re}</td>
</tr>
<tr>
<td>( \Gamma_0 = 0 )</td>
<td>( \Gamma_1 =</td>
<td>\Gamma_{im}</td>
</tr>
</tbody>
</table>

When recording holograms with this method, one can encode the phase angle corresponding to unit vectors \((\mathbf{e}_0, \mathbf{e}_{120}, \mathbf{e}_{240})\) by means of the above-mentioned 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. 5.5, 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.

![Fig. 5.5](image)

**Fig. 5.5.** 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 by bold lines.

For holograms recorded by this method according to the arrangement of Fig. 5.5, a),
images are reconstructed at angles \( \theta_\xi = \arccos \left( \frac{\lambda}{3\Delta \xi} \right) \) to the axis \( \xi \) coinciding with
the direction of the hologram raster rows and \( \theta_\eta = 0 \) to the perpendicular axis. For the
arrangements of Fig. 5.5, b) and c) the reconstruction angles are, respectively,
\( \theta_\xi = \arccos \left( \frac{\lambda}{3\Delta \xi} \right) \), \( \theta_\eta = \arccos \left( 2\lambda / 3 \right) \) and \( \theta_\xi = \arccos \left( 4\lambda / 3\Delta \xi \right) \),
\( \theta_\eta = \arccos \left( 2\lambda / 3 \right) \).

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, coded hologram components, which are projections on
basic vectors of complex numbers representing hologram samples, are recorded by
varying the size of the transparent aperture in each of the appropriate resolution cells.
5.1.3 “Symmetrization” method

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 is symmetrized, so that, according to properties of the Discrete Fourier transform, its Fourier hologram contains only real valued samples and, thus, can be recorded on amplitude-only media. As real numbers may be both positive and negative, holograms should be recorded on amplitude-only media with a constant positive bias, making all the recorded values positive.

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

\[
\overline{\hat{A}}_{k,l} = \begin{cases} 
\overline{A}_{k,l}, & 0 \leq k \leq N_1-1; 0 \leq l \leq N_2-1 \\
\overline{A}_{2N_1-k,l}, & N_1 \leq k \leq 2N_1-1; 0 \leq l \leq N_2-1
\end{cases}.
\]

(5.1.13)

In doing so, the number of samples of the object and, correspondingly, of 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. We refer to this symmetrization method as symmetrization by duplication. It is illustrated in Fig. 5.6, a).

![Figure 5.6](image)

Fig. 5.6. Symmetrization of images for recording computer-generated holograms on amplitude media: a), b) – symmetrization by duplication: symmetrized images and an example of an image optically reconstructed from computer-generated hologram; c), d) the same for symmetrization by quadruplicating.
**Symmetrization by quadruplicating** is possible as well. It consists in symmetrizing the object according to the rule of Eq. 5.1.13 with respect to both indices k and l as it is illustrated in Fig. 5.6, b). 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 method (Figs. 5.6, c) and d)). Note that duplicating and quadruplicating do not necessarily imply a corresponding increase in computation time for execution of SDFT at the hologram calculation step because, for computing SDFT, one can use combined algorithms making use of signal redundancy for accelerating computations ([3, 6]).

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

5.2.1 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

\[ \tilde{A} = |\Gamma| \exp(i\phi) = A_0 \exp(i\phi_1) + A_0 \exp(i\phi_2), \]  

(5.2.1)

where \((\phi_1, \phi_2)\) are component phase angles defined by the following equations:

\[ \phi_1 = \varphi + \arccos\left(\frac{|\Gamma|}{2A_0}\right) \quad ; \quad \phi_2 = \varphi - \arccos\left(\frac{|\Gamma|}{2A_0}\right), \]  

(5.2.2)

that can be easily derived from the geometry of the method illustrated in Fig. 5.7.

![Fig. 5.7](image)

**Fig. 5.7.** Representation of a complex vector 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 can be written as

\[ \tilde{\Gamma}_{m,n} = \tilde{\Gamma}_{2r+m_0,2s} = A_0 \exp\left\{ i \left[ \varphi + (\arccos\left(\frac{|\Gamma|}{2A_0}\right) - (-1)^{m_0}\right] \right\}, \]  

(5.2.3)
where \( \{ \tilde{A}_{r,s} = |\tilde{A}_{r,s}| \exp(i\phi_{r,s}) \} \) are samples of the mathematical hologram, indices of the recording medium resolution cells \((m,n)\) are counted as \( m = 2r + m_0, \quad m_0 = 0,1, n = s \) and \( A_0 \) is found as half of maximal value of \( \left\{ |\tilde{A}_{r,s}| \right\} \).

For hologram encoding with the double-phase method and recorded on phase media, images are reconstructed in the 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 nonzero phase shift between the rays appears, thus leading to distortions in the peripheral image. This phenomenon will be discussed in more details in Chapt. 7.

The double phase 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.5.8, 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 it is shown in Fig.5.8, b).

![Fig. 5.8. Double-phase encoding method for hologram recording on a binary phase medium: two separate cells (a); decomposition of cells into alternating sub-cells](image)

The double-phase method is readily generalized to **multi-phase encoding** by vector decomposition into an arbitrary number \( Q \) of equal-length vector components:
\[ \tilde{A} = \sum_{q=1}^{Q} A_0 \exp(i\phi_q). \] (5.2.4)

As \( \Gamma \) is a complex number, Eq.(5.2.4) represents two equations for \( Q \) unknown values of \( \{\phi_q\} \). They have a unique solution defined by Eq. 5.2.2 only for \( Q=2 \). When \( Q > 2 \), \( \{\phi_q\} \) can be chosen in a rather arbitrary manner. For example, for odd \( Q \) one can choose phases \( \{\phi_q\} \) making an arithmetic progression:

\[ \phi_{q+1} - \phi_q = \phi_q - \phi_{q-1} = \Delta \phi. \] (5.2.5)

In this case, the following equations can be derived for the increment \( \Delta \phi \):

\[ \frac{\sin(Q\Delta \phi / 2)}{\sin(\Delta \phi / 2)} = \frac{|\Gamma|}{A_0}. \] (5.2.6)

and for phase angles \( \{\phi_q\} \):

\[ \phi_q = \phi + \left[ q - \left( Q + 1 \right) / 2 \right] \Delta \phi, \quad q = 1,2,\ldots,Q. \] (5.2.7)

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

\[ \Delta \phi = 2 \arcsin \left( \frac{\sqrt{3}}{2} \sqrt{1 - |\Gamma|^2 / 3} \right). \] (5.2.8)

For even \( Q \), it is more expedient to separate all the component vectors into two groups having the same phase angles \( \phi_1 \) and \( \phi_2 \) that are defined by analogy with Eq.(5.2.2) as

\[ \phi_+ = \phi + \arccos \left( |\Gamma| / QA_0 \right); \quad \phi_- = \phi - \arccos \left( |\Gamma| / QA_0 \right). \] (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 can be used more efficiently through allocation of the component vectors according to Figs.5.4, c) and 5.5, b) and c).
5.2.2 Kinoform

One of the most popular encoding methods oriented on using phase-only media is that of kinoform ([8]). Kinoform is a computer-generated hologram in which variations of amplitude data of the mathematical hologram are disregarded, amplitudes of all hologram samples are set to a constant and only sample phases are recorded on a phase-only medium. It is used, when approximative reconstruction of only the amplitude component of the reconstructed object wave front is required in hologram reconstruction.

In synthesis of kinoform, a specially prepared array of pseudo-random numbers in the range $(-\pi, \pi)$ is assigned for the phase component of the object wave front. The array should be generated in such a way as to reconstruct, from the kinoform, in the Fourier plane of an optical reconstruction setup a certain given spatial distribution of light intensity that reproduces the amplitude of the object wave front.

Of course, disregarding variations of hologram sample amplitudes results in distortions of the hologram which manifest themselves in appearance of speckle noise in the reconstructed wave front. However, kinoforms are advantageous in terms of the use of energy of the reconstruction light because the total energy of the reconstruction light is transformed into the energy of the reconstructed wave field without being absorbed by the hologram medium. Moreover, the distortions can to some degree be reduced by an appropriate choice of artificially introduced phase component of the object wave field using the following iterative optimization procedure ([10]).

Let $\{\theta_{k,l}\}$ be an array of numbers in the range $(-\pi, \pi)$ representing the required phase component of the object wave front and $\{|A(k,l)|\}$ be samples of the object wave front amplitude, where $\{k,l\}$ are samples’ indices. This phase component must satisfy the equation

$$|A(k,l)| = C \cdot \text{abs}\left\{\text{IDFT}\left(\exp\left(\text{angle}\left\{\text{DFT}\left(|A(k,l)|\exp\left(i\theta_{k,l}\right)\right)\right\}\right)\right)\right\}$$  \hspace{1cm} (5.2.10)

where $C$ is an appropriate normalizing constant, $\text{abs}(\cdot)$ is an operator of taking absolute value of the variable and $\text{angle}(\cdot)$ is an operator for computing the phase component of the complex variable. In general, the exact solution of the equation may not exist. However, one can always replace it by a solution $\{\hat{\theta}_{k,l}\}$, taken from a certain class of phase distributions, that minimizes an appropriate measure $D(\cdot,\cdot)$ of deviation of the result of numerical reconstruction $C \cdot \text{abs}\left\{\text{IDFT}\left(\exp\left(\text{angle}\left\{\text{DFT}\left(|A(k,l)|\exp\left(i\hat{\theta}_{k,l}\right)\right)\right\}\right)\right)\right\}$ of the kinoform from $\{|A(k,l)|\}$:
\[
\{ \hat{\theta}_{k,l} \} = \arg \min_{\{ \theta_{k,l} \}} \left[ D \left( |A(r,s)| \right), C \cdot \text{abs} \left( \text{IDFT} \left( \exp \left( \text{angle} \left( \text{DFT} \left[ A(k,l) \exp \left( i\hat{\theta}_{k,l} \right) \right] \right) \right) \right) \right) \right],
\]

where \( \theta_{r,s} \) is the phase distribution of the kinoform to a certain specified number of quantization levels (operator \( \text{quant} \cdot \) in the flow diagram) in order to take into account properties of the phase spatial light modulator intended for recording of the kinoform.

Flow diagram of Fig. 5.9 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 \( \{ |A(k,l)| \} \).

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. 5.10, b) and c) along with original image (Fig. 5.10, a) used as the amplitude mask $A(k,l)$. In this simulation, mean square error criterion was used as a measure of deviation of the reconstructed image from the original one. On the bar graph of Fig. 5.10, 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 may worsen the image intensity reconstruction quality substantially.

Fig. 5.10. 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 (error values are 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 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:
\[
\tilde{\Gamma}_{m,n} = \frac{1}{2} \text{rectf} \left\{ \text{Re} \left[ \Gamma_{s} \exp(-i2\pi \frac{m}{2}) \right] \right\} = \frac{1}{2} \text{rectf} \left\{ \text{Re} \left[ \Gamma_{s} \exp(-i2\pi \frac{4r + 2m_{01} + m_{02}}{2}) \right] \right\};
\]

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

\[
\tilde{\Gamma}_{m,n} = A_{0} \exp \left\{ i \left[ \phi_{s} - \cos \left( 2\pi \frac{m}{2} \right) \arccos \left( |\Gamma_{s}|/2A_{0} \right) \right] \right\} = \frac{1}{2} \text{rectf} \left\{ \text{Re} \left[ \Gamma_{s} \exp(-i2\pi \frac{4s + n_{0}}{4}) \exp(-i2\pi \frac{4s + n_{0}}{4}) \right] \right\};
\]

\[
m = 2r + m_{0}; \quad n = 2s + n_{0}; \quad m_{0}, n_{0} = 0, 1; \quad (5.1.8')
\]

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 must be determined by hologram interpolation.

The simplest interpolation method, the zero order, or nearest neighbor, interpolation, simply repeats samples. 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 Ch. 7 it will be shown that the distortions of the reconstructed image caused by non-perfect 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 as it will be shown in Chapt. 7, Sect. 7.3.

Along with methods that introduce 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]):

\[
\hat{\Gamma}_{m,n} = 1 + |\Gamma_{m,n}| \cos(2\pi fm + \phi_{m,n}) \tag{5.3.2}
\]

and of Huang and Prasada ([12])

\[
\hat{\Gamma}_{m,n} = |\Gamma_{m,n}| \left[ 1 + \cos(2\pi fm + \phi_{m,n}) \right] \tag{5.3.3}
\]

where \( f \) is frequency of the spatial carrier.

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

\[
\hat{\Gamma}_{m,n} = \text{hlim} \left\{ \cos \left( \arcsin \left( |\Gamma_{m,n}| / A_0 \right) \right) + \cos \left( 2\pi fm + \phi_{m,n} \right) \right\}, \tag{5.3.4}
\]

where \( A_0 \) is maximal value of \( |\Gamma_{m,n}| \).

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

\[
\hat{\Gamma}_{m,n} = A_o \exp \left\{ i \left[ \phi_{m,n} - h_{m,n} \cos(2\pi fm) \right] \right\} \tag{5.3.5}
\]

where \( h_{m,n} \) depends in a certain way on \( |\Gamma_{m,n}| \) 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

\[
h_{m,n} = \arccos \left( |\Gamma_{m,n}| / 2A_0 \right) \tag{5.3.7}
\]

it coincides with the double phase encoding method according to Eq. 5.2.3.
5.4 Artificial diffusers in the synthesis of display holograms

One problem common to all hologram recording methods is extremely high dynamic range (the ratio of maximal to minimal non-zero component) 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. This phenomenon is illustrated in Fig. 5.11, b). At the same time, the dynamic range of best spatial light modulators is only of the order $10^2 \div 10^3$. Of the same order in also the number of quantization levels in digital-to-analog converters used to control spatial light modulators. One can fit the dynamic range of the hologram and hologram recording device by a non-linear compressing signal transformation such as, for instance, $P$-th law nonlinearity ([3]).

$$\text{OUTPUT} = \left[ \text{abs} \left( \text{INPUT} \right) \right]^p \text{sign} \left( \text{INPUT} \right),$$

(5.4.1)

where $P \leq 1$ is a dynamic range compressing parameter, $\text{abs}(.)$ and $\text{sign}(.)$ 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).

![Fig. 5.11. 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 and recorded using the nonlinear compressive transform; e) – a result of reconstruction of a Fourier hologram of the test image with a pseudo-random phase component.](image_url)
An alternative solution is assigning to images artificial phase component to make image spectrum almost uniform. In the 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 the object wave front spectrum dynamic range. Therefore one can select an object wave front phase distribution in such a way as to secure least possible distortions of the object’s hologram due to the 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 homogeneous over all range of spatial frequencies (see an example in Fig. 5.11, 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 d) and e).

Distortions of the hologram in the process of encoding and recording result in distortions of reconstructed images. For holograms synthesized with the use of the artificial phase component, or diffuser, image distortions exhibit themselves as speckle noise ([3]). In principle, these distortions 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 using pseudo-random phase components with correspondingly selected correlation function. We refer to such correlated pseudo-random phase component assigned to the object wave front with a purpose of imitating pre-assigned non-uniform light scattering as to **programmed diffuser**. Using holograms with programmed diffuser for synthesis of display holograms of three-dimensional objects is discussed in Chapt. 9.

Algorithm for generating pseudo-random phase masks with pre-assigned correlation function is similar to the above-described algorithm of generating the optimized diffuser for synthesis of kinoform. Flow diagram of the algorithm is presented in Fig. 5.12. In this diagram, $\{\vartheta_{k,l}\}$ is an array of numbers in the range $(-\pi, \pi)$ representing the required phase mask and $\{P(r,s)\}$ is an array of coefficients of representation of the required distribution of light intensity in the Fourier domain, or Discrete Fourier Transform of the required correlation function of the diffuser. By definition, arrays $\{\vartheta_{k,l}\}$ and $\{P(r,s)\}$ should satisfy the equation:

$$P(r,s) = C \cdot |\text{DFT}\{\exp(i\vartheta_{k,l})\}|^2$$

where $C$ is an appropriate normalizing constant. Because, in general, the exact solution of this equation for $\{\vartheta_{k,l}\}$ given $\{P(r,s)\}$ may not exist, its approximation
is sought by means of iterative minimization of an appropriate measure $D(\cdot, \cdot)$ of deviation of $C \cdot |\text{DFT}\left\{\exp\left(i2\pi\hat{\theta}_{k,l}\right)\right\}|^2$ from $\{P(r,s)\}$:

$$\{\hat{\theta}_{k,l}\} = \arg\min_{\{\theta_{k,l}\}} \left[ D\left(\{P(r,s)\}, C \cdot |\text{DFT}\left[\exp\left(i2\pi\theta_{k,l}\right)\right]|^2\right) \right].$$  (5.4.3)

Fig. 5.12. Flow diagram of an iterative algorithm for generating pseudo-random phase masks with a specified power spectrum, or discrete Fourier transform of its autocorrelation function

References