Optica Applicata, Vol. XX, No. 2, 1990

Self-imaging with nonparabolic approximation of spherical wavefronts: an extension to the objects with nonsinusoidal amplitude transmission

K. GADOŚ, K. PATORSKI

Institute of Design of Precise and Optical Instruments, Warsaw Uniwersity of Technology, ul. Chodkiewicza 8, 02-525 Warszawa, Poland.

An analytical study of the influence of the diffraction process aberrations on the self-imaging phenomenon of nonsinusoidal amplitude transmittance objects is presented. The equations describing higher harmonics of the Fresnel diffraction field intensity distribution are derived and experimentally verified.

1. Introduction

In the recent paper [1] a simple analytical approach to the problem of self-imaging under a nonparabolic approximation of spherical wavefronts was presented. In the series expansion of an optical path the terms up to the fourth (as opposed to the second) power in the lateral coordinates were included. Grating image aberrations were calculated and discussed. In the range of small magnifications, realized at relatively short distances of the observation plane from the source and the self-imaging object, a strong contrast modulation was found. It restricts the useful area of the self-image to be utilized. The zero contrast regions form the elliptically shaped fringelike structure. A departure of the image fringes from straightness observed with the increase in the field of view was caused by the aberration of the coma of interfering diffraction orders.

In the cited paper a diffraction grating with sinusoidal amplitude transmittance was assumed as an object. Although such an assumption simplified the calculations and permitted us to draw some general conclusions, the analysis should be extended to nonsinusoidal amplitude transmittances used in practice. It is of particular importance in applications of the self-imaging phenomenon utilizing the detection of higher harmonics of the diffraction field intensity distribution.

In this paper we present such an analysis focused on the three lowest spatial harmonics of Fresnel field intensity distribution. An emphasis is put on the studies of the image plane contrast variations. Binary amplitude grating is taken as an object. An experimental verification of the principles derived is given.

2. Analysis

2.1. Derivation of a general formula

Figure 1 shows a schematic representation of the configuration under study. The considered self-imaging object is a linear diffraction grating with binary amplitude transmittance

$$T_G(x, y) = \sum_{m=-\infty}^{\infty} a_m \exp(i2\pi mx/d) \tag{1}$$

where a_m denotes the amplitude coefficient of diffraction orders and d is the grating period. An infinite lateral extension of the grating was assumed since, according to the usual practice, the observation area removed from the shadow image of the object's edges is of practical importance. Our goal is to calculate the intensity distribution in the Fresnel diffraction field behind grating G.

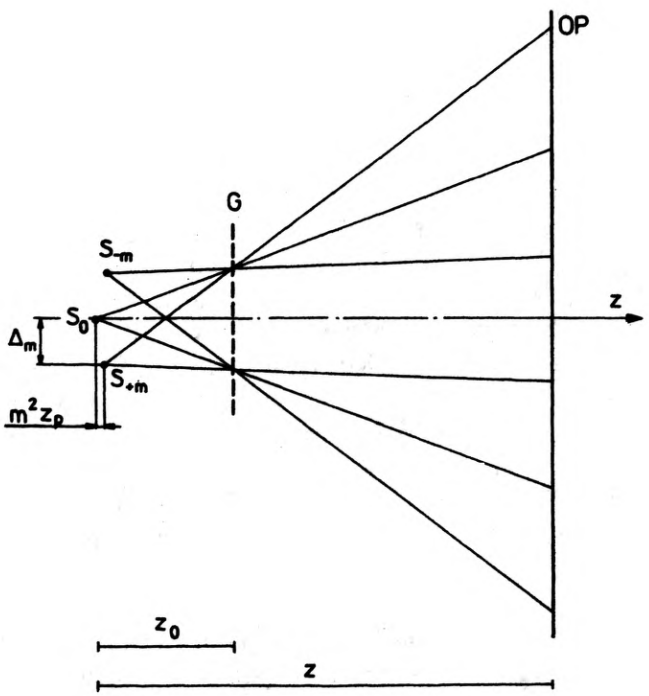

Fig. 1. The zero and the symmetrical $+n$ and $-n$ diffraction orders of a binary amplitude diffraction grating G illuminated by a point source S_0. The axial and lateral distances of the sources S_n and S_{-n} with respect to S_0 are indicated as $m^2 z_p$ and Δ_m, respectively

The distances of G and the observation plane OP from a quasi-monochromatic point source S are z_0 and z, respectively. The diffraction field in the observation plane is given by a sum of quasi-spherical wavefront diffraction orders originating

from the multiple of the sources $S_{-m}, \ldots, S_0, \ldots, S_m$. These wavefronts are described by

$$U_m = a_m \exp\{ik\alpha_m\} = a_m \exp ik\left\{\frac{(x-\Delta_m)^2+y^2}{2z} - \frac{[(x-\Delta_m)^2+y^2]^2}{8z^3} - m^2 z_p\right\} \tag{2}$$

where $|\Delta_m|$ is the distance of S_m from the optical axis

$$|\Delta_m| = mz_0 \sin\theta = mz_0(\lambda/d), \tag{3}$$

and

$$z_p = z_0\lambda^2/2d^2. \tag{4}$$

We have used the nonparabolic approximation of spherical wavefronts, i.e., taking into account the fourth (as opposed to the second) power term in the Taylor-series expansion of the optical path. In such a case the observation field is not limited to a paraxial region. When deriving Eqs. (2) and (4) the axial displacement of source S_m with respect to S_0 equal to

$$|z_0(1-\cos\theta_m)| \approx \left|z_0\left[\frac{1}{2}\left(\frac{\lambda}{d}\right)^2 + \frac{1}{8}\left(\frac{\lambda}{d}\right)^4\right]\right| \tag{5}$$

has been approximated by the first term only.

The nonparabolic approximation, corresponding to the inclusion of the fourth power term only in x and y in the Taylor series expansion, is sufficient for analysing the majority of self-imaging configurations. This term determines the main features of the introduced diffraction field aberrations, whereas the higher order terms make only small contributions. As shown in [1], this is especially true for the configuration free from the image contrast degradations.

The amplitude $E(x, y, z)$ in the observation plane is equal to

$$E(x, y, z) = \sum_{m=-\infty}^{\infty} U_m, \tag{6}$$

and the intensity distribution $I = UU^*$ becomes

$$\begin{aligned} I = a_0^2 + 2\sum_{n=1}^{\infty} a_n^2 + \sum_{N=0}^{\infty}\Bigg\{ & 2a_{N/2}^2 \cos\left[2\pi N \frac{z_0}{z}\frac{x}{d}\left(1-\frac{x^2+y^2+\delta}{2z^2}\right)\right] \\ & + 4\sum_{n=0}^{\infty} a_n a_{n+N} \cos\left[2\pi N \frac{z_0}{z}\frac{x}{d}\left(1-\frac{x^2+y^2+\delta}{2z^2}\right)\right] \\ & \times \cos\left\{2\pi[(n+N)^2-n^2]\frac{\lambda z_0}{2d^2 z}\left[z - z_0 + \frac{z_0}{2z^2}(3x^2+y^2+\delta/2)\right]\right\} \\ & + 4\sum_{n=1}^{N/2-1} a_n a_{N-n} \cos\left[2\pi N \frac{z_0}{d}\left(1-\frac{x^2+y^2+\delta}{2z^2}\right)\right] \\ & \times \cos\left\{2\pi[(N-n)^2-n^2]\frac{\lambda z_0}{2d^2 z}\left[z - z_0 + \frac{z_0}{2z^2}(3x^2+y^2+\delta/2)\right]\right\}\Bigg\} \end{aligned}$$

$$+\sum_{M=1}^{\infty}\left\{4\sum_{n=0}^{\infty}a_n a_{n+M}\cos\left[2\pi M\frac{z_0}{z}\frac{x}{d}\left(1-\frac{x^2+y^2+\delta}{2z^2}\right)\right]\right.$$
$$\times\cos\left\{2\pi[(M+n)^2-n^2]\frac{\lambda z_0}{2d^2 z}\left[z-z_0+\frac{z_0}{2z^2}(3x^2+y^2+\delta/2)\right]\right\}$$
$$+4\sum_{n=1}^{(M-1)/2}a_n a_{M-n}\cos\left[2\pi M\frac{z_0}{z}\frac{x}{d}\left(1-\frac{x^2+y^2+\delta}{2z^2}\right)\right]$$
$$\left.\times\cos\left\{2\pi\left[(M-n)^2-n^2\right]\frac{\lambda z_0}{2d^2 z}\left[z-z_0+\frac{z_0}{2z^2}(3x^2+y^2+\delta/2)\right]\right\}\right\} \tag{7}$$

where M and N denote the odd and even numbers, respectively, and $\delta = \lambda^2 z_0^2/d^2$.

Equation (7) describes a sum of intensity patterns of different spatial frequencies that represent a set of harmonics of the intensity distribution. The fringes are not straight but deformed proportionally to

$$u(x,\ y) = \frac{1}{2z^2}\,x(x^2-y^2). \tag{8}$$

As already mentioned in the cited paper [1], the deformation given by Eq. (8) corresponds to the aberration of the coma displayed in the reference beam interferometer or the spatial derivative of the spherical aberration visualized by lateral shear interferometry.

Each harmonic of the intensity distribution of Equation (7) is qiven by a sum of several fringe structures. Their contrast in the observation plane z = const is not constant and is different for each structure. The points of equal contrast value constitute ellipses described by

$$2\pi(p^2-r^2)\frac{\lambda z_0}{2d^2 z}\left[z-z_0+\frac{z_0}{2z^2}(3x^2+y^2+\delta/2)\right] = \text{const} \tag{9}$$

where p and r designate integer numbers. For each structure the ratio of the lengths of the axes is equal to $\sqrt{3}$. The calculations of the lengths of the axes as a function of the self-imaging configuration parameters z_0, z and d will be given in the following.

In the system under consideration there are no observation planes with constant intensity distribution. Such ones are encountered in two special cases only:

a) plane wavefront illumination, i.e., $z_0 = \infty$, $z = \infty$,

b) spherical wavefront illumination with $z_0 \to 0$, $z \to \infty$, that is, in the self-imaging configurations giving very large lateral magnifications.

2.2. Analysis of the intensity distribution of three lowest spatial harmonics

Fundamental harmonic

Intensity distribution of the fundamental harmonic is derived from Equation (7) in the form

$$I_I = a_0^2 + 2\sum_{n=1}^{\infty} a_n^2 + 4\sum_{n=0}^{\infty} a_n a_{n=1} \cos\left[2\pi \frac{z_0}{z}\frac{x}{d}\left(1 - \frac{x^2+y^2+\delta}{2z^2}\right)\right]$$
$$\times \cos\left\{2\pi[(n+1)^2 - n^2]\frac{\lambda z_0}{2d^2 z}[z - z_0 + z_0(3x^2+y^2+\delta/2)/2z]\right\}. \quad (10)$$

The zero contrast occurs when the value of the second cosine term is equal to zero. In this case we have

$$2\pi \frac{\lambda z_0}{2d^2 z}[z - z_0 + z_0(3x^2+y^2+\delta/2)/2z^2] = (2K-1)\frac{\pi}{2} \quad (11)$$

where K stands for an integer. This equation is valid for an arbitrary value of the binary grating opening ratio, i.e., the ratio of the slit width to the grating period.

Assuming that the cosine term under consideration attains the maximum on the optical axis, $x = y = 0$, we note that with a departure from the axis the contrast decreases and that its value becomes equal to zero when

$$2\pi \frac{\lambda z_0^2}{4d^2 z^3}(3x^2+y^2) = (2K-1)\frac{\pi}{2}. \quad (12)$$

The length of the longer axis of the K-th order ellipse is

$$y_{eI} = 2[(2K-1)d^2 z^3/\lambda z_0^2]^{1/2}. \quad (13)$$

The ratio of the lengths of the axes of the K-th and first ellipse is given by

$$Q_I = \frac{y_{eK,I}}{y_{e1,I}} = (2K-1)^{1/2} \quad (14)$$

where the subindex I refers to the first harmonic of the intensity distribution.

Second harmonic

From Equation (7) we obtain

$$I_{II} = a_0^2 + 2\sum_{n=1}^{\infty} a_n^2 + 2a_1^2 \cos\left[2\pi 2 \frac{z_0}{z}\frac{x}{d}\left(1 - \frac{x^2-y^2+\delta}{2z^2}\right)\right]$$
$$+ 4\sum_{n=0}^{\infty} a_n a_{n+2} \cos\left[2\pi 2 \frac{z_0}{z}\frac{x}{d}\left(1 - \frac{x^2+y^2+\delta}{2z^2}\right)\right]$$
$$\times \cos\left\{2\pi[(n+2)^2 - n^2]\frac{\lambda z_0}{2d^2 z}\left[z - z_0 + \frac{z_0}{2z^2}(3x^2+y^2+\delta/2)\right]\right\}. \quad (15)$$

The function determining the contrast of fringes is

$$f = 2a_1^2 + 4\sum_{n=0}^{\infty} a_n a_{n+2} \cos\{[(n+2)^2 - n^2]\gamma\} \quad (16)$$

where

$$\gamma = 2\pi \frac{\lambda z_0}{2d^2 z}\left[z - z_0 + \frac{z_0}{2z^2}(3x^2 + y^2 + \delta/2)\right]. \tag{17}$$

For practical purposes, in order to discuss the contrast it is enough to take into consideration only the first three terms of Eq. (16) as the dominating ones. The formulae describing the elliptical bands of minimum contrast depend now on the opening ratio α of the binary amplitude grating.

(a) $\alpha < 0.5$

In this case a_0, a_1 and a_2 are greater than 0, $a_3 < 0$. Simplified form of Eq. (16) becomes

$$f = 2a_1^2 + 4|a_0 a_2| \cos 4\gamma - 4|a_1 a_3| \cos 8\gamma. \tag{18}$$

The equation of minimum contrast fringe is

$$2\pi \frac{4\lambda z_0}{2d^2 z}\left[z - z_0 + \frac{z_0}{2z^2}(3x^2 + y^2 + \delta/2)\right] = (2K-1)\pi. \tag{19}$$

Neglecting the first two terms, i.e., $z - z_0$ in the square brackets (they do not depend on the off-axis coordinates x and y) we get

$$2\pi \frac{\lambda z_0^2}{d^2 z^3}(3x^2 + y^2) = (2K-1)\pi. \tag{20}$$

The length of the major axis of the K-th ellipse is

$$y_{eII} = \left[(2K-1)\frac{d^2 z^3}{\lambda z_0^2}\right]^{1/2}, \tag{21}$$

and the ratio of the lengths of the axes of the K-th and first ellipse is

$$Q_{II(\alpha<0.5)} = \frac{y_{eK,II}}{y_{e1,II}} = (2K-1)^{1/2}. \tag{22}$$

(b) $\alpha = 0.5$

This is the case of a Ronchi ruling with a_0 and a_1 greater than zero $a_2 = 0$ and $a_3 < 0$. The function (16) describing the fringe contrast becomes now

$$f = 2a_1^2 - 4|a_1 a_3| \cos 8\gamma. \tag{23}$$

The equation of minimum contrast bands is

$$2\pi \frac{4\lambda z_0}{2d^2 z}\left[z - z_0 + \frac{z_0}{2z^2}(3x^2 + y^2 + \delta/2)\right] = K\pi. \tag{24}$$

From its simplified form

$$2\pi \frac{\lambda z_0^2}{d^2 z^3}(3x^2 + y^2) = K\pi, \tag{25}$$

$$y_{eII} = \left[K \frac{d^2 z^3}{2\lambda z_0^2} \right]^{1/2}, \tag{26}$$

and

$$Q_{II(\alpha=0.5)} = K^{1/2}. \tag{27}$$

c) $\alpha > 0.5$

In this case we have a_0 and $a_1 > 0$, and a_2 and $a_3 < 0$. This leads to

$$f = 2a_1^2 - 4|a_0 a_2| \cos 4\gamma - 4|a_1 a_3| \cos 8\gamma. \tag{28}$$

The equation of elliptical minimum contrast bands is

$$2\pi \frac{4\lambda z_0}{2d^2 z} \left[z - z_0 + \frac{z_0}{2z^2} (3x^2 + y^2 + \delta/2) \right] = 2K\pi. \tag{29}$$

From its simplified form

$$2\pi \frac{\lambda z_0}{d^2 z^2} (3x^2 + y^2) = 2K\pi \tag{30}$$

we get

$$y_{eII} = \left[K \frac{d^2 z^3}{\lambda z_0^2} \right]^{1/2}, \tag{31}$$

and

$$Q_{II(\alpha>0.5)} = K^{1/2}. \tag{32}$$

Third harmonic

From Equation (7) the expression describing the third harmonic of intensity distribution is derived as

$$\begin{aligned} I_{III} = a_0^2 + 2 \sum_{n=1}^{\infty} a_0^2 + 4a_1 a_2 \cos & \left[2\pi 3 \frac{z_0}{z} \frac{x}{d} \left(1 - \frac{x^2 + y^2 + \delta}{2z^2} \right) \right] \\ & \times \cos \left\{ 2\pi 3 \frac{\lambda z_0}{2d^2 z} \left[z - z_0 + \frac{z_0}{2z^2} (3x^2 + y^2 + \delta/2) \right] \right\} \\ & + 4 \sum_{n=0}^{\infty} a_n a_{n+3} \cos \left[2\pi 3 \frac{z_0}{z} \frac{x}{d} \left(1 - \frac{x^2 + y^2 + \delta}{2z^2} \right) \right] \\ & \times \cos \left\{ 2\pi [(n+3)^2 - n^2] \frac{\lambda z_0}{2d^2 z} \left[z - z_0 + \frac{z_0}{2z^2} (3x^2 + y^2 + \delta/2) \right] \right\}. \end{aligned} \tag{33}$$

Since the amplitudes of higher order terms decrease rapidly, we will analyse the influence of the first term in the series of Eq. (33). In this case we have

$$I_{III} = a_2^2 + 2\sum_{n=1}^{3} a_n^2 + 4a_1a_2\cos\left[2\pi 3\,\frac{z_0}{z}\,\frac{x}{d}\left(1-\frac{x^2+y^2+\delta}{2z^2}\right)\right]$$
$$\times\cos\left\{2\pi\,\frac{3\lambda z_0}{2d^2 z}\left[z-z_0+\frac{z_0}{2z^2}\,(3x^2+y^2+\delta/2)\right]\right\}$$
$$+4a_0a_3\cos\left[2\pi 3\,\frac{z_0}{z}\,\frac{x}{d}\left(1-\frac{x^2+y^2+\delta}{2z^2}\right)\right]$$
$$\times\cos\left\{2\pi\,\frac{9\lambda z_0}{2d^2 z}\left[z-z_0+\frac{z_0}{2z^2}\,(3x^2+y^2+\delta/2)\right]\right\}. \tag{34}$$

The elliptical bands of minimum contrast are described by

$$2\pi\,\frac{3\lambda z_0}{2d^2 z}\left[z-z_0+\frac{z_0}{2z^2}\,(3x^2+y^2+\delta/2)\right] = (2K-1)\,\frac{\pi}{2}, \tag{35}$$

or

$$2\pi\,\frac{3\lambda z_0}{2d^2 z}\left[z-z_0+\frac{z_0}{2z^2}\,(3x^2+y^2+\delta/2) = \arccos\left\{+\left[3-\frac{a_1a_2}{a_0a_3}\right]\right\}^{1/2}. \tag{36}$$

It follows from the above equations that the dimensions of the ellipses of minimum contrast depend on the amplitude transmittance function of an object. It is not possible to derive a general expression for the lengths of the ellipse axes.

2.3. Discussion

From the above analytical considerations the following conclusions can be obtained:

(a) When calculating the lengths of the axes of the elliptical form of contrast modulation bands the factor $\delta/2$ can be omitted. It becomes significant for $z_0 > 200$ mm or for gratings of spatial frequency higher than 500 lines/mm.

(b) In the case of first, second and third harmonic detection there is a phase shift of π of the lines in the diffraction image when passing the zero contrast bands. This shift corresponds to the sign change of the contrast value. On the other hand, when detecting the second harmonic, this phase jump should not be noticeable. The negative contrast fringes are not visible because of a very small contrast value. For the Ronchi type square wave grating the contrast assumes the positive values only.

(c) The ratio of the lengths of the axes of the first and third harmonic ellipses depends on the ellipse order and the grating opening ratio α. All locations of the zero contrast ellipses for the first harmonic coincide with the zero contrast ellipses for the third harmonic. In a fixed area the number of ellipses for the third harmonic is 9 times greater than that for the first one.

(d) The ratio of the lengths of the axes of the first and second harmonic ellipses depends on the ellipse order and grating opening ratio α. All locations of the zero contrast ellipses for the first harmonic coincide with the zero contrast ellipses for the second harmonic when $\alpha \geqslant 0.5$. In a fixed area the number of ellipses for the second harmonic is 2 times greater than that for the first one.

(e) For $\alpha = 0.5$ the number of ellipses of minimum contrast for the second harmonic is 4 times greater than that for first one.

(f) Function describing the changes in contrast of the second harmonic has a second minimum (in a single period of cos 4γ), the contrast value differs from zero and depends on α.

The analysis presented concerned the three lowest harmonics of the Fresnel field intensity distribution. For the higher harmonics their contrast falls down and the ellipses of minimum contrast become very dense. Their detection is practically impossible.

3. Experimental work

The experiments were conducted to verify the principles derived in Sect. 2. The results were obtained in the simple configuration shown in Fig. 1. A point-source, spherical wavefront illumination was realized by focusing a He-Ne laser light beam by a microscope objective. A binary amplitude-type linear diffraction grating with a spatial frequency of 40 lines/mm was used as a self-imaging object. Its diffraction images were detected by a moiré fringe method.

After inspecting the grating transmittance its trapezoidal profile was assumed. The following Fourier expansion amplitude coefficients were calculated: $a_0 = 0.45A$, $a_1 = 0.81A/\pi$, $a_2 = 0.17A/\pi$, and $a_3 = -0.26A/\pi$, where A designates the maximum value of amplitude transmittance.

Several fringe patterns corresponding to different values of z_0 and z were observed and analysed. The lengths of the axes of the zero-contrast fringes were measured and checked against the theoretical values. The results are summarized in Tables 1 and 2.

T and E designate the theoretical and experimental values, respectively, x_{e1}, x_{e2} and x_{e3} correspond to the first, second and third order ellipses, respectively. Experimental values are by 5 to 9 percent lower in comparison to the theoretical

Table 1. Length of the minor axis of the zero-contrast elliptical bands. Detection of the first harmonic of the Fresnel field intensity distribution

x_e [mm] \ z_0 [mm]		50	95	100
$x_{e1,I}$	T	11.5	16	16.5
	E	10.5	15	16.5
$x_{e2,I}$	T	20	27.5	28
	E	19	26.5	27

Table 2. Length of the minor axis of the zero-contrast elliptical bands. Detection of the second harmonic of the Fresnel field intensity distribution.

x_e [mm] \ z_0 [mm]		40	95	100
$x_{e1,II}$	T	7.5	11	11.5
	E	6.5	10	11.5
$x_{e2,II}$	T	12.5	19.5	20
	E	11.5	17.5	18.5
$x_{e3,II}$	T	17	25	–
	E	15.5	22.5	–

ones. These differences are within the measurement accuracy of the axial separation distances z_0 and z and the distances between the centres of the zero contrast bands, which are not well defined.

Figure 2 shows moiré fringes in the detecting grating plane obtained with $z_0 = 75$ mm and $z = 150$ mm. The reference moiré fringes, which are necessary for visualization of the contrast modulation effect, were introduced by slightly rotating the detecting grating slightly about the optical axis (grating normal). Figures 2a and 2b show the case of the first and second harmonic detection, respectively. In both

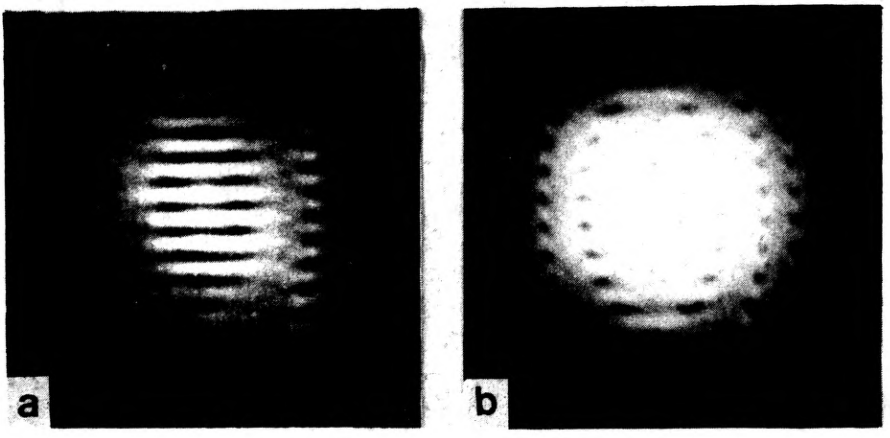

Fig. 2. Moiré fringe detection of the first (a) and second (b) harmonic of the Fresnel field intensity distribution of a binary grating with a spatial frequency of 40 lines/mm, $z_0 = 75$ mm, $z = 150$ mm. The self-imaging grating lines were vertical. Horizontal moiré fringes were obtained by a slight rotation of the detecting grating in its plane (different rotation angles were used in the two photographs)

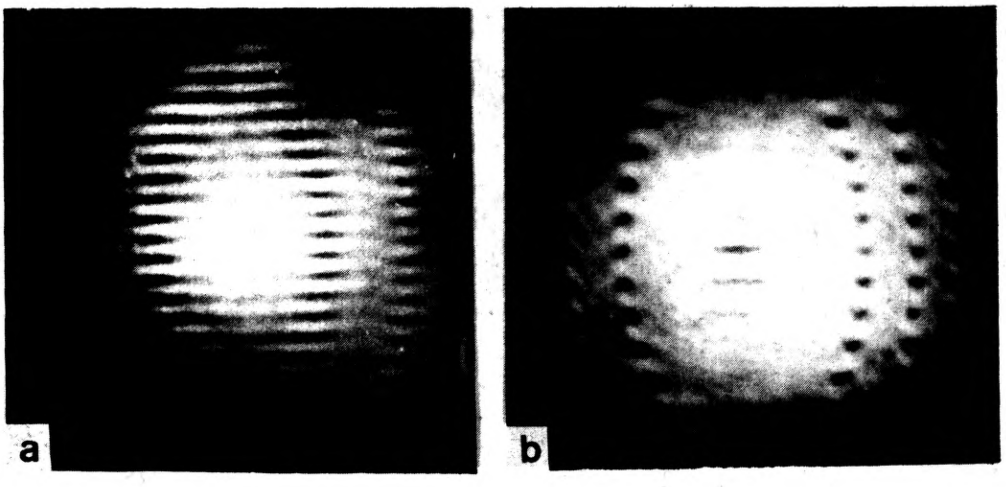

Fig. 3. Moiré fringe detection of the first (a) and second (b) harmonic of the Fresnel field intensity pattern, $z_0 = 95$ mm, $z = 190$ mm

cases the contrast modulation areas are clearly seen, as predicted theoretically. Nqte the presence and the lack of a phase change of π between the moiré fringes located on the opposite sides of the zero contrast elliptical bands in the case of the first and second harmonic detection, respectively. This is in perfect agreement with the conclusions stated above.

Figure 3 showns moiré fringes obtained with $z_0 = 95$ mm and $z = 190$ mm when detecting the first and second harmonics.

4. Summary

An analytical approach to the problem of self-imaging under a nonparabolic approximation of an optical path has been extended to detection of higher harmonics of the Fresnel diffraction field intensity distribution. The general formula has been derived, and the problem of contrast modulation which restricts the useful area of the self-image to be utilized has been studied in detail. The zero-contrast regions with the elliptically shaped fringelike structure have been calculated and verified experimentally. Good agreement has been obtained.

This paper can be treated as a contribution to a general theory of the self-imaging phenomenon [2].

References

[1] PATORSKI K., KOZAK S., J. Opt. Soc. Am. **A5** (1988), 1322.

[2] PATORSKI K., *The self-imaging phenomenon and its applications*, [In] *Progress in Optics*, Vol. 27, [Ed.] E. Wolf, Elsevier Science Publishers B.V., 1989.

Received November 27, 1989

Саморепродукция с непараболическим приближением сферического волнового фронта; расширение для объектов с несинусоидальной передаточной функцией амплитуды

Представлен теоретический анализ влияния аберрации выступающих в процессе дифракции на т. наз. явление саморепродукции (эффект Тальбота) предметов с несинусоидальной передаточной функцией амплитуды. Выведены формулы, описывающие высшие гармонические распределения интенсивности в дифракционном поле Френеля. Полученные зависимости проверены экспериментально.