Title	Holography : an interpretation from the phase-space point of view
Authors(s)	Situ, Guohai; Sheridan, John T.
Publication date	2007-12-15
Publication information	Optics Letters, 32 (24): 3492-3494
Publisher	Optical Society of America
Link to online version	http://dx.doi.org/10.1364/OL.32.003492
Item record/more information	http://hdl.handle.net/10197/3354
Publisher's statement	This paper was published in Optics Letters and is made available as an electronic reprint with the permission of OSA. The paper can be found at the following URL on the OSA website: http://www.opticsinfobase.org/abstract.cfm?URI=ol-32-24-3492. Systematic or multiple reproduction or distribution to multiple locations via electronic or other means is prohibited and is subject to penalties under law.
Publisher's version (DOI)	10.1364/OL.32.003492

Holography: an interpretation from the phase-space point of view

Guohai Situ[1,2] and John T. Sheridan[1,*]

[1]*School of Electrical, Electronic, and Mechanical Engineering, University College Dublin, Belfield, Dublin 4, Dublin, Ireland*
[2]*ghsitu@gmail.com*
*Corresponding author: john.sheridan@ucd.ie

The formation of holograms is interpreted as the consequence of the bilinearity of the ambiguity function. Reconstruction can then be regarded as the manipulation of the ambiguity function. Specifically, we show that in the case of in-line holography, the reconstruction can be regarded as phase tomography. In this way we provide a unified picture for the formulation of both noninterferometric and interferometric phase-retrieval techniques. © 2007 Optical Society of America

OCIS codes: 100.5070, 090.0090, 110.6960, 110.2990, 070.0070.

Phase-retrieval techniques (PRTs) have found important applications in various fields such as in astronomy, radiography, crystallography [1], microscopy [2,3], and security [4,5]. The determination of the phase of a scattered wave field is of interest because it carries important information about the object surface or inner structure. Generally, PRTs fall into two main categories: interferometric (InF) and noninterferometric (NInF) approaches. The latter category can be further classified into methods based on iterative technique [1,3–5], transport-of-intensity equation (TIE) [2,6], phase-space tomography [7,8], and moments of phase-space distributions [9]. All of these NInF approaches essentially use intensity measurements of the scattered field at various domains to retrieve the object phase. At each of these domains, the phase-space distribution, e.g., the ambiguity function (AF) associated with the wavefront has the same value, but the coordinates undergo an affine transform specified by the ABCD matrix [10,11]. Thus the Fourier transform of each measured intensity corresponds to a plane through the ambiguity space with a slope with respect to the propagation distance. The intensity measurements required by the NInF approaches mentioned above can then be regarded as specific samplings in the ambiguity space [12,13], and these techniques have a unified interpretation based on phase-space tomography.

InF approaches, including digital holography (DH), have received much attention in the last decades and have been successfully applied in many areas [14]. Usually the principle of holography has been primarily interpreted using communication theory [15]. It is not until recently that attempts have been made to interpret holography using the Wigner distribution function (WDF) [16]. However, as we will show, the WDF representation does not depict holography most conveniently, while the AF representation provides a much more insightful interpretation. This may be instructive in developing a sampling strategy in the reconstruction of DH. It may also provide a unified framework for the common formulation of InF and NInF PRTs.

For simplicity, we consider a one-dimensional (1D) signal $s(x)$, since the generalization to two-dimensions is straightforward. The WDF of $s(x)$ is defined as [11]

$$\mathcal{W}_s(x,\nu) = \int s\left(x + \frac{\bar{x}}{2}\right)s^*\left(x - \frac{\bar{x}}{2}\right)\exp[-j2\pi\nu\bar{x}]\mathrm{d}\bar{x}.$$

(1)

One important property of the WDF is its bilinearity, which is useful in analyzing the coherent superposition of two signals $y(x) = s(x) + r(x)$. The WDF of $y(x)$ is

$$\mathcal{W}_y(x,\nu) = \mathcal{W}_s(x,\nu) + \mathcal{W}_r(x,\nu) + \mathcal{W}_{sr}(x,\nu) + \mathcal{W}_{rs}(x,\nu),$$

(2)

where

$$\mathcal{W}_{sr}(x,\nu) = \int s\left(x + \frac{\bar{x}}{2}\right)r^*\left(x - \frac{\bar{x}}{2}\right)\exp[-j2\pi\nu\bar{x}]\mathrm{d}\bar{x},$$

(3)

and straightforward for the definition of $\mathcal{W}_{rs}(x,\nu)$. Equation (2) shows that the cross terms arise because of the bilinearity. Now let us consider the case when a plane wave of special frequency ν_0, i.e., $r(x) = \exp[j2\pi\nu_0 x]$ is incident at the interference plane whose longitudinal coordinate is $z = 0$. The WDF of $r(x)$ is $\delta(\nu - \nu_0)$, and according to Eq. (3) $\mathcal{W}_{sr}(x,\nu) = 2\exp[j4\pi(\nu - \nu_0)x]S(2\nu - \nu_0)$, $\mathcal{W}_{rs}(x,\nu) = -2 \times \exp[-j4\pi(\nu - \nu_0)x]S^*(2\nu - \nu_0)$, where S and S^* are the Fourier transforms of s and s^*. As schematically shown in Fig. 1, both cross terms occupy the same spatial frequency region centered at the coordinate $(0, \nu_0/2)$, while the WDFs of the signal and the reference are centered at origin and $(0, \nu_0)$, respectively. This is clearly not the familiar picture of off-axis holograms as interpreted using the carrier wave theory [15].

Now we consider the AF representation. The AF was initially proposed and is commonly used in radar and sonar to track moving targets [10]. The AF of the signal $s(x)$ is defined as
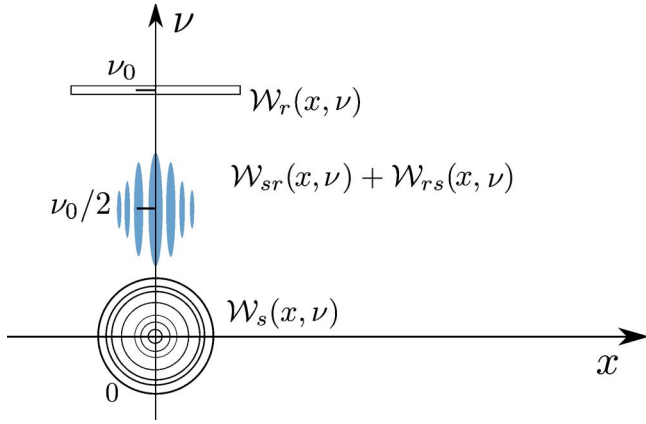
Fig. 1. (Color online) Schematic geometry of the WDF of the superposition of a band-limited signal $s(x)$ and an incline plane wave.

$$A_s(\bar{\nu},\bar{x}) = \int s\left(x + \frac{\bar{x}}{2}\right)s^*\left(x - \frac{\bar{x}}{2}\right)\exp[-j2\pi\bar{\nu}x]\mathrm{d}x, \tag{4}$$

which is the two-dimensional (2D) Fourier transform of its WDF. Like the WDF, the AF is also bilinear. The AF associated with $y(x)$ is

$$A_y(\bar{\nu},\bar{x}) = A_s(\bar{\nu},\bar{x}) + A_r(\bar{\nu},\bar{x}) + A_{sr}(\bar{\nu},\bar{x}) + A_{rs}(\bar{\nu},\bar{x}), \tag{5}$$

in which the cross terms $A_{sr}(\bar{\nu},\bar{x})$ and $A_{rs}(\bar{\nu},\bar{x})$ can be defined following Eqs. (3) and (4). Historically, these cross terms were considered undesirable in time-frequency domain signal processing [17], and efforts were made to eliminate them. However, they are of importance in our case because they result in the hologram. We again examine the case of the coherent superposition of the signal $s(x)$ with the incline plane wave $r(x)$. The AF of the cross terms can be written as $A_{sr}(\bar{\nu},\bar{x}) = \exp[+j\pi(\bar{\nu}+2\nu_0)\bar{x}]S(\bar{\nu}+\nu_0)$, and $A_{rs}(\bar{\nu},\bar{x}) = \exp[-j\pi(\bar{\nu}-2\nu_0)\bar{x}]S^*(\bar{\nu}-\nu_0)$, respectively. Substituting them into Eq. (5), we can obtain the distribution of $y(x)$ in the ambiguity space.

If the signal $s(x)$ propagates along the z axis, $A_s(\bar{\nu},\bar{x})$ then lies at the origin in the 2D ambiguity space, as does $A_r(\bar{\nu},\bar{x})$. The cross terms, $A_{sr}(\bar{\nu},\bar{x})$ and $A_{rs}(\bar{\nu},\bar{x})$ are centering at $(-\nu_0,0)$ and $(+\nu_0,0)$, and if ν_0 is sufficiently large, are well separated, as schematically shown in Fig. 2.

It is interesting to examine the line $\bar{x}=0$ in the ambiguity space. The 1D pattern with respect to this line can, according to Eq. (5), be expressed as

$$A_y(\bar{\nu},0) = A_s(\bar{\nu},0) + A_r(\bar{\nu},0) + A_{sr}(\bar{\nu},0) + A_{rs}(\bar{\nu},0) \tag{6}$$

$$= A_s(\bar{\nu},0) + \delta(\bar{\nu}) + S(\bar{\nu}+\nu_0) + S^*(\bar{\nu}-\nu_0). \tag{7}$$

Referring to the definition of the AF, Eq. (6) is the Fourier spectrum of the recorded hologram $h(x) = |y(x)|^2$. We illustrate this in Fig. 2. It can be clearly seen that the Fourier spectra of the object wavefront $S(\bar{\nu})$ and of its conjugate $S^*(\bar{\nu})$ are located on either

side of the origin, with shifts in spatial frequency of $-\nu_0$ and $+\nu_0$, respectively, while the spectra of the DC terms $A_s(\bar{\nu},0) + \delta(\bar{\nu})$ occupy the lower frequencies around the origin. This is consistent with the picture of the hologram spectrum provided by the carrier wave theory [15].

We can see from Eq. (7) that the line $\bar{x}=0$ contains the full complex amplitude and not only its intensity as in the NInF cases. Thus appropriate filtering allows extraction of the signal of interest $A_{rs}(\bar{\nu},0)$ or $A_{sr}(\bar{\nu},0)$ from which the object wavefront can be retrieved.

If the spatial frequency of the reference ν_0 is insufficiently large, $A_{rs}(\bar{\nu},\bar{x})$ and $A_{sr}(\bar{\nu},\bar{x})$ will overlap with $A_s(\bar{\nu},\bar{x})$ and filtering can no longer be satisfactorily performed. In the extreme case of $\nu_0=0$, $y(x)=s(x)+1$, all terms in Eq. (5) overlap; the AF of $y(x)$ then becomes

$$A_y(\bar{\nu},\bar{x}) = A_s(\bar{\nu},\bar{x}) + \delta(\bar{\nu})$$
$$+ \exp[j\pi\bar{\nu}\bar{x}]S(\bar{\nu}) + \exp[-j\pi\bar{\nu}\bar{x}]S^*(\bar{\nu}), \tag{8}$$

which corresponds to the case of in-line holography. To cancel out the DC and conjugate terms, one approach is to introduce $\pi/2$-stepwise phase shifting into the reference beam [14]. The signs of the last three terms in Eq. (8) then change. The reconstruction is achieved by algebraic manipulations of these AFs.

Alternatively, one can shift the CCD position and record several holograms longitudinally at the planes z_i, $(i=1,2,\ldots)$, parallel to the first plane (denoted by $z_0=0$). The Fourier transforms of these holograms can then be expressed as

$$A_y(\bar{\nu},z_i\bar{\nu}) = A_s(\bar{\nu},z_i\bar{\nu}) + \delta(\bar{\nu})$$
$$+ \exp[j\pi z_i\bar{\nu}^2]S(\bar{\nu}) + \exp[-j\pi z_i\bar{\nu}^2]S^*(\bar{\nu}). \tag{9}$$
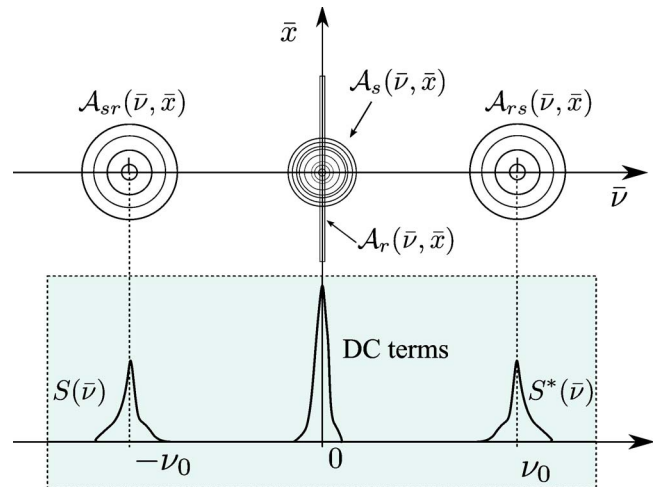

Fig. 2. (Color online) Schematic geometry of the AF of the superposition of a band-limited signal $s(x)$ and an incline plane wave $r(x)$. The lower box shows the Fourier spectrum of the hologram.

**Table 1. Comparison of Interferometric and Noninterferometric Phase-Retrieval Techniques
Based on the AF**

	Off-axis	In-line	Noninterferometric
Terms of AF being used	Cross	Cross	Auto
Information contained	Amplitude	Amplitude	Intensity
Uniqueness of solution	Yes	Yes	Algorithm dependent
CCD SBP can be used	Part	Full	Full
Number of projections	1	—	1^a
	—	2^b	2^c
	—	$>2^d$	$>2^e$

[a]Homogeneous or phase-only objects, or Fienup's method [1].
[b]With weak object assumption [18].
[c]TIE [6], GS algorithm [1].
[d]Phase shifting or position shifting.
[e]Multiple measurements [7,8,13].

It is seen from the last two terms that we have come to the Fresnel transform. Note that in the paraxial approximation $\mathcal{A}_y(\bar{\nu},\bar{x})$ corresponding to $y(x)$ at different planes have the same distribution, subject to a coordinate transform specified by an ABCD matrix [10], $\mathcal{A}_y(\bar{\nu},z_i\bar{\nu})$ therefore can be regarded as the samplings of the ambiguity space $\mathcal{A}_y(\bar{\nu},\bar{x})$ along the lines $\bar{x}=z_i\bar{\nu}$, $(i=0,1,\dots)$. Note that on the right-hand side of each of Eqs. (9) the variable $\mathcal{A}_s(\bar{\nu},z_i\bar{\nu})$ changes with respect to z_i, while the other two $S(\bar{\nu})$ and $S^*(\bar{\nu})$ are independent on z_i. However, $\mathcal{A}_s(\bar{\nu},z_i\bar{\nu})$ can be known by measuring the intensities of the object wavefront at planes z_i. There are, actually, two unknown variables in Eqs. (9). Normally, it may be possible to use four samples or more [that is, to record at least two holograms and object intensities at longitudinal positions; their Fourier spectra give $\mathcal{A}_y(\bar{\nu},z_0\bar{\nu})$, $\mathcal{A}_y(\bar{\nu},z_1\bar{\nu})$, $\mathcal{A}_s(\bar{\nu},z_0\bar{\nu})$, and $\mathcal{A}_s(\bar{\nu},z_1\bar{\nu})$] to solve for $S(\bar{\nu},0)$. This $S(\bar{\nu},0)$ is essentially the Fourier spectrum of the full-complex object wavefront at the first interference plane $z_0=0$. If the weak-object approximation is applicable [18,19], Eq. (9) can be reduced to simple algebraic equations. Two samples are sufficient to retrieve the object phase. Thus this reconstruction approach can also be regarded as phase-space tomography.

In conclusion, we have shown that the formation of holograms can be interpreted as the result of the bilinearity of the AF. Compared with the WDF representation [16], the present approach provides a picture consistent with the carrier wave theory. One important prediction of this interpretation is that the reconstruction of a hologram may be possible by intensity measurements along the z axis. This may result in a new reconstruction algorithm by solving Eq. (9). Compared with the NInF techniques, such as the phase-space tomography of the autoterm of the AF, the holographic approaches can be regarded as the phase-space tomography or filtering of the cross term of the AF in terms of wavefront reconstruction. This provides a unified picture for the formulation of both these two categories of PRTs. However, as shown in detail in Table 1, each of these techniques has its own applicability. Generally speaking, InF techniques have simple reconstruction algorithms, while NInF techniques have simpler measurement setups.

References

1. J. R. Fienup, Appl. Opt. **21**, 2758 (1982).
2. A. Barty, K. A. Nugent, D. Paganin, and A. Roberts, Opt. Lett. **23**, 817 (1998).
3. J. W. Miao, T. Ishikawa, B. Johnason, E. H. Anderson, B. Lai, and K. O. Hodgson, Phys. Rev. Lett. **89**, 088303 (2002).
4. G. Situ and J. Zhang, Proc. SPIE **5642**, 108 (2005).
5. G. Situ, U. Gopinathan, D. S. Monaghan, and J. T. Sheridan, Appl. Opt. **46**, 5257 (2007).
6. M. R. Teague, J. Opt. Soc. Am. **73**, 1434 (1983).
7. M. G. Raymer, M. Beck, and D. F. McAlster, Phys. Rev. Lett. **72**, 1137 (1994).
8. J. Tu and S. Tamura, Phys. Rev. E **55**, 1946 (1997).
9. T. Alieva and M. J. Bastinaans, IEEE Trans. Signal Process. **51**, 112 (2003).
10. A. Papoulis, J. Opt. Soc. Am. **64**, 779 (1974).
11. A. W. Lohmann, J. Opt. Soc. Am. A **10**, 2181 (1993).
12. A. Semichaevsky and M. Testorf, J. Opt. Soc. Am. A **21**, 2173 (2004).
13. K. A. Nugent, J. Opt. Soc. Am. A **24**, 536 (2007).
14. U. Schnars and W. Jüeptner, *Digital Holography* (Springer, 2004).
15. E. N. Leith and J. Upatnieks, J. Opt. Soc. Am. A **52**, 1123 (1962).
16. K. B. Wolf and A. L. Rivera, Opt. Commun. **144**, 36 (1997).
17. F. Hlawatsch and G. F. Boudreaux-Bartels, IEEE Signal Process. Mag. **9**, 21 (1992).
18. W. L. Bragg and G. L. Rogers, Nature **167**, 190 (1951).
19. Y. Nishino, T. Ishikawa, K. Hayashi, Y. Takahashi, and E. Matsubara, Phys. Rev. B **66**, 092105 (2002).