Diffraction and holography from a signal processing perspective

Levent Onural
Haldun M. Ozaktas
Diffraction and Holography from a Signal Processing Perspective
Levent Onural and Haldun M. Ozaktas
Department of Electrical Engineering, Bilkent University, TR-06800 Bilkent, Ankara, Turkey

ABSTRACT
The fact that plane waves are solutions of the Helmholtz equation in free space allows us to write the exact solution to the diffraction problem as a superposition of plane waves. The solution of other related problems can also be expressed in similar forms. These forms are very well suited for directly importing various signal processing tools to diffraction related problems. Another signal processing-diffraction link is the application of novel sampling theorems and procedures in signal processing to diffraction for the purpose of more convenient and efficient discrete representation and the use of associated computational algorithms. Another noteworthy link between optics and signal processing is the fractional Fourier transform. Revisiting diffraction from a modern signal processing perspective is likely to yield both interesting viewpoints and improved techniques.

Keywords: diffraction, signal processing, sampling, fractional Fourier transform

1. REVIEW OF DIFFRACTION FROM A SIGNAL PROCESSING PERSPECTIVE
The relationships between signals and systems concepts and basic diffraction and optical systems is well established in classic works. We begin with a brief review of these basics in order to prepare the ground for more advanced concepts.

Probably the best way to show how optical wave propagation fits into a signals and systems framework is the planar wave decomposition approach:

$$\psi(x) = \int B(k) \exp(jk^T x) \, dk$$

where $x$ represents the space coordinates $[x, y, z]^T$, $k$ is the wave vector $[k_x, k_y, k_z]^T$, $B(k)$ is the amplitude of the plane wave propagating along the $k$ direction, and $\psi(x)$ is the corresponding three-dimensional (3D) field. Various restrictions reduce the domain of integration. For example, for monochromatic waves with wavelength $\lambda$, the integration set becomes the sphere whose radius is equal to $k = |k| = (2\pi)/\lambda$ (the Ewald sphere). It is also common to limit the propagation direction only along the positive $z$-axis, and therefore, the domain of integration reduces to the positive $z$ semisphere. For the monochromatic case, Eq. 1 becomes,

$$\psi(x) = \int_{k_x^2 + k_y^2 + k_z^2 = k^2} B(k) \exp(jk^T x) \, dk$$

$$= \int_{k_x^2 + k_y^2 \leq k^2} B(k) \frac{k}{k_z} \exp(jk_z z) \exp[j(k_x x + k_y y)] \, dk_x dk_y$$

where the term $\frac{k}{k_z}$ is due to the Jacobian as a consequence of the change of integration variables; $k_z = (k^2 - k_x^2 - k_y^2)^{1/2}$, is now a function of $(k_x, k_y)$ due to the monochromaticity constraint.

Author contact info: onural@bilkent.edu.tr, haldun@ee.bilkent.edu.tr. More info: www.3dtv-research.org
A planar cross-section of \( b(x) \) at \( z = 0 \) yields the 2D field which is usually called the "object mask":

\[
\psi_{2D_0}(x,y) \triangleq \psi(x,y,0) = \int \int B(k) \frac{k}{k_z} \exp \left[ j(k_z x + k_y y) \right] dk_x dk_y
\]

\[
= \mathcal{F}^{-1} \left\{ \frac{4\pi^2 B(k)}{k_z} \right\}
\]  

(3)

where \( \mathcal{F}^{-1} \) represents the inverse Fourier transform from the \((k_x, k_y)\) domain to the \((x, y)\) domain. It is quite common to absorb the Jacobian into a new function by defining \( A(k_x, k_y) = 4\pi^2 B \left( k_x, k_y, (k^2 - k_x^2 - k_y^2)^{1/2} \right) \frac{k_z}{k} \).

A similar cross-section of the 3D field at the \( z = Z \) plane yields the expression for the corresponding diffraction pattern over that plane as,

\[
\psi_{2D_z}(x,y) \triangleq \psi(x,y,Z) = \int \int B(k) \frac{k}{k_z} \exp \left[ j(k^2 - k_x^2 - k_y^2)^{1/2} \right] \exp \left[ j(k_x x + k_y y) \right] dk_x dk_y
\]

\[
= \mathcal{F}^{-1} \left\{ A(k_x, k_y) H_Z(k_x, k_y) \right\}
\]

(4)

where \( H_Z(k_x, k_y) = \exp \left[ j(k^2 - k_x^2 - k_y^2)^{1/2} \right] \) represents the transfer function of a linear shift-invariant system.

Eqs. 1, 2, 3, 4 are exact solutions for a homogeneous linear isotropic medium since the plane wave is a solution of the Helmholtz equation. The impulse response corresponding to this system is the kernel of the well-known Rayleigh-Sommerfeld solution.\(^2\,^3\) For \( Z \gg \lambda \), the impulse response reduces to the kernel representing wave propagation in free space due to a point source as,

\[
h_z(x,y) \approx \frac{Z}{j\lambda(x^2 + y^2 + Z^2)} \exp \left[ \frac{j2\pi}{\lambda} \left( x^2 + y^2 + Z^2 \right)^{1/2} \right]
\]

(5)

which is known as the Rayleigh-Sommerfeld diffraction formula.\(^2\)

As a result of the simple linear shift-invariant model with exactly specified transfer function, diffraction between parallel planes can be conveniently modeled and therefore associated signal processing techniques can be immediately applied. Furthermore, issues associated with discretization, etc. can be rather easily understood due to this simple product form in the Fourier domain. The analytically known transfer function and the impulse response provide additional benefits.

A discrete simulation of diffraction for the 1D case (2D field), corresponding to Eq. 4, is shown in Figure 1. Here the bottom line is the input mask, which is a simple pulse (transparent opening) at the center of an otherwise opaque object. The vertical axis is the \( z \)-axis representing the distance between the object and the diffraction plane. Please note that the spatial scale is not the same for the horizontal and the vertical axes.

There has been significant efforts reported in the literature to compute the diffraction pattern between tilted planes efficiently.\(^4\,^6\) The presented plane wave decomposition approach provides an efficient and elegant solution based on signal processing techniques: It is easy to find out the 2D functions over two different arbitrarily oriented 2D planes by intersecting a single 3D plane wave by those planes: each intersection is merely a 2D complex sinusoidal function (i.e., a 2D plane wave) whose orientation and frequency are different and dependent on the positions and orientations of the intersecting planes. Therefore, by just forming the superposition of plane waves in 3D space and the corresponding superposition of the 2D plane waves over the two 2D intersecting planes, we can easily find the diffraction relation between the tilted planes:

\[
g(x,y) = \mathcal{F}^{-1}(k_x, k_y) \rightarrow (x,y) \left\{ \mathcal{F}(x,y) \rightarrow (k_x, k_y) \{ f(x,y) \} \right\}_{k_x \rightarrow k_x'} H(k_x', R, b)_{k_x'}
\]

(6)

where \( \mathcal{F} \) represents the 2D Fourier transform, and the arrowed subscripts under the symbol \( \mathcal{F} \) denote the variables of the pre- and post-Fourier transform domains, and the function \( H(k_x', R, b) \) provides the kernel of the corresponding system, represented by

\[
H(k_x', R, b) = e^{jk_x' (R \cdot b)}.
\]

(7)
Figure 1. Simulation of diffraction using Eq. 4. A 1-D object case is presented for better visualization. (The scale is not the same along the horizontal (x), and the vertical (z) axes. The numerical values are for \( \lambda = 633nm \).) (The simulation was conducted by Ali Özgür Yöntem.)

Here, the relation between \( k \) and \( k' \) is given by a rotation as \( k = Rk' \). Physically, \( k \) is the direction of propagation with respect to the object plane, and \( k' \) is the same direction with respect to the orientation of the tilted diffraction plane. A similar rotation and translation relation exists between the coordinates of the tilted planes as \( x' = Rx + b \) where the matrix \( R \) represents the rotation of the two planes and the vector \( b \) represents the translation.

2. OPTICAL PROPAGATION AS FRACTIONAL FOURIER TRANSFORMATION

The purpose of this section is to briefly review fractional Fourier transforms and linear canonical transforms, as they are related to wave propagation, diffraction, and holography. It is not the purpose of this section to attempt any review of the fractional Fourier transform in general; other sources are available for this purpose. Excellent references on linear canonical transforms exist. The fractional Fourier transform was introduced into signal processing and optics during the early nineties.

A key result is that relating free-space propagation in the Fresnel approximation (namely the Fresnel integral or the Fresnel transform), to the fractional Fourier transform. Several papers deal with this relationship. Extensions of this result relate arbitrary linear canonical transforms (quadratic-phase integrals) to the fractional Fourier transform. Linear canonical transforms are a three-parameter family of integral transforms. This family of transforms includes the Fourier and fractional Fourier transforms, simple scaling including the identity and parity operations (corresponding to imaging in optics), chirp multiplication and convolution operations (corresponding to passage through a thin lens and free-space propagation in the Fresnel approximation respectively), and hyperbolic transforms as special cases. Since optical systems consisting of arbitrary concatenations of lenses and section of free space can be modeled as linear canonical transforms, it follows that propagation through such systems, as well as free-space propagation can be viewed as an act of
continual fractional transformation. The wave field evolves through fractional Fourier transforms of increasing
order as it propagates through free space or the multi-lens system.

In order to mathematically express the above result, we first give the definition of the fractional Fourier
transform (FRT). The FRT \( f_a(x) \) of \( f(x) \) is defined as

\[
f_a(x) = A_{a\pi/2} \int_{-\infty}^{\infty} \exp[i\pi(\alpha^2 - 2\alpha x' \csc(\alpha \pi/2) + x'^2 \cot(\alpha \pi/2))]f(x') \, dx',
\]

where \( A_{a\pi/2} \) is a factor depending on \( a \) whose exact form is not of importance here. Restricting ourselves to
one-dimensional notation for simplicity, the output \( g(x) \) of a quadratic-phase system is related to its input \( f(x) \)
through

\[
g(x) = \sqrt{\beta}e^{-ix/4} \int_{-\infty}^{\infty} \exp[i\pi(\alpha x^2 - 2\beta x + \gamma x^2)]f(x') \, dx',
\]

where \( \alpha, \beta, \gamma \) are the three parameters of the system. When all three of these parameters equal 1/\( \lambda Z \),
this expression reduces to the Fresnel integral (within an inconsequential phase factor). The same relationship
can also be written in terms of an alternate set of parameters \( \alpha, M, R \) as follows:

\[
g(x) = e^{ix^2/\lambda R} \frac{1}{\sqrt{s^2 M}} A_{a\pi/2} \int_{-\infty}^{\infty} \exp \left[ \frac{i\pi}{s^2 M^2} \left( \frac{x^2}{M^2} \cot(\alpha \pi/2) - 2x' \frac{x'^2}{M} \csc(\alpha \pi/2) + x'^2 \cot(\alpha \pi/2) \right) \right] f(x') \, dx',
\]

where \( s \) is an arbitrary scale factor. This relationship maps a function \( s^{-1/2} f(x/s) \) to \( \exp(i\pi x^2/\lambda R) / \sqrt{1/s M f_a(x/s M)} \). That is, \( g(x) \) is essentially the \( a \)th order fractional Fourier transform of \( s^{-1/2} f(x/s) \), scaled by \( M \), and multiplied by a residual quadratic-phase factor. In optics scaling corresponds to magnification of the distribution of light in the transverse direction. The existence of the quadratic-phase factor means that the magnified fractional Fourier transform is observed on a spherical reference surface, rather than on a plane. Comparing Eqs. 9 and 10, we can relate the two sets of parameters as follows:

\[
\alpha = \frac{\cot(\alpha \pi/2)}{s^2 M^2} + \frac{1}{\lambda R},
\]

\[
\beta = \frac{\csc(\alpha \pi/2)}{s^2 M},
\]

\[
\gamma = \frac{\cot(\alpha \pi/2)}{s^2}.
\]

These equations allow us to switch between the two sets of parameters and thus interpret any quadratic-phase
integral and thus the wide class of optical systems they represent as fractional Fourier transforms. Since the FRT
has a much broader set of properties mirroring those of the ordinary Fourier transform, and is geometrically and
numerically much better behaved, formulating the propagation of light through optical systems in terms of the
FRT has several advantages. As a special case, when \( \alpha = \beta = \gamma = 1/\lambda Z \) corresponding to ordinary free-space
propagation, we have

\[
\tan(\alpha \pi/2) = \frac{\lambda Z}{s^2},
\]

\[
M = \sqrt{1 + (\lambda Z/s^2)^2},
\]

\[
\frac{1}{\lambda R} = \frac{1}{s^4} + \frac{\lambda Z}{1 + (\lambda Z/s^2)^2}.
\]

There is no doubt that digital processing of signals paved the way to otherwise impossible techniques in almost
every field. Audio and video technology, together with digital telecommunications methods affected all aspects
of daily life, including business styles, home living, and leisure. The close interaction of digital technologies with
optics has been rather late, at least at the visible consumer products level.

Coupling optical signals with computers, and thus, migrating the benefits of the digital technology to optics,
require the digitization of these signals. At the input, where the analog optical signal is captured and converted

Proc. of SPIE Vol. 6252  625219-4
to a digital signal for subsequent digital processing, usually the light hits an array of sensors located at the
surface of a chip. The size and shape of the aperture, the size, shape and the number of sensing elements, their
noise characteristics, speed, etc., affect the quality of the captured signal. Currently the common technology is
to use CCD arrays. Such captured discrete signals are then digitized and stored in computer memory and can
be reconstructed for further digital processing, or are fed to a SLM for optical reconstruction.2831 Another area
of widespread interest is the generation of various types of holograms and other diffractive optical elements by
digital means. Processing of captured optical signals by digital means, for purposes like digital reconstructions
from holograms, analysis of holographic signals, nondestructive testing, technical measurements, etc., are of
considerable interest. Creation of synthetic 3D data by computer graphics means, and then displaying these 3D
data using holographic technology also require digital processing. Finally, simulations of optical phenomena for
scientific purposes or for computer-aided design and perfection of optical components require handling of optical
and related signals by digital means.

All of the primary applications above, as well as other related tasks, require the sampling (discretization) and
quantization of associated signals: these signals could be the diffracted wave from an object, propagated light
from a scene to a sensor, a hologram, and mathematical functions describing these physical phenomena.

3. A REVIEW OF SAMPLING THEORY

Our review here, while including the fundamentals of general sampling theory, emphasizes those aspects that are
more relevant to optics and diffraction.

The classical formulation of sampling starts with the influential approach published by Shannon.32-34 It is
well known that, a band-limited function can be fully recovered from its uniformly-positioned samples taken at
the Nyquist rate. The recovery of the original continuous function is accomplished by superposing weighted and
shifted sinc interpolators. It is known that the roots of sampling of band-limited functions are quite older.35,36
Unser provides an excellent survey on sampling.37

Actually, band-limitedness is just one constraint which then leads to representation of continuous functions
by their samples. Other constraints may lead to totally different discrete representations, based on the sam-
ple, together with corresponding recovery procedures of the originals. Indeed it seems that applying common
Shannon-type (band-limited original signal) sampling is neither appropriate, nor desirable in optics and diffrac-
tion.

For optical signals arising in diffraction and holography, the so called α-Fresnel limited functions are far
more convenient and appropriate than band-limited functions, at least as far as sampling and recoverability
is concerned.38 It is shown that functions which are not necessarily bandlimited can be fully recovered from
their finite rate samples, provided that they are α-Fresnel limited. A special case of the theorems proven in
the literature38 indicate that the Fresnel transform of a space-limited function can be fully recovered from its
samples. This was also proven later independently by Onural,39 where it is shown that a space-limited function
can be fully recovered from the samples of its Fresnel diffraction pattern. Since the space-limited function is
not band-limited, and since diffraction is essentially an all-pass linear operator, the diffraction pattern is not
band-limited either. Therefore, Shannon’s theorem is not applicable but nevertheless full recovery is possible
from a rather sparse set of uniform samples. Another work on sampling appears in the recent literature: it is
shown by Stern and Javidi40 that neither band-, nor space-limited functions can be fully recovered from their
samples if the replicas of their Wigner distributions due to sampling do not overlap.

Another interesting approach to sampling may be realized through viewing diffraction phenomena as a con-
tinuous wavelet transform.41 It is shown in that work that if the field is considered to be produced in accordance with
the Fresnel approximation, the light field at different distances (along propagation direction) may be regarded
as the result of an inner product of the light distribution at some initial plane (orthogonal to the propagation
direction) and a function (chirp) scaled by the square root of the distance. The difference from the conventional
wavelet analysis is that the scaling functions are not limited either in the spatial or the frequency domain. Thus
a question arises whether such a transform is actually legitimate (the answer to which is actually expected to be
positive due to the fact that the distribution at a plane determines the field in the whole space). The transform
is named as a “scaling chirp transform” and is shown to be legitimate in.42 A number of inversion formulas

3536
are provided with a discussion of their redundancy (when a volume is considered for inversion where a plane is sufficient) and ways to possibly exploit this redundancy. Despite its difference in the underlying wavelet function, the scaling chirp transform seems to suggest a way to sample the light field. Interestingly, this sampling is to be performed throughout the space at different distances (since the inner product following the scaling of the chirp corresponds to obtaining the light field at a farther plane). Actually the need at this point is to discretize the scaling chirp transform to avoid redundancies. Once this is done, the samples that one obtains would be the samples of the light field throughout the space.

Digital reconstructions from captured diffraction patterns or holograms require the algorithmic digital implementations of the underlying continuous mathematical models representing diffraction. There are two common implementations for the Fresnel case. One of the implementations is based on the implementation of the convolution of the input with the Fresnel kernel which represents a linear shift-invariant system. The specific form of the kernel (the two-dimensional chirp, which is also called the quadratic-phase function, or the zone-plate term) makes it possible to convert the convolution to a single Fourier transform together with pre and post array multiplications with the quadratic phase function. Inevitably, either the kernel which represent wave-propagation (diffraction), or its analytically known Fourier transform (the transfer function) should be discretized for performing digital reconstructions by converting the continuous convolutions to their discrete counterparts. Similar discretization of the quadratic-phase function is an issue for the single Fourier transform model. Therefore, understanding the properties of the discrete function obtained from the continuous Fresnel function is essential for both successful simulations of diffraction, and for proper interpretations of computer generated results. Some well known properties of the continuous Fresnel kernel, together with some rather overlooked properties are presented in the literature.

One such interesting property is the invariance of the sampling operation under the Fourier transform for such a function: the continuous Fourier transform of the sampled chirp function, under some conditions, is another sampled (conjugate) chirp function. This observation leads to desirable interpretations of the outputs obtained from commonly used simulation algorithms based on the circular convolution implementations: for example, it is possible to compute, very efficiently, the exact Fresnel transform of some periodic input (object) functions at a number of discrete distances. Another observation regards the perfectly discrete and periodic nature of the continuous Fresnel transform, at some distances, of periodic and discrete input functions.

The fractional Fourier transform formulation provides another approach to handle discretization and computation issues. Sampling issues related to the fractional Fourier transformation, in the sense of the conventional Nyquist sampling results, have been discussed, and will not be reviewed here. Instead, referring to Eq. 8, we observe that naive application of the Nyquist-Shannon approach may require very large sampling rates due to the highly oscillatory nature of the kernel. However, by careful consideration of sampling issues, it is possible to accurately and efficiently compute this integral with a number of samples close to the space-bandwidth product of $f(x)$. This leads to a fast (order of $N \log N$) algorithm for computing the samples of the continuous fractional Fourier transform of a function from the samples of that function. The continuous function obtained by interpolating the computed output samples is an approximation of the actual continuous function in the same sense and to the same degree of approximation as the discrete Fourier transform (DFT) approximates the continuous Fourier transform. Here $N$ is the space-bandwidth product of the signal whose transform is to be computed. It is important to note that $N$ is not allowed to be artificially large as a consequence of the wide bandwidth of the chirp function constituting the kernel of the fractional Fourier transform. These issues are related to the investigation of efficient sampling strategies in diffraction, and the effort to apply the results to efficient computational algorithms.

4. CONCLUSION

The very nature of diffraction phenomena is conveniently suited for the immediate application of various signal processing approaches and algorithms to optics. The coupling of optics and diffraction with digital environments naturally leads to issues related to sampling and quantization. Both established results as well as novel interpretations and improvements in sampling theory provide a rich potential for interesting solutions and applications and efficient implementations in optics.
This work is supported by EC within FP6 under Grant 511568 with the acronym 3DTV.

REFERENCES
43. L. Onural, Digital Decoding of In-line Holograms-PhD Diss., SUNY at Buffalo, Buffalo, N.Y., USA, 1985.