CHAPTER 2

HOLOGRAPHY

2.1 INTRODUCTION

The art of holography has undergone several changes since its invention. The techniques and principles of holography have been modified and improved through years to achieve better efficiency and quality. As explained earlier the aim of this work is also the same, which is to devise a new calculation method for cylindrical digital holography. Among the vast number of methods available, each has its own merits, and drawbacks. None of them are superior than all others in every aspects. So a suitable method should be chosen based on the problem in hand. For this, a thorough analysis of all the evolved methods, their properties including difficulties, is necessary. Hence this chapter will present in detail, the basic theory of holography and the available methods and techniques for holographic recording and reconstruction. The discussion in this chapter will also make clear the need for a cylindrical shaped hologram, and explain its advantages and disadvantages.

2.2 OPTICAL HOLOGRAPHY

The method of holography in which the recording and reconstruction is done using laser light on a holographic plate is called as optical holography. This was the earliest and popular method for making holograms which demands the presence of real object and highly stable vibration free recording environments. The reconstruction is done either with laser or white light. The best quality holograms that very closely resemble the
object were made using this procedure. No digital electronic devices were used for recording or reconstruction. Hence the name optical holography.

Later on with the development in opto-electronic devices digital electronics began to take part in holography. This gave rise to an interesting field of research namely, Computer generated holography which is also called as digital holography now. Holograms of non-existing objects are made using this method. Real time dynamic imaging and reconstruction were also made possible using these methods, which was a major drawback in optical holography.

Eventhough the method was digitised, the theory and methods of optical holography still apply for digital holography. Computer generated holography and digital holography were built on the principles of optical holography. In other words, optical holography is the forefather of digital holography. The work reported in this thesis is also an attempt to digitise such an optical holographic technique proposed by Jeong (1967). Therefore a discussion on various optical holographic methods will throw more light on the foundations of this research work. The following sections explains the various methods available for optical holography and their merits and demerits.

2.2.1 Principle

Light is electromagnetic in nature and hence the theory of holography entirely revolves around the equations of electromagnetic wave propagation and their solutions (Jackson 1998, Born and Wolf 1999). A typical holographic recording and reconstruction setup considered for explaining the theoretical foundations of holography is shown in Figure 2.1.

The principle optical holographic recording is shown in Figure 2.1(a) which consists of the light source, object (to be recorded) and a recording device, e.g. a photographic plate. Light with sufficient coherence
Figure 2.1: Optical system for a) recording and b) reconstructing a hologram
length is split into two partial waves using a beam splitter. The first wave illuminates the object and is called as the object wave. It is scattered at the object surface and reflected to the recording medium. The second wave, named the reference wave, illuminates the light sensitive medium directly. Since they are coherent, both waves interfere to create a standing wave pattern. The interference pattern is recorded by chemical development of the photographic plate. The recorded interference pattern is known as the hologram. The hologram has recorded all the information that came from the object, i.e., both phase and amplitude.

To get back the recorded information, the hologram is illuminated with the same reference beam alone. In other words, the original object wave is reconstructed by illuminating the hologram with the reference wave as shown in Figure 2.1(b). An observer viewing through the hologram sees a virtual image of the object, which resembles the original object itself. There is also a real image and other wavefronts reconstructed, which will be explained below. The reconstructed image exhibits all effects of perspective and depth of focus. The above mentioned recording and reconstruction process can be explained in the language of Mathematics as follows.

The complex amplitude of the object wave is described by

\[ U(x, y) = A_0(x, y)e^{i\phi(x, y)} \]  \hspace{1cm} (2.1)

with real amplitude \( A_0(x, y) \) and phase \( \phi(x, y) \).

\[ R_r(x, y) = A_r(x, y)e^{i\psi(x, y)} \]  \hspace{1cm} (2.2)

is the complex amplitude of the reference wave with real amplitude \( A_r(x, y) \) and phase \( \psi(x, y) \). Both the waves interfere at the surface of the recording medium resulting in an intensity distribution (fringe pattern) across the medium. This
intensity distribution can be calculated as follows.

\[
I(x, y) = |U(x, y) + R_r(x, y)|^2
\]

\[
= |A_0(x, y)|^2 + |A_r(x, y)|^2 + 2A_0(x, y)A_r(x, y) \cos(\psi(x, y) - \phi(x, y))
\]

(2.3) (2.4)

where the last term equals \( UR^*_r + U^*R_r \) and includes both the amplitude and phase of the object wave front, i.e., \( A_0(x, y) \) and \( \phi(x, y) \)

The transmission function of optical recording devices including photographic film is sensitive to intensity. We will assume that the sensitivity is linear in intensity. The reference \( A_r(x, y) \) will be assumed to be constant, and equal to \( A \), which is a plane wave incident perpendicular to the hologram. The transmission function of such a device can be written as

\[
t(x, y) = t_0 + \beta \tau \left[ |A_0(x, y)|^2 + |A_r(x, y)|^2 + UR^*_r + U^*R_r \right]
\]

(2.5)

where \( \beta \) and \( t_0 \) are constants. The constant \( \beta \) is the slope of the amplitude transmittance versus exposure characteristic of the light sensitive material. For photographic emulsions \( \beta \) is negative. \( \tau \) is the exposure time and \( t_0 \) is the amplitude transmission of the unexposed plate. \( t(x, y) \) represents the stored information and is known as the hologram function. In Digital holography where CCD’s are used as recording medium the term \( t_0 \) can be neglected. Now suppose that the generated hologram is illuminated by another reference wave \( R(x, y) \) as shown in Figure 2.1(b). The wave emanating from the hologram can be written as

\[
u(x, y) = \left( t_0 + \beta \tau \left[ |A_0(x, y)|^2 + |A_r(x, y)|^2 + UR^*_r + U^*R_r \right] \right) R(x, y)
\]

(2.6)

\[
R_t = U_1 + U_2 + U_3 + U_4
\]

(2.7)
where,

\[ U_1 = (t_0 + \beta \tau |A_r(x,y)|^2)R(x,y) \]  
\[ U_2 = \beta \tau |A_0(x,y)|^2 R(x,y) \]  
\[ U_3 = \beta \tau R_r(x,y) U(x,y)^* R(x,y) \]  
\[ U_4 = \beta \tau R_r(x,y) U(x,y)^* R(x,y) \]

Suppose that \( R_r \) and \( R \) are the same and are constant, as in a plane wave case perpendicular to the direction of propagation. Then, \( U_3 \) is proportional to \( U \), and \( U_4 \) is proportional to \( U^* \).

The first term \( U_1 \) refers to the intensity reduction of the reconstruction wave by the factor \( t_0 + \beta \tau |A_r(x,y)|^2 R(x,y) \) during reconstruction. The second term is small assuming that we choose \( A_0(x,y) < A_r(x,y) \) during recording. This term is distinguished from the first term by its spatial variation \( |A_0(x,y)|^2 \). The \( |A_0(x,y)|^2 \) term contains low spatial frequencies which have small diffraction angles and create a so-called halo around the reconstruction wave. The size of the halo is given by the angular dimension of the object. These first two terms form the zeroth diffraction order in equation. The third term \( U_3 \) in Equation (2.6) denotes the object wave \( U(x,y) \) multiplied with the constant factor \( \beta \tau R_r^2 \). An observer who registers this wave in his eye therefore sees the virtual image of the (not present) object. The third term is the most important and represents the first diffraction order. The wave travels divergent from the hologram thus creating a virtual image at the position of the original object. It is a virtual image because the wave is not converging to form a real image. This image cannot be captured on a screen. The intensity (square of amplitude) of the image does not depend on the sign of \( \beta \). Therefore it is unimportant whether the hologram is processed “positive” or “negative”.

The fourth term \( U_4 \) is essentially the complex conjugate of the object
wave $U^*$ except for a multiplicative term. This represents the $-1^{st}$ diffraction order. Since it is complex conjugated wave, the phase changes its sign with respect to $U(x,y)$. As a consequence the wave $U^*(x,y)$ travels convergent and forms a real image. The conjugated real image $U_4$ is usually located at the opposite side of the hologram with respect to $U_3$. $U_3$ and $U_4$ are called twin images or also represented as virtual and real images, respectively. All these reconstructed wavefronts are represented in Figure 2.1(b). However, which image is virtual and which is real actually depend on the properties of the reference waves used during recording and reconstruction. These issues are further discussed in the next section.

The virtual image appears at the position of the original object itself, if the hologram is reconstructed with the same parameters like those used in the recording process. However, if one changes the wavelength or the coordinates of the reconstruction wave source point with respect to the coordinates of the reference wave source point used in the recording process, the position of the reconstructed image moves. The coordinate shift is different for all points, thus the shape of the reconstructed object is distorted. The image magnification can be influenced by the reconstruction parameters, too.

The Equations (2.12 to 2.17) are called imaging equations that relate the coordinates of an object point $O$ with that of the corresponding point in the reconstructed image. Only the final equations are mentioned here. The detailed derivations are given by Hariharan (2002), Kreis (2005).

The coordinate system is shown in Figure 2.2. $(x_o, y_o, z_o)$ are the coordinates of the object point $O$, $(x_r, y_r, z_r)$ are the coordinates of the source point of the reference wave used for hologram recording $R$ and $(x_p, y_p, z_p)$ are the coordinates of the source point of the reconstruction wave $P$. $\mu = \lambda_2/\lambda_1$ denotes the ratio between the recording wavelength $\lambda_1$ and the reconstruction wavelength $\lambda_2$. The coordinates of that point in the reconstructed virtual image, which corresponds to the object point $O$, are:
Figure 2.2: Formation of image point object by hologram during a) recording b) reconstruction

(a) Hologram recording

(b) Image reconstruction

\[ x_1 = \frac{x_p z_o z_r + \mu x_o z_p z_r - \mu x_r z_p z_o}{z_o z_r + \mu z_p z_r - \mu z_p z_o} \]  
\[ (2.12) \]

\[ y_1 = \frac{y_p z_o z_r + \mu y_o z_p z_r - \mu y_r z_p z_o}{z_o z_r + \mu z_p z_r - \mu z_p z_o} \]  
\[ (2.13) \]

\[ z_1 = \frac{z_p z_o z_r}{z_o z_r + \mu z_p z_r - \mu z_p z_o} \]  
\[ (2.14) \]
The coordinates of that point in the reconstructed real image, which corresponds to the object point $O$ are:

$$x_2 = \frac{x_p z_o z_r - \mu x_o z_p z_r + \mu x_r z_p z_o}{z_o z_r - \mu z_p z_r + \mu z_p z_o} \quad (2.15)$$

$$y_2 = \frac{y_p z_o z_r - \mu y_o z_p z_r + \mu y_r z_p z_o}{z_o z_r - \mu z_p z_r + \mu z_p z_o} \quad (2.16)$$

$$z_2 = \frac{z_p z_o z_r}{z_o z_r - \mu z_p z_r + \mu z_p z_o} \quad (2.17)$$

An extended object can be considered to be made up of a number of point objects. The coordinates of all the surface points are described by the above mentioned equations. The lateral magnification of the entire virtual image is described as

$$M_{\text{lat,1}} = \frac{dx_1}{dx_o} \left[ 1 + z_o \left( \frac{1}{\mu} - \frac{1}{z_r} \right) \right]^{-1} \quad (2.18)$$

The lateral magnification of the real image is

$$M_{\text{lat,2}} = \frac{dx_2}{dx_o} \left[ 1 - z_o \left( \frac{1}{\mu} + \frac{1}{z_r} \right) \right]^{-1} \quad (2.19)$$

The longitudinal magnification of the virtual image is given by

$$M_{\text{long,1}} = \frac{dz_1}{dz_o} = \frac{1}{\mu} M_{\text{lat,1}}^2 \quad (2.20)$$

The longitudinal magnification of the real image is

$$M_{\text{long,2}} = \frac{dz_2}{dz_o} = \frac{1}{\mu} M_{\text{lat,2}}^2 \quad (2.21)$$
Apart from all these, there is a very important difference between the real and virtual image. Since the real image is formed by the conjugate object wave $U^*(x,y)$, it has the curious property that its depth is inverted. Corresponding points of the virtual image (which coincides with the original object points) and of the real image are located at equal distances from the hologram plane, but at opposite sides of it. The background and foreground of the real image are therefore exchanged. The real image appears inverted. This image is called “pseudoscopic” contrary to the normal image which is called “orthoscopic”.

2.2.2 Methods in Optical Holography

As mentioned earlier there are a lot of methods available for recording and reconstructing holograms optically. Each method has its own merits and demerits. This section discusses briefly each method and its significances. For a detailed description the reader may refer to Ackermann and Eichler (2007) and Hariharan (2002).

2.2.2.1 Inline hologram

We consider the optical system shown in Figure 2.3(a) in which the object (a transparency containing small opaque details on a clear background) is illuminated by a collimated beam of monochromatic light along an axis normal to the photographic plate.

The light incident on the photographic plate then contains two components. The first is the directly transmitted wave, which is a plane wave whose amplitude and phase do not vary across the photographic plate. The second is a weak scattered wave which emanates from the object. Both these waves superimpose on the photographic plate giving rise to fringe pattern which is the hologram. A detailed mathematical derivation for the fringe pattern formed is given by Hariharan (2002).
If the object is chosen as an axial point $O$ emitting spherical wave, the resulting hologram for a plane reference wave is a Fresnel zone lens. During reconstruction the hologram is illuminated with a plane reference wave as shown in Figure 2.3(b). A virtual image point is formed at the original object position and additionally a real image point appears at the same distance to the right of the hologram. The phenomenon also holds for extended objects which can be divided into single points.

This was the earliest method and was developed by Gabor (1948). It was named after him as Gabor’s inline holography. This method has certain demerits. During observation the two images lying on the same axis interfere
which leads to image disturbances. Moreover, the observer looks directly into the reconstruction wave, which is not always safe.

But this method has its own advantages. A single laser beam is used for the recording which constitutes both the object and reference beam without splitting the beam. This technique is also called as “single beam holography”. The fringe density is very low compared to the other methods where there is an angle between the object and reference beams. This significantly reduces the computation load and is of great help to digital holographers.

2.2.2.2 Off-Axis Hologram

The recording arrangement for off-axis holography is shown in Figure 2.4. The reference beam is a collimated beam of uniform intensity, derived from the same source as that used to illuminate the object. The reader may refer to Hariharan (2002) and Ackermann and Eichler (2007) for the detailed mathematical equations for the generated fringe pattern.

By tilting the reference wave (or shifting the object) the three diffraction orders, namely the image, the conjugated image, and the illumination wave, are spatially separated (Leith and Upatnieks 1962, 1963). Hence the unwanted overlapping of the real and virtual image suffered by the inline recording method is avoided. This also has the advantage that, holograms of opaque objects can be produced since the reference wave is not obstructed by the object.

This is the most popular and widely used method for recording holograms optically. But on the contrary this is the least used method by digital holographers. This is due to the fact that, the fringe density increases with angle between the reference and object. Hence the number of samples should be increased in order to completely record all information which in turn affects the computation speed. Hence digital holographers always prefer
Gabor’s inline recording setup. A similar inline recording setup is used for the work reported in this thesis.

![Diagram of optical system for recording and reconstruction of off-axis holograms]

**Figure 2.4:** Optical system to a) record and b) reconstruct off-axis holograms

### 2.2.2.3 Fourier Hologram (Lensless)

If the object and the reference are within the same plane parallel to the hologram, then the so called “Fourier holograms” are generated. It is
also necessary that the reference should be a point source and the object is illuminated with a plane wave. Then a hologram which is similar and has the same properties as that of a Fourier hologram is generated. A schematic of the recording and reconstruction of lensless Fourier holograms is shown in Figure 2.5. This is a Fourier hologram generated without a lens and hence is called as lensless Fourier hologram (Stroke 1965).

![Diagram of lensless Fourier hologram]

**Figure 2.5:** Optical system to a) record and b) reconstruct lensless Fourier holograms
The special property of this hologram is that, like in all thin holograms two images appear during reconstruction but both are virtual now. The regular image is in the position of the original object, while the conjugated one appears in the same plane parallel to the hologram. The point light source that represents the reference will be the center of point symmetry for the two images. The other properties of these holograms are same as a Fourier hologram and are discussed in the next section.

2.2.2.4 Fourier Hologram

The optical setup for recording a Fourier hologram is shown in Figure 2.6(a). A plane object is placed in the first focal plane of the lens. The reference wave emerges from a point light source in the same plane. The photographic layer is placed in the back focal plane of the lens during recording (Vander, L.A. 1964). The reconstruction is done by illuminating the hologram with an axially parallel plane wave as shown in Figure 2.6(b). The hologram is again placed in the first focal plane of a similar second lens. The primary and the conjugated images appear in the second focal plane symmetric to the optical axis. The undiffracted reference wave forms an axial light spot representing the zeroth diffraction order. It can be shown that the reconstructed image remains stationary when the hologram is shifted in its plane.

Fourier holograms have the useful property that the reconstructed image does not move when the hologram is translated in its own plane. This is because a shift of a function in the spatial domain only results in its Fourier transform being multiplied by a phase factor which has no effect on the intensity distribution. This setup is most liked by digital holographers because the simulation is very easy which only requires an FFT calculation. The Fourier holograms are limited only to plane objects and 3D perception is not possible.
2.2.2.5 Fraunhofer Hologram

As explained in the previous section, Fourier holograms are formed by the superposition of spherical waves whose centers have the same distance from the holographic layer. If the layer is moved far away as in Figure 2.7, the centers depart and in the limit plane waves are created. This kind of holograms are called “Fraunhofer holograms”.

A hologram of this type is especially used for the measurement and investigation of aerosols (Thompson et al. 1967, Trolinger 1975). A setup for this is shown in Figure 2.7. The object with radius $r_0$ has to be so small that a diffraction pattern will appear in the far field. The condition for the distance between the object and hologram is $z_0 \ll r_0^2/\lambda$. 

Figure 2.6: Optical system to a) record and b) reconstruct Fourier hologram
2.2.2.6 Image plane Hologram

It has a lot of advantages to record the real image of an object instead of the object itself. For image-plane holograms the object is imaged into the plane of a hologram by a large lens as shown in Figure 2.8. During reconstruction, the real image of extended objects is partly in front of and behind the hologram.

Due to the hologram plane being in the middle of the image the differences in path lengths are smaller than those in other techniques. Hence minimal demands are made regarding the coherence of the light source. If the depth of the object is small even white light sources can be used. Another
advantage is that image-plane holograms are relatively bright and brilliant, though the observation angle is limited by the lens aperture.

2.2.2.7 Rainbow Hologram

Rainbow holograms can be reconstructed in transmission using white light (Benton 1975, 1977). Depending on the viewing direction the reconstructed image appears in different colors, exhibiting the whole light spectrum. The recording and reconstruction setup are shown in Figure 2.9. The technique for the recording of rainbow holograms consists of two steps. In the first step an off-axis hologram is created in the usual manner (Figure 2.9(a)). In the second step, a photosensitive layer is positioned inside the real image and a second hologram $H_2$ is created as shown in Figure 2.9(b). By this process the information of the pseudoscopic image is recorded.

To reconstruct the images of rainbow holograms, they are rotated by $180^\circ$ to create an orthoscopic image from the pseudoscopic one. The image is reconstructed using monochromatic light as shown in Figure 2.9(c). The observer looks through a horizontal slit which is the image of the aperture that was used. A high intensity is achieved since the diffracted light is concentrated on the slit. The viewing angle is limited and the three-dimensional impression exists only in the horizontal direction.

When using the white light for reconstruction, the image of the horizontal slit appears under a different diffraction angle, see Figure 2.9(d). For each spectral color there exists a different viewing slit. If the observer moves the head in the vertical direction he or she will see the image successively in red, orange, yellow, green and blue, i.e., in the spectral colors of the rainbow. Hence this method has the name rainbow holography.
Figure 2.9: Optical system to record and reconstruct a rainbow hologram
2.2.2.8 Double sided Hologram

Figure 2.10: Optical system to record and reconstruct double sided hologram
Usually only the information of the front side of a three-dimensional object can be recorded on a plane hologram. The double-sided holograms can be viewed from both sides (Figure 2.10(d)) and the reconstructed image shows front and backside of the object.

The production of a double-sided hologram starts with the recording of a transmission master hologram H1 of side (1) of the object, see Figure 2.10(a). After that a second hologram H2 of the wavefront from the other side (2) of the object recorded, see Figure 2.10(b). This one is not developed at first and a latent reflection hologram is created. Starting from this hologram H2 the third step consists in creating a double sided hologram. In doing so a pseudoscopic real image of side (1) of the object is generated from the master hologram by inverting the direction of the reference wave. On the second hologram H2 another exposure is made and the wavefront from the master is recorded. The direction of the reference wave is different from the first exposure, see Figure 2.10(c). For the reconstruction the illumination wave has to be inverted again since the image of side (1) was pseudoscopic as shown in Figure 2.10(d). Two independent reflection holograms are obtained which display both sides of the object by a virtual (illumination by \( r_2 \)) and a real image (illumination by \( r_3^* \)).

2.2.2.9 Reflection Hologram

Until now thin holograms were discussed where the object and reference wave impinges from the same side on the photographic layer. In the case of reflection holograms, the reference wave and the object wave impinges from the opposite sides of the photographic layer. Later the reconstruction wave impinges from the observer’s side onto the hologram during reconstruction (Denisyuk 1962, Hariharan 1976, Koch and Petros 1998).
The optical setup for recording a reflection hologram is shown in Figure 2.11(a). The holographic layer is positioned in between the light source and the object. This results in the interference planes being almost parallel to the light sensitive layer. The distance of the grating planes when using a He-Ne laser is $\lambda/2 \approx 0.3 \ \mu m$. So, for 20 grating planes to fit into the recording material, it has to be of almost $6 \ \mu m$ in thickness. Hence the system behaves like thick grating.
The common diffraction theory has to be modified for the reflection holograms because thick gratings exhibit a totally different behavior. During reconstruction the illumination wave which is ideally identical to the reference wave is reflected at the grating planes. The virtual image of the object appears in the reflected light, see Figure 2.11(b). If white light is used for illumination only the wavelength used for the recording is reflected due to the Bragg effect. Therefore a sharp monochromatic image appears although white light is used for reconstruction. This is the advantage of thick reflection holograms which are called “white light holograms”. This holograms are of large importance especially in the field of graphics and art.

2.2.2.10 **Cylindrical Hologram**

A drawback experienced by all the holographic methods discussed so far is the limited angle over which they can be viewed. This is because they were all made only on plane surfaces. Making a hologram in a cylindrical shape can solve this. The cylindrical holography is supposed to make the complete geometry of the object viewable. It can be recorded by using a cylindrical film surrounding the object. Figure 2.12 shows the setup for a single-beam transmission hologram, proposed by Jeong (1967). The object is placed at the center of a glass cylinder which has a strip of photographic film taped to its inner surface with the emulsion side facing inwards, and the expanded laser beam is incident on the object from above. The central portion of the expanded laser beam illuminates the object gets scattered and reaches the cylindrical hologram surface, which constitute the object beam. While the outer portions, which fall directly on the film, constitute the reference beam. The object and reference beams interfere in the cylindrical hologram surface to generate the hologram.
To view the reconstructed image, the processed film is replaced in its original position and illuminated with the same laser beam. When illuminating, all perspectives of the object are displayed when walking around the hologram. Multiplex holograms are also often reconstructed using the 360° geometry. With an illumination from above, all recordings of the multiplex hologram can be reconstructed simultaneously and the recorded changing images can be observed when walking around the complete circle. Hence this method is also called as 360° holography. But the optical recording setup imposes serious difficulties during recording. The geometric shape of the object may prevent some portion of the object being illuminated. Mounting and aligning the optical setup in a vertical direction on the optical table has difficulties. However, generating the hologram using computer can solve all these problem. Accordingly, generating a cylindrical hologram on a computer is the aim of the research work reported in this thesis.

Various holographic recording setups and the merits and demerits of using each one has been discussed. It is clear that none of the holographic setups are not able to reconstruct an object with the “look around property”, except the cylindrical holography. But as explained earlier it faces serious
difficulties with the optical recording and reconstruction setups. The difficulties can be overcome if the object is modeled on the computer and the hologram can be generated in the computer itself. Hence it was intended to do some research on generation of cylindrical holograms on computer and devise a more efficient computation method.

2.3 DIGITAL HOLOGRAPHY

Technologies like photography, signal processing, terrestrial television broadcasting etc, which were born as analog ones, are all digitized today. This is mainly due to the efficiency and convenience that these digital systems offer when compared to their analog counterparts. The other reason is the revolution in electronic and computer research sectors, which kept pouring solutions and advanced instrumentation to any kind of problems in the digitization process. Holography was no exception to this trend and holographists also tried to digitize conventional optical holography. The first successful attempt was reported by Lohman and Paris (1967). Thus was born digital holography and it had many advantages and some disadvantages over its conventional (optical) counterpart. The research work reported in this thesis is also an attempt to digitize a conventional optical holographic method called as cylindrical holography. In order to appreciate the usefulness of the work reported in this thesis, it is worth discussing the basics of digital holography. Hence this section explains the basic principles of digital holography and its various implementation methods with their advantages and disadvantages.

2.3.1 Principle

During the holographic recording or reconstruction process, if the hologram is in digital form, then the technique is called as Digital holography. It is usually done by simulating the optical holographic recording or reconstruction setup on a computer. The schematic of this process is shown in
Figure 2.13: Digital holography - Schematic

Figure 2.13. Figure 2.13(a) represents the process where the object to be recorded is modeled in the computer. Then the wave propagation from object to the hologram plane is simulated. The same simulation is done for the reference, and the interference between the object and reference is calculated at the hologram plane. Hence the required hologram is generated. This generated hologram is in digital format and can be transferred to a photographic film using image setter or other methods. Then the photographic film can be reconstructed using optical methods. Thus hologram of an object that does not exist or cannot be optically recorded, can be produced using this
method. This method is usually used to make display holograms and is also known by the term computer generated holography.

Figure 2.13(b) shows the process in which the hologram is recorded optically, but a CCD is used for recording instead of holographic plate. Thus the recorded hologram is in digital format. The hologram is fed into a computer and reconstructed on the computer. For this the wave propagation is simulated from hologram plane to object plane. The object that was recorded can be viewed in the computer screen.

Simulating wave propagation from object to hologram plane or vice versa, occurs in both the process and is the most important step in digital holography. This is basically a signal processing problem. The research work reported in this thesis basically deals with display holography. Hence the following sections will discuss only problems related to display holography.

Two fundamental signal processing problems in holographic display are referred to as forward and reverse problems. The forward problem is the computation of the light field distribution which arises over the entire 3-D space from a given 3-D scene or object. In traditional holography, this light field would have been optically created and recorded by interferometric and other techniques, but in digital holographic systems the associated field must be computed. This is considerably more difficult problem because the 3-D scene consists of nonplanar surfaces. In other words simulation of wave propagation is the heart of digital holography.

Once the desired field is computed, physical devices will be used to create it at the display end. The field generated by these devices will propagate in space and reach the viewer, creating the perception of the original 3D-scene. These devices impose many constraints on the 3-D light distributions they can generate, as a consequence of their particular characteristic and limitations. Therefore, given a physical device, such as a specific SLM, finding driving signals to get the best approximation to the
desired time varying 3-D light field is a challenging inverse problem. A precise definition of this, so called synthesis problem and some proposed solutions can be found in the literature (Piestun 2001, Piestun and Shamir 2002, Piestun and Miller 2000, Lohmann 2002). The various display devices available for this purpose are discussed in section 2.3.7.

Computation of propagating electromagnetic field depends on the foundations of diffraction theory (Goodman 2004, Born and Wolf 1999, Iizuka 1987). Approaches in solving diffraction problems can be investigated under four categories. From rather simple to more complicated categories, these categories are ray optics, wave optics, electromagnetic optics and quantum optics. Ray optics describes the propagation of light by using geometrical rules and rays (Saleh and Teich 1991). In wave optics, the propagation of light is described by a scalar wave function which is a solution of the wave equation. The work reported in this thesis also uses the wave optics for simulating wave propagation of light. Hence the theory of wave optics is presented in detail in section 2.3.2 and the various wave optic techniques and corresponding fast algorithms are reviewed in section 2.3.3. Based on the computation models many methods have been proposed for Computer generated holography which are explained in section 2.3.4.

Other signal processing approaches have also been extensively employed in problems related to wave optics. However the present state-of-the-art does not seem to be sufficient for solving some of the problems arising in real-time holographic, 3-D display. In order to facilitate further developments, several signal processing tools which has the potential of advancing the state-of-the-art has been discussed in section 2.3.6.

Another problem of fundamental nature is the discretization of signals associated with propagating optical waves. At the acquisition stage, CCD or CMOS arrays capture holographic patterns and convert them into digital signals (Yaroslavsky 2003, Kreis 2002, 2005). While sampling and
quantization is an extensively studied and mature field in the general sense, direct application of the general results will not be efficient in most diffraction related problems. Instead, systematic approaches which take the specific properties of the underlying signals into consideration and merge them with modern digital signal processing methods are highly desirable. The literature dealing with discretization and quantization issues in diffraction and holography are reviewed in section 2.3.5.

2.3.2 Electromagnetic Wave Propagation

Light is electromagnetic in nature and electromagnetic field anywhere in space is well defined by the Maxwell’s equations. The propagation of electromagnetic field is defined by the wave equation. Analytical solution to wave equation describes the wavefield due to a propagating wave front anywhere in space. But in digital holography, the object has arbitrary shape and size and hence analytical solutions to the wave equation is not possible. So numerical solution to the wave equation is sought to calculate the wavefield in the hologram plane or reconstruction plane. Wave equation is a vectorial differential equation and numerically solving it is very time consuming. Moreover sampling errors and discretization errors creep in when the distance of propagation increases, affecting the results very badly. To overcome these issues, approximations have been induced into the equation based on the problem in hand. The approximated equations are integral equations derived from the Helmholtz differential equation using a suitable Greens function. These integral equations are scalar in nature and hence are also called as scalar diffraction formulas. These approximated solutions make calculation much easier and faster, but at the same time give satisfying results in holography. The scalar diffraction formulae are most used ones in Digital holography. The research work reported in this thesis is also an attempt to derive out a new scalar diffraction formula for digital cylindrical holography. Hence it is worth discussing the various scalar diffraction theories, the
approximation conditions and their significances.

Maxwell’s equations in terms of \( E(r) \) and \( H(r) \) can be written as

\[
\begin{align*}
\nabla \cdot \epsilon E &= 0 \quad (2.22) \\
\nabla \cdot \mu H &= 0 \quad (2.23) \\
\nabla \times H &= \frac{\epsilon}{\mu} \frac{\partial E}{\partial t} \quad (2.24) \\
\nabla \times E &= -\mu \frac{\partial H}{\partial t} \quad (2.25)
\end{align*}
\]

where,

\[
\begin{align*}
E &\rightarrow \text{electric field (V/m)} \\
H &\rightarrow \text{magnetic field (A/m)} \\
\epsilon &\rightarrow \text{permittivity (F/m)} \\
\mu &\rightarrow \text{permeability (H/m)}
\end{align*}
\]

The field vectors \( E \) and \( H \) both are functions of position \((x, y, z)\) and time \( t \). As seen from Equations (2.22 and 2.25), Maxwell’s equations relate the field vectors by means of simultaneous differential equations. On elimination we obtain differential equations which each of the vectors must satisfy separately. For this, we apply the \( \nabla \times \) operation to the left and right sides of Equation (2.25).

\[
\nabla \times (\nabla \times E) = \nabla (\nabla \cdot E) - \nabla^2 E \quad (2.26)
\]

Let the propagation medium be linear, isotropic, homogeneous and nondispersive. Substituting the two Maxwell’s equations for \( E \), Equations (2.22 and 2.25) into Equation (2.26) yields

\[
\nabla^2 E - \epsilon \mu \frac{\partial^2 E}{\partial t^2} = 0 \quad (2.27)
\]
The permeability ($\mu$) and permittivity ($\varepsilon$) are related with the wave velocity $v$ and refractive index $n$ as follows:

$$v = \frac{c}{\sqrt{\mu \varepsilon}}$$  \hspace{1cm} (2.28)

where, $c$ is the velocity of light in vacuum.

$$n = \frac{c}{v}$$  \hspace{1cm} (2.29)

Again, from Equation (2.28) and Equation (2.29) it could be derived that

$$n = \frac{1}{\sqrt{\mu \varepsilon}}$$  \hspace{1cm} (2.30)

Substituting Equations (2.28, 2.29 and 2.30) in Equation (2.27) yields,

$$\nabla^2 E - \frac{n^2}{c^2} \frac{\partial^2 E}{\partial t^2} = 0$$  \hspace{1cm} (2.31)

In the similar way we can obtain an equation for $H$ alone, which can be written as follows

$$\nabla^2 H - \frac{n^2}{c^2} \frac{\partial^2 H}{\partial t^2} = 0$$  \hspace{1cm} (2.32)

Equation (2.31) and Equation (3.3) are the standard equations of electromagnetic wave motion propagating with a velocity $c$.

Since the same vector wave equation is obeyed by both $E$ and $H$, it is possible to summarize the behavior of all components of $E$ and $H$ ($E_x, E_y, E_z, H_x, H_y, H_z$) through a single scalar wave equation.

$$\nabla^2 V(x, y, z, t) - \frac{n^2}{c^2} \frac{\partial^2 V(x, y, z, t)}{\partial t^2} = 0$$  \hspace{1cm} (2.33)
Hence, if the medium of propagation is linear, isotropic, homogeneous and nondispersive, all components of electric and magnetic field behave identically and their behavior is fully described by a single scalar wave equation as shown in Equation (2.33). However there is coupling between the components of electric and magnetic field at the boundaries. Hence even if the medium is homogeneous, the use of scalar theory entails some degree of error. But the error will be small and satisfactory results could be obtained, if the boundary conditions have effect over an area that is a small part of the area through which a wave may be passing. The wave propagation very well satisfies this condition in this research work and hence we turn our interest towards the scalar wave equation. The scalar wave equation can still be simplified on inducing certain approximating conditions. These approximated equations are integral equations and are much easier for numerical evaluation. These approximated scalar wave equations are generally known as the scalar diffraction theories. The following explains the various diffraction theories and their approximating conditions.

For a monochromatic wave, the scalar field may be written explicitly as,

$$V(x, y, z, t) = U(x, y, z)e^{-i2\pi vt}$$  \hspace{1cm} (2.34)

where,

$$U(x, y, z) = U(P) = A(x, y, z)e^{i\phi(x, y, z)}$$  \hspace{1cm} (2.35)

where $A(x, y, z)$ and $\phi(x, y, z)$ are the amplitude and phase, respectively, of the wave at position $(x, y, z)$. $v$ is the frequency of the propagating wave. If this scalar field represents a propagating optical field, then it must satisfy the scalar wave equation represented in Equation (2.33) at each source free point. The complex function $U(x, y, z)$ serves as an adequate description of the wave, since the time dependence is known a priori. Accordingly, when Equation (2.34) is substituted in Equation (2.33) it follows
that \( U(x, y, z) \) shown in Equation (2.35) must obey the time-independent equation.

\[
(\nabla^2 + k^2)U = 0 \tag{2.36}
\]

where

\[
k = 2\pi n \frac{\nu}{c} = \frac{2\pi}{\lambda} \tag{2.37}
\]

Equation (2.36) is known as the Helmholtz Equation. It can be very well stated that the complex amplitude of any monochromatic optical disturbance propagating in vacuum \((n = 1)\) or in a homogeneous dielectric medium \((n > 1)\) must obey Equation (2.36). The Helmholtz equation is the starting point for the derivation of scalar diffraction theories.

Before exploring the different diffraction theories, it is worth introducing the concept of diffraction. Diffraction is a phenomenon of considerable importance in the fields of physics and engineering whenever wave propagation is involved. Sommerfeld defined diffraction as “any deviation of light rays from rectilinear paths which cannot be interpreted as reflection or refraction” (Sommerfeld 1954). In 1665, the first account of diffractive phenomena was published by Grimaldi when he observed the shadow resulting from an aperture in an opaque screen illuminated by a light source. He observed that the transition from light to shadow was gradual rather than sharp. Sommerfeld’s definition implies that diffraction only applies to light rays. In reality, diffraction occurs with all types of waves including electromagnetic, acoustic, and water waves, and is present at all frequencies. The content of this thesis deals exclusively with electromagnetic radiation at optical frequencies.

Diffraction was initially considered to be a nuisance when designing optical systems because diffraction at the apertures of an optical imaging system is often the limiting factor in the system’s resolution. However, by the mid 1900’s, methods and devices utilizing the effects of
Diffraction began to emerge. Examples include analog holography, synthetic aperture radar, computer-generated holograms, digital holography and kinoforms, (also known as diffractive optical elements). As mentioned earlier, among these Digital holography is the main topic of this thesis.

The propagation of waves can often be described by rays which travel in straight lines (geometric optics). However, the behavior of wave fields encountering obstacles cannot be described by rays. Some of the wave encountering an obstacle will deviate from its original direction of propagation causing the resulting wave field to differ from the initial field at the obstacle. This is called diffraction. In other words “diffraction is a general characteristic of wave phenomena occurring whenever a portion of a wavefront be it sound, matter wave or light obstructed in some way”. Classic examples include diffraction of light from a knife’s edge and a wave field passing though an aperture in an opaque screen.

## 2.3.2.1 Huygens Fresnel Principle

The initial step in the evolution of a theory that would explain diffraction was made by Christian Huygens in the year 1678. Huygens expressed an intuitive conviction that if each point on the wavefront of a light disturbance was considered to be a new source of “secondary” spherical disturbance, then the wavefront at a later instant could be found by constructing the “envelope” of the secondary wavelets. But the technique ignores most of the secondary wavelets, retaining only that portion common to the envelope. As a result of this inadequacy, Huygens principle by itself was unable to account for the details of diffraction process.

The difficulty was resolved by Fresnel by the addition of the concept of interference. The corresponding Huygens-Fresnel Principle states that “every unobstructed point of wavefront, at a given instant, serves as a source of spherical secondary wavelets. The amplitude of the optical field at
any point beyond is the superposition of all these wavelets (considering amplitudes and relative phases)”. Fresnel was able to calculate the distribution of light in diffraction patterns with excellent accuracy. The calculations are worked out by Hect (2003). Huygens-Fresnel Principle had a few shortcomings. First of all the whole thing is rather hypothetical. Again, according to the principle at any instant every point on the primary wavefront is envisioned as a continuous emitter of spherical secondary wavelets. But if each wavelet radiated uniformly in all directions, in addition to the generating and ongoing wave, there would also be a reverse wave traveling back toward the source. No such wave is found experimentally.

The ideas of Huygens and Fresnel were put on a firm mathematical foundation by Gustav Kirchhoff. He showed that the amplitudes and phases ascribed to the secondary sources by Fresnel were indeed logical consequences of wave nature of light. He developed his rigorous theory based directly on the solution of Helmholtz wave equation Equation (2.36) using Green’s theorem. The complete derivation is given by Born and Wolf (1999). Accordingly the complex amplitude $U(P)$ defined in Equation (2.35) is given by

$$U(P) = \frac{1}{4\pi} \iiint_S \left[ U \frac{\partial}{\partial n} \left( \frac{e^{ikS}}{S} \right) - \frac{e^{ikS}}{S} \frac{\partial U}{\partial n} \right] dS$$ (2.38)

where $S$ is the field boundary.

Thus Kirchoff showed that, Huygens-Fresnel principle is an approximate form of a certain integral theorem which expresses the solution of the homogeneous wave equation at an arbitrary point in the field, in terms of the values of the solution and its first derivatives at all points on the arbitrary closed surface surrounding P. Equation (2.38) is one form of the integral theorem of Helmholtz and Kirchoff. This integral theorem embodies the basic idea of Huygens-Fresnel principle but the laws governing the contributions from different elements of the surface are more complicated that
Fresnel assumed. Kirchoff showed that, in many cases the theorem can be reduced to an approximate more simpler form. This resulted in the Kirchoff diffraction theory.

2.3.2.2 Kirchoff Diffraction theory

Consider a wave propagating from a point source $P_0$ through an opening in a plane opaque screen as shown in Figure 2.14(a). Let $P$ be the
point at which the light disturbance is to be determined. The linear dimensions of the opening are large compared to the wavelength but are small compared to the distances of both \( P_0 \) and \( P \) from the screen.

To find the disturbance at \( P \) we take the Kirchoff’s integral over a surface \( S \) formed by (1) the opening \( A \), (2) the non illuminated portion \( B \) (3) the portion \( C \) which is a large sphere of radius \( R \) as shown in Figure 2.14(a). All the portions together form a closed surface. Now applying Kirchoff’s law expressed by Equation (2.38) gives

\[
U(P) = \frac{1}{4\pi} \left[ \iint_A + \iint_B + \iint_C \right] \left[ U \frac{\partial}{\partial n} \left( \frac{e^{iks}}{s} \right) - \left( \frac{e^{iks}}{s} \right) \frac{\partial U}{\partial n} \right] ds \tag{2.39}
\]

Kirchhoff accordingly adopted the following assumptions to the problems.

1. Across the surface \( A \), the field distribution \( U \) and its derivative \( \frac{\partial U}{\partial n} \) are exactly the same as they would be in the absence of the screen.

2. Over the portion of \( B \), that lie in the geometrical shadow of the screen, the field distribution \( U \) and its derivative \( \frac{\partial U}{\partial n} \) are identically zero.

These conditions are commonly known as Kirchhoff boundary conditions. The first allows us to specify the disturbance incident on the aperture by neglecting the presence of the screen. The second allows us to neglect all of the surface integration except that portion lying directly within the aperture itself. On applying these conditions to Equation (2.39) it reduces to

\[
U(P) = -\frac{iA}{2\lambda} \iint_A \frac{e^{ik(r+s)}}{rs} [\cos(n,r) - \cos(n,s)] dS \tag{2.40}
\]

The detailed derivation of the result shown in Equation (2.40) is given by Goodman (2004). This result, which applies only for an illumination
consisting of a single point source, is commonly known as the Fresnel-Kirchhoff diffraction formula. It allows one to calculate the optical disturbance at a point in space due to a diffracting object. Kirchhoff mathematical development demonstrated that Huygen-Fresnel assumptions were in fact natural consequence of wave nature of light.

Fresnel-Kirchhoff formula in Equation (2.40) closely resembles Huygens-Fresnel way of computation that involves superposition of secondary wavelets except for the fact that, the secondary sources

1. Differ in amplitude from the primary by a factor $\lambda - 1$.
2. Reduce in amplitude from the primary by a factor $\frac{1}{2}[\cos(n, r) - \cos(n, s)]$ called the obliquity factor.
3. Lags in phase by $90^\circ$ from the primary.

The Kirchhoff theory has been found experimentally to yield remarkably accurate results and is widely used in practice. However, numerically evaluating this integral in a PC is tedious and also time consuming when compared with other formula (which will be discussed shortly). So this formula is usually not used to simulate wave propagation in digital holography.

More over there are certain internal inconsistencies in this theory also. It is a well known theorem of Potential theory that if a two dimensional potential function and its derivative vanish together along any finite curve segment, then that potential function must vanish over the entire plane. Similarly, if a solution of the three dimensional wave equation vanishes on any finite surface element, it must vanish in all space. Thus the Kirchhoff’s two boundary conditions imply that the field is identically zero everywhere behind the aperture, a result that contradicts the physical situation. A further indication of these inconsistencies is the fact that the Fresnel-Kirchhoff
diffraction formula can be shown to fail to reproduce the boundary condition as the observation point approaches the screen or aperture.

2.3.2.3 Rayleigh-Sommerfeld diffraction formula

The inconsistencies of the Kirchhoff theory were removed by Sommerfeld, who eliminated the necessity of imposing boundary values on both the disturbance and its normal derivative simultaneously.

Suppose the Kirchhoff theory was modified in such a way that, either $U$ or $\frac{\partial U}{\partial n}$ vanishes over the entire surface $B$, and not both. Then the necessity of imposing simultaneous boundary conditions on $U$ and $\frac{\partial U}{\partial n}$ would be removed, and hence the inconsistencies eliminated. Sommerfeld pointed out that Greens function with the required property do indeed exist.

Accordingly the Kirchhoff boundary condition may now be applied to $U$ alone (not $\frac{\partial U}{\partial n}$), which yields the following result (Born and Wolf 1999).

$$U(P) = -\frac{iA}{\lambda} \int_A \int e^{ik(r+s)} \frac{\cos(n,r)}{rs} dS$$

(2.41)

This expression is known as Rayleigh-Sommerfeld diffraction formula. It yields wonderful results and has also removed the inconsistencies suffered by Fresnel- Kirchhoff. This formula is also not usually used in digital holography due to the complexity in numerical evaluation. It should be noted that in Kirchhoff and Sommerfeld theories, light is treated as a scalar phenomenon; i.e. only the scalar amplitude of one transverse component of either the electric or the magnetic field is considered. Any other component of interest can be treated independently in the similar manner. Such an approach entirely neglects the fact that the various components of electric and magnetic field vectors are coupled through Maxwell's equations and cannot be treated independently. But experiments in the microwave regions (Silver 1962) have
shown that scalar theory yields very accurate results if two conditions are met

1. The diffracting aperture must be large compared with the wavelength.

2. The diffracted fields must not be observed too close to the aperture.

Born and Wolf (1999) have presented a complete discussion on the applicability of scalar diffraction.

The vectorial nature of the fields must be taken into account if reasonably accurate results are to be obtained. Vectorial generalizations of diffraction theory do exist. The first satisfactory one was proposed by Kottler (1923). The first truly rigorous solution of a diffraction problem was given in 1896 by Sommerfeld (1896). These theories are of no interest with regard to the work in this project.

2.3.2.4 Convolution Integral

It is also possible to formulate scalar diffraction theory in a framework that closely resembles the theory of linear, invariant systems. Accordingly the Rayleigh-Sommerfeld formula given by Equation (2.41) can also be expressed as follows (Goodman 2004).

\[ U(P) = \frac{e^{i k r}}{\lambda} \int_A U(P_0) \frac{e^{i k r}}{r} \cos(\theta) ds \]  

(2.42)

where \( \theta \) is the angle between the vectors \( n \) and \( r \). Now, Equation (2.42) is no more than a superposition integral. To make the point clear, Equation (2.41) can be re-written as (Goodman 2004)

\[ U(P) = \int_A h(P, P_0) U(P_0) ds \]  

(2.43)
where \( h(P, P_0) \) is given explicitly by (Goodman 2004)

\[
h(P, P_0) = \frac{iA}{2\lambda} \int_A e^{i\lambda(r)} \left[ \frac{\cos(\theta)}{r} \right]
\]

Equation (2.43) can be calculated using a convolution operation. The primary ingredient required for such a result is the property of linearity. The function \( h(P, P_0) \) is called as the point response function or the impulse response function. If the system is space-invariant, then Fast Fourier Transform can be used to evaluate this equation. Hence the numerical computation becomes more fast and easy. Hence this formula holds an important place in Digital holography especially when the propagation distances are very small. Thus Huygens-Fresnel principle is nothing but a convolution integral.

2.3.2.5 Angular Spectrum of plane waves

Another formulation of the scalar diffraction theory in the framework of linear invariant systems theory is the angular spectrum of plane waves. It is very important to discuss this method because the research work in this thesis is based on this method. Hence the following discusses this in detail.

Let us consider the same situation where the wave field \( U(x, y, z) \) travels in the \( z \) direction. The wavefield is assumed to have a wavelength \( \lambda \) such that \( k = \frac{2\pi}{\lambda} \). Let \( z = 0 \) initially. The 2-D Fourier representation of \( U(x, y, 0) \) is given in terms of its Fourier transform \( A(f_x, f_y, 0) \) by (Ersoy 2007).

\[
U(x, y, 0) = \int_{-\infty}^{\infty} A(f_x, f_y, 0)e^{i2\pi(f_x x + f_y y)} \, df_x \, df_y
\]
where
\[ A(f_x, f_y, 0) = \int_{-\infty}^{\infty} U(x, y, 0)e^{-j2\pi(f_x x + f_y y)} \, dx \, dy \] (2.46)

Including time variable to the integrand of Equation (2.46) gives
\[ A(f_x, f_y, 0)e^{i2\pi(f_x x + f_y y + f t)} \]. This represents a plane wave at \( z = 0 \) propagating with direction cosines \((\alpha, \beta, \gamma)\). Such a plane wave has a complex representation of the form (Goodman 2004).

\[ p(x, y, z, t) = e^{i(k \cdot r - 2\pi vt)} \] (2.47)

where \( r = x\hat{x} + y\hat{y} + z\hat{z} \) is the position vector and \( \vec{k} = \frac{2\pi}{\lambda}(\alpha\hat{x} + \beta\hat{y} + \gamma\hat{z}) \). The direction cosines are interrelated through

\[ \gamma = \sqrt{1 - \alpha^2 - \beta^2} \] (2.48)

Thus across the plane \( z = 0 \) the complex exponential function \( e^{i2\pi(f_x x + f_y y + f t)} \) can be regarded as representing a plane wave propagating with direction cosines \( \alpha = \lambda f_x , \beta = \lambda f_y , \gamma = \sqrt{1 - (\lambda f_x)^2 - (\lambda f_y)^2} \). In the Fourier decomposition of \( U \), the complex amplitude of the plane-wave component with spatial frequencies \((f_x, f_y)\) is simply \( A(f_x, f_y; 0) \) (with the time components discarded) evaluated at \((f_x = \alpha/\lambda, f_y = \beta/\lambda)\). For this reason the function

\[ A(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}; 0) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(x, y, 0)e^{-j2\pi(\frac{\alpha}{\lambda} x + \frac{\beta}{\lambda} y)} \, dx \, dy \] (2.49)

is called as the angular spectrum of the disturbance \( U(x, y, z) \).
Consider now the wavefield \( U(x, y, z) \) parallel to \((x, y)\) but at a distance \( z \) from it given by

\[
U(x, y, z) = \iint_{-\infty}^{\infty} A(f_x, f_y; z) e^{i2\pi(f_xx + f_yy)} df_x df_y \tag{2.50}
\]

Let the function \( A(f_x, f_y, z) \) represents the angular spectrum of \( U(x, y, z) \). That is

\[
A(f_x, f_y; z) = \iint_{-\infty}^{\infty} U(x, y, z) e^{-i2\pi(f_xx + f_yy)} dx dy \tag{2.51}
\]

Now if the relation between \( A(f_x, f_y) \) and \( A(f_x, f_y, z) \) can be found, then the effects of the wave propagation on the angular spectrum can be determined. To find the relation, let us consider the fact that, \( U(x, y, z) \) satisfies the Helmholtz equation at all source-free points namely,

\[
\nabla^2 U(x, y, z) + k^2 U(x, y, z) = 0 \tag{2.52}
\]

Substitution of \( U(x, y, z) \) from Equation (2.51) into Equation (2.52) yields

\[
\int_{-\infty}^{\infty} \left[ \frac{d^2}{dz^2} A(f_x, f_y; z) + (k^2 - 4\pi^2(f_x^2 + f_y^2)) A(f_x, f_y; z) \right] e^{i2\pi(f_xx + f_yy)} df_x df_y = 0 \tag{2.53}
\]

This is true for all the waves only if the integrand is zero.

\[
\frac{d^2}{dz^2} A(f_x, f_y; z) + (k^2 - 4\pi^2(f_x^2 + f_y^2)) A(f_x, f_y; z) = 0 \tag{2.54}
\]
An elementary solution to this differential equation can be written of the format

$$A(f_x, f_y, z) = A(f_x, f_y, 0)e^{i\mu z} \quad (2.55)$$

where

$$\mu = \sqrt{k^2 - 4\pi^2(f_x^2 + f_y^2)} \quad (2.56)$$

or in terms of direction cosines the solution can be written as

$$A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}; z\right) = A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}; 0\right)e^{i\mu z} \quad (2.57)$$

where

$$\mu = \frac{2\pi}{\lambda} \sqrt{1 - \alpha^2 - \beta^2 z} \quad (2.58)$$

These results demonstrate that when the direction cosines \((\alpha, \beta)\) satisfy

$$\alpha^2 + \beta^2 < 1 \quad (2.59)$$

i.e., when \(\mu\) is real, the effect of propagation over distance \(z\) is simply a change of the relative phases of the various components of the angular spectrum by a phase factor \(e^{i\mu z}\). Since each plane wave component propagates at a different angle, each travels a different distance between two parallel planes, and relative phase delays are thus introduced. Plane wave components satisfying this condition are known as homogeneous waves.

However when \((\alpha, \beta)\) satisfy

$$\alpha^2 + \beta^2 > 1 \quad (2.60)$$

then \(\alpha\) and \(\beta\) are no longer interpretable as direction cosines. The
square root in Equation (2.58) and Equation (2.56) becomes imaginary and hence the Equation (2.55) becomes

\[ A(f_x, f_y, z) = A(f_x, f_y, 0)e^{-\mu z} \quad (2.61) \]

Since \( \mu \) is a positive real number, the wave components are strongly attenuated by the propagation in the \( z \)-direction. They are called as evanescent waves. These evanescent waves carry no energy from the aperture.

Hence knowing \( A(f_x, f_y, z) \) in terms of \( A(f_x, f_y, 0) \) allows us to find the wavefield at \( (x, y, z) \) by using Equation (2.61) in Equation (2.50);

\[ U(x, y, z) = \int_{-\infty}^{\infty} A(f_x, f_y, 0)e^{iz\sqrt{k^2 - 4\pi^2(f_x^2 + f_y^2)}}e^{i2\pi(f_x x + f_y y)} df_x df_y \quad (2.62) \]

Thus, if \( U(x, y, 0) \) is known, \( A(f_x, f_y, 0) \) can be computed, followed by the computation of \( U(x, y, z) \). The limits of integration in Equation (2.62) can be limited to a circular region given by

\[ 4\pi^2(f_x^2 + f_y^2) \leq k^2 \quad (2.63) \]

provided the distance \( z \) is at least several wavelengths long so that the evanescent waves may be neglected. Under these conditions, Equation (2.55) shows that wave propagation in a homogeneous medium is equivalent to a linear 2-D spatial filter with the transfer function given by

\[ H(f_x, f_y) = e^{iz\sqrt{k^2 - 4\pi^2(f_x^2 + f_y^2)}} \quad (2.64) \]
or, in terms of direction cosines as

\[ H(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}) = e^{i \frac{\pi}{2} \sqrt{1 - \alpha^2 - \beta^2}} \]  

(2.65)

Equation (2.62) can also be represented in another form using the wavenumber, which provides more insight and easy understanding. For this, the relations \( f_x = 2\pi k_x \), \( f_y = 2\pi k_y \) and \( k_z = \sqrt{k^2 - k_x^2 - k_y^2} \) are used in Equation (2.62) which reduces to

\[
U(x, y, z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(k_x, k_y; 0) e^{i(k_z z)} e^{i(k_x x + k_y y)} dk_x dk_y \]  

(2.66)

Propagation of a wavefield in the \( z \)-direction in a source-free space is correctly described by the propagation of the angular spectrum in the near field as well as the far field. This is also often described as wave propagation in spectral domain. Two other ways to characterize such propagation are in terms of the Fresnel and Fraunhofer approximations which will be discussed in the next section.

Let \( F \) and \( F^{-1} \) denote the forward and inverse Fourier transform operators respectively. In terms of these operators, Equation (2.62) can be represented as

\[
U(x, y, z) = F^{-1} \left[ F[U(x, y, 0)] e^{ikz \sqrt{1 - \alpha^2 - \beta^2}} \right] \]  

(2.67)

The above equations show that the numerical computation of the formula is also very easy and can be done using Fast Fourier transform if the shift invariant property exists. Hence this formula is more frequently used in
digital holography especially when working in the near field. The research work reported in this thesis is to develop the same formula for wave propagation in cylindrical co-ordinates and apply it to digital cylindrical holography.

2.3.2.6 Fresnel diffraction

The Rayleigh-Sommerfeld formula can be made easier to compute by making approximations based on the size of the aperture, the distance $z$ of the output plane from the aperture, and the observed region of the output plane. For that, we consider a rectangular reference system as shown in figure 2.15. Accordingly, the screen with aperture is assumed to be planar with a rectangular coordinate system $(\xi, \eta)$ attached. The region of observation is also assumed to be a plane, standing parallel with the plane of the screen at a normal distance $z$. A coordinate system $(x, y)$ is attached to the plane of observation.

![Figure 2.15: Diffraction geometry](image)

According to Equation (2.42) the Huygens-Fresnel principle can be stated as follows (Goodman 2004).

$$U(P) = \frac{iA}{\lambda} \int_A \int_A U(P_0) e^{ikr} \cos(\theta) ds$$

(2.68)

Now, we begin the approximations
1. The distance between the points $P_0$ and $P$ are large compared to the linear dimensions of the aperture. Then the factor $\cos(\theta)$ in Equation (2.68) will not vary appreciably over the aperture. Hence it can be approximated to be $\cos(\theta) = \frac{z}{r}$.

2. The distance $r$ between the points $P_0$ and $P$ is given exactly by

$$r = \sqrt{z^2 + (x - \xi)^2 + (y - \eta)^2} \tag{2.69}$$

By expanding Equation (2.69) using binomial expansion theorem and neglecting the higher order terms, $r$ turns out to be

$$r \approx z \left[ 1 + \frac{1}{2} \left( \frac{x - \xi}{z} \right)^2 + \left( \frac{y - \eta}{z} \right)^2 \right] \tag{2.70}$$

On applying these two approximations to Equation (2.42), it becomes

$$U(x, y) = e^{ikz} \frac{e^{i k z}}{j \lambda z} \int_{-\infty}^{\infty} U(\xi, \eta) \exp \left( j \frac{k}{2z} \left[ (x - \xi)^2 + (y - \eta)^2 \right] \right) d\xi d\eta \tag{2.71}$$

On factorizing Equation (2.71) yields,

$$U(x, y) = e^{ikz} \frac{e^{i k z}}{j \lambda z} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[ U(\xi, \eta) e^{i \frac{1}{2z} \left( \xi^2 + \eta^2 \right)} \right] e^{-j \frac{2\pi}{\lambda z} (\xi x + \eta y)} d\xi d\eta \tag{2.72}$$

From Equation (2.72) we could recognize that, (apart from the multiplicative factors) it is nothing but the Fourier Transform of the product of the complex field just to the right of the aperture and a quadratic phase
exponential. Both the results, Equation (2.71) and Equation (2.72) are referred to as the Fresnel diffraction Integral.

It can also be seen from Equation (2.71) that, it is nothing but a convolution integral expressible as

\[
U(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [U(\xi, \eta)h(x - \xi, y - \eta)] d\xi d\eta
\]  

(2.73)

where the convolution kernel is

\[
h(x, y) = \frac{e^{ikz}}{j\lambda z} exp \left[ \frac{jk}{2z} (x^2 + y^2) \right]
\]  

(2.74)

This convolution integral can also be computed using Fast Fourier transform, if the shift-invariance relation holds. Due to these facts, Fresnel diffraction formula is more frequently used in digital holography in the near field.

2.3.2.7 Fraunhofer diffraction

Fresnel Diffraction-pattern calculations can be further simplified by applying more stringent approximations. In the Fresnel diffraction the observed field strength \( U(x, y) \) can be found out from the Fourier transform of the product of the aperture distribution \( U(\xi, \eta) \) with a quadratic phase function \( e^{i(l/2z)(\xi^2+\eta^2)} \). If in addition the stronger assumption (Goodman 2004).

\[
z \gg \frac{k(\xi^2 + \eta^2)_{\text{max}}}{2}
\]  

(2.75)

is satisfied, then the quadratic phase factor is approximately unity over the entire aperture. Accordingly the diffraction formula reduces to
U(x, y) = \frac{e^{j kz} e^{\frac{j}{j \lambda z} (x^2 + y^2)}}{\int_{-\infty}^{\infty} U(\xi, \eta) \exp \left[ -j \frac{2\pi}{\lambda z} (x\xi + y\eta) \right] d\xi d\eta}

The observed field distribution can now be directly found from the Fourier Transform of the aperture distribution itself. The transform must be evaluated at frequencies \( f_x = \frac{x}{\lambda z} \) and \( f_y = \frac{y}{\lambda z} \) to assure the correct scaling in the observation plane. The diffraction pattern in this region is called as Fraunhofer diffraction and the formula according to Equation (2.76) is called as Fraunhofer diffraction integral.

Equation (2.76) also reveals that it is just the Fourier transform (apart from the factors outside the integral), which can be computed easily and with greater speeds using well developed Fast Fourier Transform algorithms. Even the factors outside the integral can be omitted if we are interested only in the intensity distribution in the screen. This formula is well suited for the simulation of diffracted wave from a 2D object in the far field. But from condition given in Equation (2.75) it is clear that Fraunhofer diffraction will be formed only at larger distances from the aperture. So it will be inconvenient to perform the reconstruction experiment in the lab. A convex lens is usually used in the optical setup to overcome this potential problem.

2.3.2.8 Summary

The various theories that explain wave propagation and the corresponding formulas were discussed starting from the very basic electromagnetic theory. Accordingly electromagnetic field at any point in space is defined by the Maxwell’s equations as seen from Equations (2.22 - 2.25). Light is also electromagnetic in nature and hence the disturbance (complex amplitude) of a propagating light field can be
Figure 2.16: Diffraction theories

determined any where in space and at any time by the Maxwell’s equations. Numerically evaluating the complex amplitude using Maxwell’s equation is very difficult due to its vectorial nature and also the discretization errors that creep in. But when the medium of propagation becomes, linear, isotropic, homogeneous and nondispersive the Maxwell’s equations can be greatly simplified to get rid of the vectorial nature. This gives rise to the scalar wave equation as seen from Equation (2.31). Experiments in digital holography satisfy these requirements and hence scalar diffraction theories are the mostly used ones to simulate wave propagation. The scalar wave equation which is a differential equation can also be expressed as an integral equation based on a particular Green’s function. It also turns out that the Huygens-Fresnel postulate on diffraction can be mathematically expressed using this integral equation as seen from Equation (2.41). These non-vectorial integral equations constitute the Scalar diffraction theory. These equations can be further approximated based on the propagation distance which results in the Fresnel and Fraunhofer diffraction theories. These approximated theories reduce the computation complexity greatly and also fit well into most practical situations. Hence the Fresnel and Fraunhofer diffraction theories are the mostly used
ones in digital holography. But they can be used with FFT only when the object surface and hologram surface are parallel to each other and perpendicular to the optical axis. There can be situations where the hologram surface is tilted or curved to the object surface. In these situations, only direct integration is possible which is very time consuming. This work is an attempt to device a wave propagation formula which can use FFT even when the hologram surface is a curved (cylindrical) one. All the scalar diffraction theories explained in this chapter define wave propagation in real space except the angular spectrum formula Equation (2.62) which defines propagation in spectral domain. This formula can also be evaluated by FFT, only if the object and hologram surface are plane and parallel to each other i.e., are shift-invariant. The work reported in this thesis is also to device a spectral propagation formula where the hologram surface is cylindrical but still could use FFT. All the diffraction theories explained in this chapter can be summarized graphically as show in Figure 2.16.

### 2.3.3 DH Techniques and Algorithms

This section presents an overview of some of the available techniques and algorithms which facilitate, or offer solutions to the problem of wave propagation based on wave optics.

(Sherman 1967) gave an elegant proof of the equivalence of the Rayleigh diffraction integral and the exact scalar solution based on the plane wave superposition of waves propagating in the direction. This is an important contribution because the fast direct calculation of the Rayleigh integral is difficult but efficient procedures based on FFT can be developed by using the plane wave decomposition. Grella (1982) examined diffraction and free-space propagation of an optical scalar field by using the Fresnel approximation. Grella states that Fresnel approximation can be represented as a superposition of plane waves besides the original approach based on the series expansion of
the spherical wavelet exponent. He has provided an unified approach for Fresnel approximation.

Ganci (1981) gave a simplified representation of diffraction of a plane wave through a tilted slit by using Fraunhofer approximation. Rabal et al. (1985) generalized the method proposed by examining the amplitude of diffraction patterns due to a tilted aperture. They use the Fourier transform to calculate the intensity pattern from a tilted plane onto another plane perpendicular to the initial optical axis. Leseberg and Frere (1988) were interested in the computation of the diffraction pattern between tilted planes, and they generalized the approach proposed by Ganci (1981), Rabal et al. (1985). Frere and Leseberg (1989) used their proposed method to obtain computer-generated holograms of larger objects.

Tommasi and Bianco (1992) investigated the relation of the angular spectra between rotated planes. They also proposed a solution to the diffraction problem between tilted and shifted planes (Tommasi and Bianco 1993), which employs the FFT. The mathematical and physical basis of the method together with several simulation results and their physical meaning were given by Esmer (2004), Esmer and Onural (2004). Another method is proposed by Matsuhima et al. (2003) to compute diffraction pattern on tilted planes, but the presented method is essentially based on the method given in Tommasi and Bianco (1992, 1993), Delen and Hooker (1998). The significance of the procedure proposed by (Matsuhima et al. 2003) compared several interpolation algorithms together with their effects on the computed diffraction patterns.

Mas et al. (1999) compared fast Fourier transform methods and fractional Fourier transform methods for calculation of diffraction patterns. They state that discrete Fourier transform methods are valid only for a specific range of distances. On the other hand, fractional Fourier transform methods provide an accurate and easy implementation and give much better results in reproducing the amplitude patterns. In another paper, Mas et al. (2003)
investigates the diffraction pattern calculation under convergent illumination. They conclude that fractional Fourier transform gives a unified solution of calculation of diffraction field in all ranges of distances. Mendlovic et al. (1997) undertook similar investigations, comparing different numerical approaches and identifying the more advantageous one as a function of the distance of propagation. Hennelley and Sheridan (2005) provided a very general and uniform framework to compare most such approaches. Ozaktas et al. (2006) proposed an algorithm based on the fractional Fourier transform that solves most of the problems associated with earlier algorithms applicable to the Fresnel regime, and is also applicable to a broader family of integrals.

Sypek (1995) compares the two computational approaches associated with the Fresnel diffraction. One of them is the direct convolution, whereas the other one is based on a single Fourier transform with pre and post multiplications with chirp functions. Two modifications on the convolution based approach are proposed. The first one uses length $2N$ vectors instead of length $N$ vectors. The second one divides the propagation distance into several segments. This reduces aliasing errors.

Veerman et al. (2005) proposed a method that integrates the Rayleigh-Sommerfeld diffraction integral numerically. They exploit the slow varying nature of the envelope of the highly oscillatory quadratic phase function in diffraction patterns. However, the method is not as fast as methods based on the plane wave decomposition or Fresnel approximation. An FFT-based computation of the Rayleigh-Sommerfeld diffraction is also presented by (Shen and Wang 2006).

Optical diffraction can also be represented by using wavelet transformation (Onural 1993). Sheng et al. (1998) have shown that optical wavelets proposed by Onural (1993), Onural and Kocatepe (1995) are the Huygens spherical wavelets under Fresnel approximation.

Some basis functions have been designed to deal especially with
holographic signals. The wavelet-like fresnelets, have been constructed for Fresnel hologram processing (Liebling et al. 2003, Liebling and Unser 2004). A Fresnel transform is applied to a standard B-spline biorthogonal wavelet basis to simulate the propagation in the hologram formation process. The obtained basis functions are well localized in the sense of the uncertainty principle for the Fresnel transform and have excellent approximation characteristics. The fresnelet transform allows for the reconstruction of complex scalar waves at several user-defined, wavelength-independent resolutions.

Cywiak and Santoyo (2001) used the linearity of the Fresnel transform for fast computation. They first decomposed the input function into Gaussian functions. Since it is easy to compute the Fresnel transform of a single Gaussian function, a final superposition of the individual results gives the desired Fresnel transform. The Gaussian (more generally, the Hermite polynomials times the Gaussian) functions are the eigenfunctions of the Fourier, and therefore the fractional Fourier transforms. Thus we can associate the easy computation of their Fresnel transform to this property.

Onural and Scott (1987) mainly concentrated on eliminating the twin-image in in-line holograms. Since the twin image and the desired image overlap with each other in in-line holography, twin-image elimination is more important compared to the off-axis case. Moreover Onural (1985) presented and compared the two digital Fresnel diffraction computation algorithms: one based on direct convolution with a chirp, and the other one based on a single Fourier transform with pre and post multiplications by a chirp. Allebach et al. (1976) discussed the application of DFrT for hologram computation and the associated aliasing effects due to sampling.

Esmer et al. (2007) presented algorithms based on pseudo matrix inversion, projections onto convex sets and conjugate gradient methods, together with performance comparisons for computing the diffraction pattern
over a reference plane due to distributed discrete data in 3-D space.

Mapping from a 3-D problem into its 2-D counterpart, and other issues associated with resolution, accuracy and degrees of freedom in optics were also discussed (Shamir 1999). Some algorithms based on optimization techniques are proposed by Piestun and Shamir (2002). Wave field synthesis methods found applications in synthesis of some important beams and unconventional waves (Durnin 1987, Spektor et al. 1996).

Efficient and effective computation of holograms using modern computer graphics procedures and hardware are also reported (Ahrenberg et al. 2006). Furthermore, 3-D objects are extracted from holograms digitally and displayed on conventional 2-D displays using computer graphics methods (Ziegler et al. 2007).

Compression of holographic signals require special techniques for improved compression performance due to the specific form and nature of such signals (Naughton et al. 2002). It is also shown that the 3-D objects can be reconstructed only from the phase information of the optical field calculated from the phase-shifting digital holograms (Matoba et al. 2002). Compression of holographic signals by constructing the hologram by pre-computed, indexed, stored small-size fringe patterns is demonstrated to yield real-time operation for horizontal parallax only (HPO) holograms (Lucente 1997).

2.3.4 Methods in Holography

Instead of optical recording, the hologram associated with the wavefront representing the object is generated by employing different computational techniques and numerical approaches by mathematically simulating the optical wave propagation. An ideal CGH should achieve complex light modulation at a high diffraction efficiency and precise reconstruction of the target image. The CGH’s outperform conventional refractive and diffractive components as a consequence of their ability to
create any desired wavefront and thus to modify the input wavefront with much better flexibility (Flury et al. 2005). For this reason CGHs find a wide range of application as display elements, optical interconnects, aberration compensators in optical testing, spatial filters for optical signal processing and computing, beam manipulators and array generators etc. CGH’s can be considered as thin optical elements with a complex amplitude transmittance. However, in many cases, they are phase only elements (Davis et al. 1999). There are different classification of CGHs depending on the complex amplitude representation on the recording media (binary, phase, amplitude and combined phase-amplitude media), and the encoding method (Yaroslavsky 2003). The algorithm to form a CGH is chosen according to the desired image characteristics and the associated computational complexity. Analytical approaches such as phase-detour method, kinoform method, double or multiple phase methods, explicit spatial carrier methods, 2-D simplex representation, representation by orthogonal and bi-orthogonal components, coding by symmetrization, etc., can be used for computing digital holograms (Yaroslavsky 2003). There are cell-oriented and point-oriented methods. In cell-oriented CGH’s the hologram plane is divided into small resolution elements. The number of resolution cells needed depends on the complexity of the wavefront that is to be produced. Iterative approaches such as iterative Fourier transform algorithm (Wyrowsky and Bryngdahl 1988), direct binary search (Seldowitz et al. 1987), simulated annealing (Kirk and Hall 1997) have been proposed and used. These methods are computationally demanding.

However, CGHs which are intended for dynamic displays need faster algorithms. It is difficult to realize SLM’s which can provide the desired complex phase (Arrizn et al. 2005). SLM’s with only binary modulation are particularly desirable for display of CGHs. Computer generated binary reflection holograms may be displayed using micro mirror devices (DMD) (Kreis et al. 2001). The SLM properties are crucial for the quality of
the optical reconstruction of digital holograms. A comparison of the optical reconstruction of phase and amplitude holograms by different modulators in terms of diffraction efficiency and recovery quality is presented by Kohler et al. (2006). A Fourier transform based algorithm for fast calculation of diffractive structures, which permits image reconstruction on cylindrically and spherically curved surfaces, is developed by Sando et al. (2005). Another popular approach is to calculate the CGH as a superposition of analytic distributions by decomposing the object surface into a certain number of discrete independent point sources, line segments or higher-order image elements. The modeled underlying physical phenomenon is the interference between the light waves coming from the analytically defined holoprimitives constructing the object and the reference wave to form the resulting complex amplitude distribution on the hologram plane (Leseberg 1992). Hardware (Ritter et al. 1999) and look-up table based computations were also proposed (Plesniak 2003). Representation of image elements at different locations by scaling and translation of similar elemental diffractive structures permits fast updating of the CGH by the so called incremental computing method (Matsuhima and Takai 2000). Real color fractional Fourier transform holography is proposed by Jin et al. (2006).

2.3.5 Sampling and Quantization

Sampling and quantization are inevitable when discretising any data. Sampling and quantization should be done properly to avoid any loss of interesting information and hence should be handled with a lot of care. Digital holography is no exception to this and discretization issues hold an important role in the process. Discretization issues like sampling and quantization are old topics, explored in detail with electrical signals and has well developed general theorems for loss free discretization. However diffraction related signals has additional special features and hence the general approaches are not efficient nor adequate enough. Hence discretization of diffraction related
signals should be done by exploiting their special characteristics. This is an interesting and fruitful area of research which has improved the speed and efficiency of diffraction calculations. This section will review through the special characteristics of diffraction signals and their sampling and quantization requirements.

2.3.5.1 Sampling

A signal can be space- or band-limited but never both. For optical signals, the so called Fresnel limited functions turned to be more convenient and efficient than the band-limited functions in terms of sampling and recover ability. Fresnel limited functions are defined to have finite extent in their Fresnel transform domain associated with the parameter $\alpha$. Such functions are not band-limited, however, they can be reconstructed from their samples taken at a rate. The proof of this result is given by Gori (1981). Another theorem proven by Gori (1981) indicates that the Fresnel transform of a space-limited function can be fully recovered from its Fresnel domain samples. The same result was also proven later independently by Onural (2000) who also stated the prefect reconstruction conditions for both band- and space-limited cases. In particular, it is shown that for Fresnel transform, full recovery of space-limited signals is possible even when sampled below Nyquist rate. It is also shown by Stern and Javidi (2004a) that it is possible to reconstruct objects from hologram samples obtained below the Nyquist rate. Real-time applications by considering finite number of samples and finite (nonimpulsive) area of the capturing charge coupled devices (CCD) array elements were also discussed. Furthermore, the effect of sampling in noisy conditions is also analyzed. Thus the possibility of full recovery from under sampled holographic signal is observed.

The effects of the shape of the sensing elements and the overall array size to the CCD captured optical data and subsequent digital reconstruction of off-axis holography are examined by Kreis (2002). A frequency domain
analysis of the overall transfer function is carried out for both the planar and the spherical reference beam cases.

In a work by Stern and Javidi (2004b), it is shown 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. Several nonuniform sampling schemes have been suggested based on the observation that the bandwidth of the object remains unchanged as a consequence of the all-pass nature of the linear system that represents the diffraction. Another approach observes that the information of interest in a hologram is carried in the complex envelope of the fringe pattern and not in the carrier (Khare and George 2003). Based on this, Khare and George (2003) have suggested sampling the recorded hologram about twice the Nyquist rate for the object (or baseband) signal. This may be regarded as a generalization in the shift-invariant space. In the work by Liebling et al. (2003), where the modulation is replaced by the Fresnel transform, can be noted as well. Wavelets have inspired several interesting approaches in the area of optical signal sampling and reconstruction. The diffraction integral is viewed as a continuous wavelet transform by Onural (1993). The light field at different distances is regarded as the result of an inner product of the light distribution at some initial plane with scaled and shifted chirp functions. In contrast to conventional wavelet analysis, these scaling functions however, are not limited in neither the spatial nor the frequency domain. The transform has been named scaling chirp transform and shown to be valid and reversible by Onural and Kocatepe (1995). A number of inversion formulas are provided with a discussion on their redundancy and ways to possibly exploit this redundancy. For fixed scale, the scaled and shifted chirp functions form a complete orthogonal set, while they form a redundant frame over different scales. This also suggests a way to sample the light field throughout the space by using scaled chirp expansions.

Some related wavelet-like functions, called chirplets have been
suggested by Mann and Haykin (1992), and used for instantaneous frequency measurements. A chirplet is a compact support signal with increasing (decreasing) frequency (Mann and Haykin 1992). It is band and time localized version of the scaling chirp function mentioned above. Chirplets are rather attractive for representation of holograms since they have minimal energy spread for the Fresnel transform in a similar sense as Gabor functions (Liebling et al. 2003). Methods for finding a sparse chirplet signal representation were suggested by Qian et al. (1998).

An interesting strategy to construct bases suitable for processing digital holograms is presented by Liebling et al. (2003). Based on the observation that digital holography tends to spread out sharp details such as object edges over the entire imaging plane, standard wavelets have been ruled out as directly applicable to holograms. Instead, a Fresnel transform is applied to a wavelet basis to simulate the propagation in the hologram formation process. In contrast to classical wavelets, where multi resolution spaces are generated through dilation of one single function, in the fresnelets case there is one generating function for each scale. B-spline biorthogonal wavelets have been used to construct the fresnelet dictionaries due to their excellent approximation characteristics and analytical expression in spatial domain. Subsequently, their Fresnel transform associated wavelets are derived explicitly (Liebling et al. 2003). Thus, this new diffracted basis can be used to analyze the light field distribution at some distance and once a decomposition is obtained, the field can be calculated immediately in the original (initial) plane.

Digital reconstructions of diffraction patterns or holograms require algorithmic digital implementations of the underlying continuous mathematical models which represent diffraction. Common implementations of the Fresnel case are either based on convolution, or on a single Fourier transformation (Yaroslavsky 2005). Inevitably, either the kernel which represent the wave-propagation (diffraction), or its analytically known Fourier
transform (the transfer function) should be discretized when the convolution is implemented digitally. This problem is in the focus of the paper (Onural 2004) where some well known properties of the continuous Fresnel kernel, together with rather overlooked ones are presented. Furthermore, efficient computation of the exact Fresnel transform of some periodic input (object) functions at some specific discrete distances is given, too. Another observation is the perfectly discrete and periodic nature of the continuous Fresnel transform of periodic and discrete input functions for certain distances.

2.3.5.2 Quantization

From a theoretical point of view, the diffraction is an operation which disperses the information content of simple object patterns over the entire space. Therefore, it is quite immune to noise or loss of information. Reconstructions from partial holograms could be pretty much satisfactory, with some bearable quality degradation. Therefore, it is expected that grossly digitized holograms would still yield reasonable reconstructions. Indeed, this fact was utilized for the computer-generation of holographic masks, going all the way to binary holograms. It might be interesting to look at oversampled, but coarse digitized cases.

Mills and Yamaguchi (2005) discuss the quantization effects in phase-shifting holography. It provides both numerical simulations and experimental quality assessment and concludes that, for both uniform (specular) and random (diffuse) objects a 4-bit quantization is sufficient to recognize the reconstructed objects and the difference between 6 and 8 bits is not perceivable. Above 4 bits, the effect of quantization on the reconstructed image quality seems to be independent of the object phase distribution. Neal and Gallagher (1978) have demonstrated that the quality of the reconstructed images from recorded holograms is more influenced by the phase information than the magnitude information. The paper assumes, with relevant arguments, that the magnitude has a Rayleigh distribution, whereas the phase is uniformly
distributed over the interval. Then, a solution for minimum-mean-squared-error quantizer in polar form is formulated and numerically solved for some quantization levels. The allocation of bits between phase and magnitude is discussed.

It is observed that even though the phase and magnitude are statistically independent, the optimum magnitude quantization scheme depends on the number of phase quantization levels. The effects of phase quantization in Fourier holography is discussed by Dalla and Lohmann (1972). It is concluded that phase quantization results in ghost images located at different depths. It is further concluded that these ghost images are less disturbing particularly for high-contrast images, due to their different depths. Nonuniform quantization through computing of complex numbers by employing nonuniform grid patterns over the complex plane is shown to be efficient for digital holograms with a reconstruction quality comparable to that obtained by quantization by the $k$-means algorithm (Shrott et al. 2006). Quantization issues associated with holographic signals are discussed by Naughton et al. (2002). It is shown that degradation in reconstructed image quality is minimal for 10 bits or more, and the distortion becomes severe below 5 bits. Numerical error plots together with reconstructed images are presented.

2.3.6 Signal Processing Tools for DH

2.3.6.1 Plane wave decomposition

As elementary as it is, plane wave decomposition remains a key tool for understanding optical diffraction. Plane wave decomposition is directly related to Fourier decomposition, with planewaves propagating in different directions corresponding to different spatial frequencies. Therefore, the Fourier transform has been the most natural tool for space-frequency analysis of optical signals. Algorithms for fast implementation of its discrete
version, the discrete Fourier transform (DFT), the so-called FFT algorithms are extensively studied and well established. The famous Cooley-Tukey algorithm is just one from this family. Among others are prime-factor (Good-Thomas) FFT algorithm, Bruun’s FFT algorithm, Rader’s FFT algorithm, and Bluestein’s FFT algorithm. A rather new approach to the efficient implementation of Fourier transforms is computing it via the Walsh-Hadamard transform (WHT). The approach is based on the Good’s theorem, which suggests factorizing a Kronecker product structured transform matrix into a product of several sparse matrices. Since the WHT matrix has exactly such a recursive Kronecker product structure the WHT coefficients can be computed very efficiently, and then converted into FT coefficients by a special conversion matrix.

A number of newer transforms have been found applicable or at least promising for analysis of signals modeling optical diffraction. They can be unified under the notion of atomic decompositions. More specifically, these are signal representations in terms of basis sets with particular features especially suited to an application, allowing the capture of the signal characteristics by only a few significant coordinates. A selection of references to the most important atomic decompositions are given below.

2.3.6.2 Wavelets

Wavelets have been perhaps the most inspirational constructions due to their ability to represent transient signals by offering a trade-off between space and frequency (scale) resolution. As basis functions, they separate the space of square-integrable functions into a set of nested subspaces. To improve the time-frequency (space-scale) resolution, wavelets have been extended also to overcomplete schemes such as wavelet packets and bases with improved directional and translational-invariant properties, such as Gabor wavelets and Dual tree-complex wavelets. Other bases, such as
ridgelets, curvelets, beamlets, brushlets have been designed for effective representation of ridges, curves, lines or oriented textures, respectively.

2.3.6.3 Chriplets

The chirplets, which are also already commented in the light of holographic signal sampling, can be useful in digital holography for space-frequency analysis since they are known to be good instantaneous frequency estimators.

2.3.6.4 Fractional Fourier Transforms

The fractional Fourier transform (FRT) is a generalization of the ordinary Fourier transform with a fractional order parameter such that the zeroth order transform is the identity operation, the first order transform is the ordinary Fourier transform, and the fractional transform interpolates between them in an index-additive manner. It has found a large number of applications in signal processing and optics. The relationship of the FRT to optical propagation and diffraction rests on the result relating free-space propagation in the Fresnel approximation (namely the Fresnel integral or the Fresnel transform) to the FRT. Extensions of this result relate arbitrary linear canonical transforms to the fractional Fourier transform. Optical systems consisting of arbitrary concatenations of lenses and section of free space can be modeled as linear canonical transforms, and thus propagation through such systems, including 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 multilens system. While these results are directly relevant to holography, relatively few works have explicitly applied the FRT to holographic problems. Sampling and periodicity issues related to the fractional Fourier transformation have been discussed by Xia (1996)
2.3.6.5 Linear Canonical transform

These take the form of quadratic-phase integrals which are equivalent or related to linear canonical transforms. During numerical evaluation of these integrals, naive application of the Nyquist-Shannon approach may require very large sampling rates due to the highly oscillatory nature of the kernels. It has been shown that by careful consideration of sampling issues, the number of samples need not be allowed to be larger than the space-band width product of the signals. A fast algorithm for computing the samples of the continuous FRT of a function from the samples of that function is presented Ozaktas et al. (1996). This algorithm has been extended, with the same properties, to arbitrary LCTs. This approach employs the smallest possible number of samples implied by the space-bandwidth product of the output signal. Recalling that linear canonical transforms can model systems consisting of arbitrary concatenations of lenses and sections of free space, this algorithm can be used to compute the input-output relation for such systems with an efficiency and accuracy comparable to the use of the FFT in computing the Fourier transform. Comparisons of different approaches to calculating Fresnel integrals was given by Mendlovic et al. (1997) which shed light onto the limitations of certain earlier methods. Some of these have received greater attention in optics, such as the windowed/short time Fourier transform, which is closely related to Gabor expansions and the Wigner distribution and ambiguity function. The relationship between the Wigner distribution and linear canonical transforms and fractional Fourier transforms is of key importance. Fractional Fourier transformation corresponds to rotation of the Wigner distribution (Mustard 1996). The Wigner distribution and linear canonical transforms were established as a standard tool in optics primarily by Bastiaans (1978).

Recently, the diffraction problem is revisited and formulated using the projection-slice theorem as a tool using impulse functions defined over
Linear canonical transforms (LCT) are a class of integral transforms which include the fractional Fourier and Fresnel transforms and other important operations as special cases. They are also known as quadratic-phase systems or integrals, generalized Huygens integrals, generalized Fresnel transforms, ABCD integral transforms, or similar names. Fast numerical algorithms for LCTs exist and is described by Hennelley and Sheridan (2005).

2.3.7 3D Display Devices

2.3.7.1 Holographic display

If a hologram is illuminated with the reconstructing light which is same as the recorded original, any observer interacting with the reconstructed light will see the same scene as original. Therefore in principle, holography creates true 3-D images, with all correct color, depth, shape information and parallax relation. This broad sense of definition involves all classical holographic techniques where coherent light is used to record the complex valued wavefront via interference and other true 3-D imaging techniques like ideal integral imaging (Lippmann 1908). However in digital holography, holograms are stored as digital images in the computer which can be streamed into any display devices one after the other. Hence a dynamic display device is required to realize the video capabilities of digital holography. Candidate technologies for holographic display units include dynamically writable/erasable chemical films (Beev et al. 2006), or on electronically controllable arrays of pixels that can alter the phase and amplitude of light passing through (or reflected by) them, called spatial light modulators (SLMs) (Amako et al. 1993). Specific forms, like deflectable mirror array devices (DMADs) are also among potential technologies that can be adapted for 3-D display, which can also be considered as special forms of
SLMs (Kreis et al. 2001). Currently, dynamic chemical film technology is not mature enough for acceptable performance. Unfortunately, the size, quality and geometries of SLMs are currently not sufficient for acceptable quality 3-D displays, either. However, it is expected that both technologies will develop in time to yield the desired quality. There are also other techniques which are based on interaction of light with acoustic signals. Some experimental holographic 3DTV systems usually choose to sacrifice from the ultimate true 3-D display quality, for example by eliminating vertical parallax, and thus achieve higher resolution and fidelity in other features, or reduction in computational complexity. Ability to steer light from each point of a display device to arbitrary directions provides solutions to the dynamic display problem. Speckle noise in case of coherent illumination is another disadvantage, and there are proposed techniques to cope with this problem (Iwai and Asakura 1996).

2.3.7.2 Integral Imaging

Integral photography has been revitalized after the progress in active pickup devices and microlens manufacturing processes (Okano et al. 1997). It relies on a capture device based on a microlens array to encode a true 3-D optical model of the object as a planar intensity distribution which can then be reconstructed by reversing the direction of incident optical rays. Analysis of integral imaging devices can be carried out both by ray and diffractive optics. Improvements in related computational procedures and the incorporation of novel techniques like moving lens arrays solved many problems in integral imaging. Solutions for viewing-zone enhancement have been tested using dynamic barrier arrays, microconvex-mirror arrays or lens switching techniques. Very large-scale and projection based integral imaging systems with increased resolution and viewing angle are reported. Techniques have been proposed to improve the depth of viewing field based on amplitude modulated microlenses, a change of the optical path length, synthesis of real
and virtual image fields, or on the use of microlenses with nonuniform focal lengths and aperture sizes. The issues of scene occlusion as well as removal of the multifacet structure or suppression of color moire in the reconstructed images have been successfully resolved. The maximum information capacity of integral imaging can be increased by using Karhunen-Loeve transform for image compression. Holograms can be computed from captured images during integral photography.

2.3.7.3 Stereoscopic Displays

Past and present implementations of most 3DTV systems rely on stereoscopy, or multiview video. In these approaches, no attempt is made to duplicate the original optical field. Instead, two or more 2-D images are captured at slightly different viewing angles. The human visual system interprets the received images. 3-D perception relies on the processing of several depth cues. Older type systems require special goggles to direct different images to each eye; however, newer systems utilize autostereoscopic systems to guide different 2-D views to different angles (Sexton and Surman 1999). Systems based on stereoscopic principles usually create a feeling like motion sickness especially when some associated alignments are not perfect. Signal processing issues related with such display schemes are discussed in the review paper by Isgro et al. (2004). While the stereoscopy-based techniques are the most popular 3-D imaging techniques to date, holography-based techniques will most likely be the ultimate choice for digital holographic in the future.

2.4 SUMMARY

It is very clear from the above discussions that, very little amount of work has been done with regard to wave propagation from cylindrical surfaces. But the basic theories governing wave propagation from cylindrical surfaces resembles its planar counter part. Hence it is possible that cylindrical
holographic techniques should fit well into the architecture of the existing techniques discussed above. With this motivation, it was planned to do some research on the fast computational method for wave propagation from cylindrical surface and apply it to digital cylindrical holography. Accordingly a plane wave decomposition method for cylindrical surface was proposed and demonstrated with results. However, from the above discussion it can also be understood that, realizing cylindrical digital holography has its own difficulties too. The major one being the unavailability of a suitable dynamic display device in cylindrical shape. But fold-able polymer display devices are out in the market and hence we are not far from the days to have cylindrical display devices that could display fringe pattern in high resolution. The other major problem is the optical arrangement needed to illuminate the whole hologram which is very complex. This greatly reduces the portability of the hologram.