CHAPTER 1

INTRODUCTION

1.1. What Is Holography?

From physical point of view, holography is a special recording and retrieval process. What is recorded in a hologram is a *wave front*, and this is the unique feature which distinguishes holography from other recording processes. The physical nature of the wave could be anything, but the mostly used ones are electromagnetic, acoustical and matter waves. Among these the visible light is far more dominant. A wave front is characterized by its magnitude and phase. Therefore, in order to reproduce it both the magnitude and the phase must be recorded. Holography achieves this to a remarkable extent: a wave front recorded in a hologram can be fully reconstructed in the ideal case, together with some extra undesirable components.

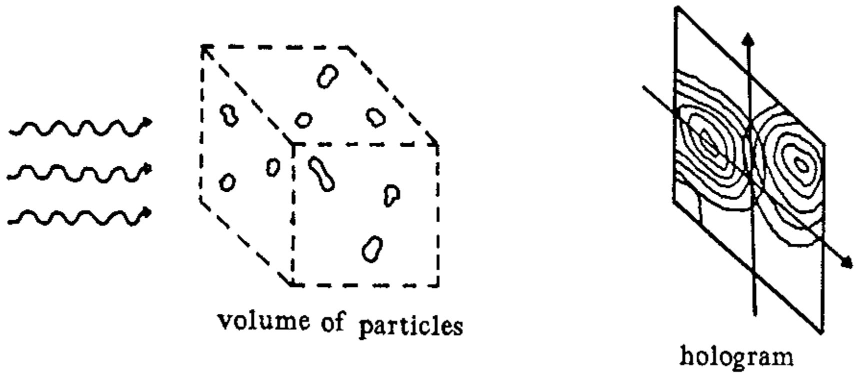
Obviously, the ability to reconstruct an optical wavefront has very significant consequences. Every optical interaction occurs through light waves. For instance, we see every feature of our environment including its three-dimensional nature, because of the biological and psychological effects induced by the light coming into our eyes. If the original wave front is reproduced somehow, then the optical interactions caused by either the reproduced light or the original one will be the same. In other words, we will be able to see the same three-dimensional space and objects as the original ones, even though the originals do not exist. The reproduced light can be used as if it is the original, for any purpose, such as taking its photograph.

From the signal processing point of view, holography is a coding-decoding process. The optical information related to three-dimensional objects and space is coded in a hologram. Then, the hologram can be decoded to get the original signal back.

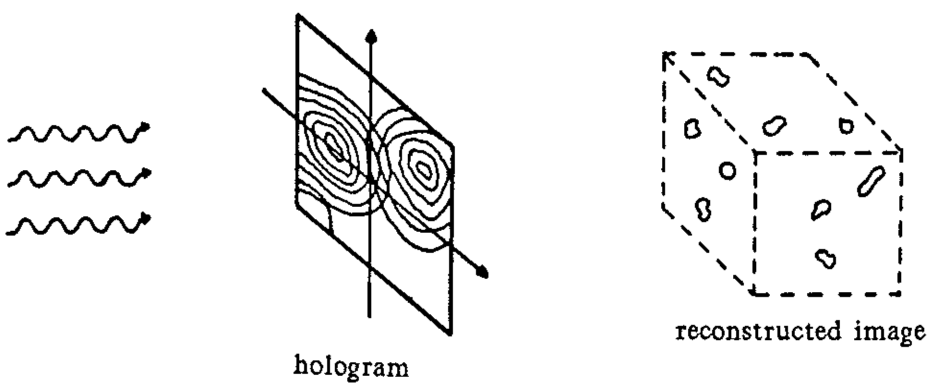
1.2. Motivation, Purpose and Approach

One of the important applications of holography is the observation of dynamic aerosols [1-14]. A volume of aerosol typically contains hundreds of particles whose sizes are on the order of microns. Since the volume under investigation is on the order of centimeters, there is no optical system which can record the desired information because of intrinsic focusing difficulties. Optical microscopes have a small depth of focus, and therefore the particles of this size, distributed to such a large volume, can not be focused all at once. It has been known for about two decades that the solution is holography. A typical optical in-line hologram recording and reconstruction set-up is shown in Fig.1. The recording phase of holography does not require any focusing, since it records the wave front itself. The hologram stores the wave front, which can be reconstructed afterwards, coming from the particle volume. The reconstructed wave front can then be focused to image individual particles. Thus, in a sense, the dynamic particles in a volume can be frozen by recording the hologram, and then this three-dimensional space can be analyzed at any pace using the reconstructed wave.

An aerosol analyzer is intended to be used in many diverse environments. Therefore, the operation of such a device has to be relatively insensitive to working conditions. Since the hologram is the recording of the intensity of an interference pattern, its quality depends on the stability of the pattern during the exposure time. Off-axis holography has much better imaging properties compared to in-line holography, but its stability requirement is very high. Since the recording of dynamic patterns is intended, less sensitive in-line holography has been used to analyze particle volumes. Its practical use has



a)



b)

Figure 1. Typical optical in-line holography set-up for particle analysis.

been justified for many years.

Two major drawbacks of in-line holography are the existence of a twin-image component in the reconstructions and the tedious task of focusing individual particles during information retrieval. Because of the first problem, a large class of in-line holograms (Fresnel holograms) is practically abandoned. Far-field in-line holography is used since it has a less severe twin-image effect. To overcome the second problem, computer controlled servo systems are used to move optical components, to achieve automated focusing.

From a practical point of view, this dissertation proposes a solution to the twin-image problem, and simplifies the solution of full automation. However, the approach used in this dissertation is fundamentally different from previous attempts. In this work, the reconstruction phase of in-line holography is treated completely as a signal processing problem. The physical reconstruction phase is *totally* eliminated. Instead, the reconstruction is performed completely by computation, using digital signal processing techniques, directly from the digitized hologram. Therefore, the overall dissertation can be considered as a special purpose image processing topic, where the images happen to be in-line holograms. Then the purpose of processing is to decode the object information from the hologram.

Since the outcome of this dissertation is intended for immediate practical use, the work is oriented towards concrete results. What is achieved in this work can be summarized by Fig.2, where the picture on the right (particles) is obtained from the picture on the left (hologram), totally by a computer, and the computation requires only one parameter as input to accomplish the task. The theoretical considerations of the work are given together with practical applications, throughout the dissertation.

1.3. Historical Developments and Present Scientific and Technological Ground

In this section some important historical contributions to the field of this dissertation are briefly summarized. Since the dissertation is exclusively related to holograms, a historical development of related area of optics is emphasized. Because of the dominant image processing nature of the dissertation, some important contributions to the general

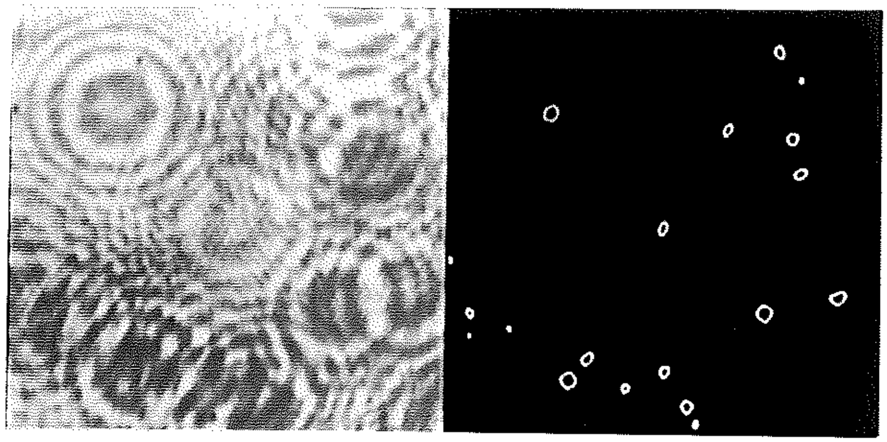


Figure 2. An in-line hologram (left) and its decoding (right) by a computer.

area of digital signal processing are also mentioned. In fact, it will be soon presented that one of the two most important contributions to holography comes directly from communication theory, which is very closely tied to signal processing.

It is known that Leonardo da Vinci was aware of diffraction phenomena, but it was first described accurately by Grimaldi in 1665. The wave theory of light was first advocated by R. Hooke, and greatly improved by C. Huygens in 1690. He stated the principle, which is named after him, that every point of the disturbance (wave) causes a new disturbance which propagates spherically and these secondary disturbances combine to form another wave front. Then it was A. J. Fresnel who formulated the Huygen's principle, by associating it with Young's principle of interference, in 1817. Fresnel's model completely rejected corpuscular theory of light, and succesfully explained diffraction. Huygens-

Fresnel formulation for diffraction is the basis of the formulation given in this dissertation for holography. Later, a more rigorous mathematical theory is presented by G. Kirchhoff which is based only on the scalar wave assumption of light. It was also shown that in many cases the formulation can be reduced to a more simpler approximate form, which is equivalent to Fresnel approach. J. C. Maxwell's succesful theoretical work proved the existence of electromagnetic waves, and it was understood that light is an electromagnetic wave. More rigorous treatments of diffraction, therefore, require the solution of vector wave equations with boundary values. However, it is known that Kirchhoff's formulation is enough for almost all optical problems.

Holography was invented by Hungarian-born British scientist Dennis Gabor, and introduced as "a new two-step method for optical imagery", in 1948 [15-17]. The basic idea was the two step diffraction process. The magnitude and the phase of a diffracted wave is stored in a photographic plate. The storage of phase is accomplished by the introduction of a reference wave, which formed an interference pattern with the diffracted wave. Since the interference pattern is related to both the phase and the magnitude of the diffracted wave, the information storage part is solved. The other nice property of the method is that when this interference pattern is illuminated, it diffracts the light in such a way that one component of it exactly duplicates the original diffracted wave, with its phase and magnitude. Thus, for the first time in history, a practical method of storing and retreiving phase information of a wave was formulated. The storage of only intensity has been known for about a century (photography). Gabor's invention can be considered as the most important contribution to holography. But this important work did not get the attention it deserved for many years. The main reason for this ignorance is technological, since it was impossible to make decent holograms until the laser was invented and brought to practical use in early 60's. Another important contribution to holography is the work of Leith and Upatnieks (1961) [18]. They used the systems model as used in this dissertation to explain the holographic process. Then they added the modulation to the formulation and thus invented off-axis (split beam) holography. As stated before, split beam holography is able to separate the twin-image from the desired component, thus its imaging quality is much better. It is this form of holography which attracted a lot of attention in many fields, including arts. The importance of Leith and Upatniek's work lies in their fundamental approach to the problem of holography, by formulating it as a coding and decoding process. Fortunately, the underlying physical principles gave rise to a coding which has been well known in the field of communication and signal processing. They realized the drawback of the in-line holography in a straight forward manner from the model, and gave the proper solution which has been known in the area of signal processing for many decades: modulation. The modulation process shifted the frequency band of the signal, thus demanded higher film resolution and fringe stability. As mentioned in the previous section, off-axis holography is not preferred for particle measurement applications because its high stability requirement. B. J. Thompson used holography for the first time as a tool for another scientific purpose in 1963 [1-8]. He used far-field in-line holography to analyze collections of small dynamic particles, which is also one of the motivations for this dissertation.

Signal processing existed at all times, since every form of information transmission requires one form or another of it. This includes human communication through languages, which is one of the most advanced coding-decoding patterns. Signal processing can be defined as the science of manipulating signals. Often, it is hard to separate signal processing from communication theory. The technology of how the signal processing algorithms are implemented has changed as new technologies came to use. Major developments occured around late 19'th - early 20'th century when electronics as a technology made it possible to implement many signal processing methods easily. As a consequence of this, analog signal processing theory has developed very well. The arrival of computers started another era of signal processing. Many signal processing algorithms were implemented with this new technology of computation. (Although the computers are themselves electronic devices, the concept of computation has nothing to do with how the numbers are stored, thus is totally different from analog electronic processing of signals.) During the past two decades many digital signal processing techniques and algorithms have been developed [38-40]. Most of the effort, however, has gone to development of digital counterparts of already known analog techniques. It has been recently realized that it is possible to create many new techniques using digital processing which do not have analog counterparts. Many fascinating results in this field could be expected in the near future. In fact, many digital techniques are much simpler conceptually then analog methods. If there was not the motivation induced by the early electronic technology, it is quite probable that many digital techniques would have preceeded analog ones. A recent article shows that the discrete Fourier transform was known before Fourier's original work on the continuous case [55]. The fast Fourier transform can be considered as the most influential algorithm in digital signal processing, which is also used extensively in this dissertation. First formulation of discrete Fourier transform was given by A. C. Clairaut in 1754, but only for cosine series, and by L. Lagrange in 1762 for sine series. J. B. J. Fourier's formulation related to continuous case, which is definitely the most used technique in signal processing and analysis, came in 1807. Thus, the continuous Fourier transform was formulated *after* the discrete Fourier transform. The fast Fourier transform algorithm, which is an efficient method to compute discrete Fourier transform, was a very important step in digital signal processing. It was published in 1965 by J. W. Cooley and J. W. Tukey. But it is recently discovered that the fast Fourier algorithm of Cooley and Tukey was known in 1805, and developed by C. F. Gauss [55].

Representation of an arbitrary function by infinite summation of trigonometric functions has been very important since Fourier started the idea in 1807. As the mathematics developed, the properties of trigonometric series became better understood. In addition to trigonometric series expansions, other classes of functions also were investigated and used in many areas. As a consequence, many different ways to represent a signal are known, and available to be used in many practical problems. Fourier's harmonic functions, however, still dominate the field.

One final word is to emphasize the role of computer technology to the work done in this dissertation. An image is represented in the computer as an array of numbers. A typical *image* used in this dissertation (see any one of the pictures in the text, for example) consists of 262144 numbers. In order to process an image, arithmetic and logical operations are performed. The number of those operations is much higher than the array size given above. It has been only possible in the last decade to handle that many numbers and operations in a meaningful way, thanks to the advanced electronics which sped up computers and to the software which made computers run efficiently. Otherwise, this work would not be possible.

1.4. Structure

In order to achieve successful decoding of any coded information, a complete knowledge of the code is necessary. In addition to this, it is desirable to have simple and fast decoding schemes to reduce the effort. Good mathematical models for in-line holography (both for recording and reconstruction) have already been given, therefore, the relation of the hologram to the object distribution is already well known [1]. The complete model is reformulated in Chapter 2 to emphasize the underlying system science concepts. The model given in Chapter 2 is complete and makes the holography much more easily understandable by those whose background is in signal processing. The simple form of the model readily implies many efficient discrete techniques for the digital simulation of the physical process. In fact, the off-axis holography was invented by the help of this kind of formulation as mentioned in previous section. Digital extensions to the formulation are also given in Chapter 2. Digital reconstructions of optical holograms are significantly superior in quality compared to optical reconstructions. Many simulation results are given in Chapter 2, which are implemented based on the system model.

Once the system model for in-line hologram recording is given, it is straightforward to see that the reconstruction is a typical image processing problem. Conceptually, the decoding is the inverse operation of coding. Therefore, the complete reconstruction of the objects from their holograms is possible by designing an inverse operation to the recording. But it is also well known that not every process is invertible. In Chapter 3, an approximate inverse operator is given for a certain class of input objects. The operator significantly suppresses the twin-image, and can be easily implemented digitally in a computer. Details of implementation and many results are also presented. If it is known *a priori* that the objects have some properties, such as having well defined sharp edges, then the non-linear post processing techniques given in Chapter 4 can complete perfect decoding.

Thus, as a result of the combined digital procedure given in Chapters 3 and 4, *both* Fresnel and far-field holograms can be reconstructed succesfully, without worrying about the twin-image. As a product of this work, the range of possible applications of in-line holography to particle field measurements is significantly increased with the inclusion of Fresnel holograms.

Since the overall reconstruction is done computationally after digitizing the holograms, many problems of automation can be easily solved. There are no more moving mechanical components for automated focusing. Instead, the focusing is a simple change of one parameter during computations. This parameter can be changed algorithmically to achieve good focusing of a specific particle, or more than that, the overall volume can be scanned for particles located at different planes. Furthermore, any subsequent image analysis can be automatically performed by the same computer to achieve full automated measurements such as size distributions, classification, location, etc. Note that nothing requires any operator intervention after the digitized hologram is fed into the computer. This dissertation is not concerned with either automated updating of the focusing parameter or final artificial intelligance type image analysis. But the image processing, which is necessary prior to any image analysis, is completed for a class of particles.

1.5. General

Astronomy had been one of the major source of philosophical thinking in ancient civilizations. Thousands of years ago Babylonians, Egyptians and Greeks made accurate observations of orbital paths of many stars and planets. Attempts to find explanations to these observations could be considered as first applications of mathematics to natural science. Ancient Greeks had developed a quite sophisticated axiomatic geometry. They used their geometrical tools to create theories about the motion of stars. For many centuries they explained the motion of the sun, moon and the planets (five of which were known at that time) by assuming that they were fixed on rotating perfect spheres around the earth. However, to match the observations to the theory, they did not use a single sphere for a single planet. Instead, they used a system of spheres consisting of different sizes and rotating at different speeds; and the pole of one sphere was fixed to the surface of another sphere. According to them, the spheres were perfect geometrical structures, and they never imagined any other geometrical shape to explain the motion of planets. Even more, some of the well known philosophers of that time (for instance Aristotle) considered the spheres as physical reality beyond just mathematical convenience. A few centuries later an Alexandrian astronomer, Ptolemy (c. A.D. 140) replaced the spheres with circles. His system used perfect circles. It was considered that the circles rotated with different speeds, and the trajectories were explained by superposing the circles, such that the center of one circle was on another circle. He thought that the earth was the center of universe, and used about 80 circles to explain the orbits of known planets of his time. Europeans believed Aristotle's universe until about 15'th century. Although it was considered by some philosophers during the time of Aristotle, the idea that earth rotates around the sun was either rejected or ignored for about two thousand years until the time of Copernicus. Copernicus (a Pole) stated around 1512 that it is the earth which rotates around the sun. But he still believed that a perfect circle was the heart of the system. He used the same idea of superposition of circles as Ptolemy, but by assuming that the sun is in the center of universe, he reduced the number of circles from 80 to 34. The Copernican theory had been assumed as physical reality for some time and by well known scientists as Galileo. Finally, it was Kepler (a German) who came to the conclusion that the orbits are ellipses. The need for superposition of many circles suddenly disappeared, and the system of planets was very clearly explained just by replacing the circles with ellipses.

For a moment the historical note above may seem totally irrelevant to the topic of this dissertation. However, the idea of superposing circles to explain a more complex structure is one of the extensively used techniques of modern signal processing. Namely, the Fourier series expansion and its conceptual extension, Fourier transform, is nothing but decomposing a signal into harmonic functions which are inherently related to the circle. Fourier's work is dated 1807, as mentioned before. The Fourier transform (or similarly Laplace transform) is very useful especially in the analysis and design of time-invariant linear systems, since it converts a differential equation with constant coefficients into an algebraic one. In a short time the idea of expanding an arbitrary function into infinite series using basis functions of the form $e^{-jw_i x}$ was extended to other basis functions as well. However, the harmonic expansion is still the most widely used expansion. As mentioned above, the extensive use of harmonic Fourier series expansion (the terminology of Fourier series expansion is sometimes used for any basis function to honor Fourier) in many fields is based on a convenient property. In fact, the harmonic Fourier series expansion and Fourier transform is extensively used in this dissertation. However, it is useful to ask ourselves a question: Is it not possible that the reason for application of Fourier series (or transform) to some problems might be more habitual than its usefulness? It is the author's intuition that the answer is *yes*. There are many problems, physical or purely conceptual, where a series expansion using other basis functions than harmonic functions may significantly reduce the complexity, just like replacing circles by ellipses! The reduction of complexity may not be essential for actual operation of a system. However it will increase the comprehensibility, and thus may trigger many new ideas which could stay hidden otherwise. Furthermore, today, many systems are designed and constructed within the computers. Even in this case, it is still very useful to design shift-invariant linear systems because of the convenience of mathematical tools available. And one of the most important mathematical tool which almost dictates the superiority of such a design is the harmonic Fourier transform, and its nice property given above. However, the computational realization of systems is a very flexible phenomenon where there are no physical constraints. So, why bias ourselves in favor of one of many possible type of basis functions?

Such an approach is taken in the series expansion of the approximate inverse filter of Chapter 3. The expansion uses basis functions of the form $\cos\left[-j\alpha_k x^2\right]$, which are related to harmonic functions, but still significantly different. These basis functions are not imposed on the system, but naturally arise from the model of the physical system as explained in Chapter 3. The harmonic Fourier series, however, is still used to prove some properties of it, simply because of the availability of more literature on that subject.

Another extension of the disscussion above can be done in the field of filtering theory. Most of the filters today are characterized by their "frequency domain" nature. For instance, a low-pass filter means a filter which passes lower harmonic contents of a signal while blocking the higher ones. However, there may be many other ways of characterizing a filter, based on different signal representations. For instance, if quadratic phase functions as given above are used for signal representation, then it will be natural to talk about α domain characteristics and behavior of filters. The author strongly believes that such an approach would conceptually simplify the twin-image elimination problem to a great extent, but because of technological limitations will not be practical at this time. Because of such reasons this approach is not covered in this work.