An example of double exposure holographic interferometry. A hologram of the c-clamp is exposed. The light source is then shuttered and the clamp tightened. A hologram of the strained c-clamp is exposed onto the same recording material. After processing, illumination of the doubly exposed hologram results in the reconstruction of two wavefronts—one corresponding to each amount of strain on the c-clamp. These two waves interfere so as to form the fringes pictured here.
1 Historical Introduction

Ever since 1900 man has been able to record and retain as a permanent record almost any scene that his eyes perceived—through the process of photography. The optical lens had been invented and used several centuries before, and the formation of optical images with lenses was well understood by 1900. With the invention of the photographic process the importance of the lens in scientific investigation was greatly enhanced. The fortunate combination of lens and photographic emulsion made possible the charting of stars, planets, and galaxies; the recording of optical spectra; the picturing of minute microscopic specimens; the storage of large amounts of data in the form of small recorded images; and myriad other uses. Because of the vast scope of its scientific importance, the science of photography has advanced steadily over the past 70 or more years; even today new and important uses are being found.

Now science has at its disposal a new method of forming optical images: holography.

Holography is a relatively new process which is similar to photography in some respects but is nonetheless fundamentally different. Because of this fundamental difference, holography and photography will not be competing in the same areas. There are several applications for which holography is more suitable than photography, whereas most of the more important uses for photography remain unchallenged. Further, there are several tasks which can be performed with holography but not at all with conventional photography.

In order to point out the fundamental differences between holography and photography, we should understand in a general way how each works.

Photography basically provides a method of recording the two-dimensional irradiance distribution of an image. Generally speaking, each “scene” consists of a large number of reflecting or radiating points of light. The
waves from each of these elementary points all contribute to a complete wave, which we will call the "object" wave. This complex wave is transformed by the optical lens in such a way that it collapses into an image of the radiating object. It is this image which is recorded on the photographic emulsion.

Holography is quite different. With holography, one records not the optically formed image of the object but the object wave itself. This wave is recorded in such a way that a subsequent illumination of this record serves to reconstruct the original object wave, even in the absence of the original object. A visual observation of this reconstructed wavefront then yields a view of the object or scene which is practically indiscernible from the original. It is thus the recording of the object wave itself, rather than an image of the object, which constitutes the basic difference between conventional photography and holography.

A brief description of how the object wave is recorded will be useful before tracing the history of holography. One starts with a single monochromatic beam of light which has originated from a very small source. The requirements that this beam of light be monochromatic and that it originate from a small source together form the condition that the light be coherent. The requirement of coherence means that the light should be capable of displaying interference effects that are stable in time. This single beam of light is then split into two components, one of which is directed toward the object or scene; the other is directed to a suitable recording medium, usually a photographic emulsion. The component beam that is directed to the object is scattered, or diffracted, by that object. This scattered wave constitutes the object wave, which is now allowed to fall on the recording medium. The wave that proceeds directly to the recording medium is termed the reference wave. Since the object and reference waves are mutually coherent, they will form a stable interference pattern when they meet at the recording medium. This interference pattern is a complex system of fringes—spatial variations of irradiance which are recorded in detail on the photographic emulsion. The microscopic details of the interference pattern are unique to the object wave; different object waves (objects) will produce different interference patterns.

The detailed permanent record of this interference pattern on the photographic emulsion is called the "hologram," from which the word "holography" is derived. This photographic record, or hologram, now consists of a complex distribution of clear and opaque areas corresponding to the recorded interference fringes. When the hologram is illuminated with a beam of light which is similar to the original reference wave used to record the hologram, light will only be transmitted through the clear areas, resulting in a complex transmitted wave. Because of the action of the recorded interference fringes, however, this transmitted wave conveniently divides into three separate components, one of which exactly duplicates the original object wave. By viewing this reconstructed wavefront, one sees an exact replica of the original object, even though the object is not present during the reconstruction process. Thus holography is a two-step process by which images can be formed. In the first step a complex interference pattern is recorded and becomes the hologram. In the second step the hologram is illuminated in such a way that part of the transmitted light is an exact replica of the original object wave. The fundamental difference between holography and conventional photography is now quite evident.

This method of optical imagery is not really new. Nearly two decades ago British research scientist Dennis Gabor first conceived of, as he called it, "a new two-step method of optical imagery" [1]. It is only in the past few years, however, that the method has become widely known and used. The modern renaissance in holography had to await the general availability of the laser with the great temporal and spatial coherence of its light, but the really significant contributions to Gabor's original idea were more basic in nature.

The general idea of this two-step imaging process was suggested to Gabor by Bragg's x-ray microscope [2]. Bragg had been able to form the image of a crystal lattice by means of diffraction from the photographically recorded x-ray diffraction pattern of the lattice. The basic idea behind Bragg's method is a double-diffraction process, which is the crux of the holographic process. Image formation by double diffraction becomes clear if we note that the field diffracted by an object can be represented as a Fourier transform of the light distribution at the object [3]. Thus the second diffraction becomes a Fourier transform of the Fourier transform of the object, which is an image of the object itself. This means that diffraction from the hologram will reproduce the object wave, provided that all the amplitudes and phases of both diffractions are preserved.

It was just this question of phase preservation that represented the basic limitation to Bragg's method. Since he was able to record only irradiances, phase information was discarded. He was thus limited to applying the method only to a restricted class of objects, such as crystal lattices, for which the absolute phase of the diffracted field could be predicted. It is preserving the phase information, or at least rendering the recording of the phase unimportant, which represents the crux of Gabor's method. Bragg was able to circumvent the phase problem by using a class of objects for which a known phase change occurs between the incident and diffracted radiation. By his use of crystals having a center of symmetry, all the scattered radiation was either in phase or 180° out of phase with the incident radiation. Hence by recording the diffraction pattern photographically and
the object in the electron beam, just in front of a reduced image of the electron source. A hologram formed by the electrons diffracted from the object was recorded on a photographic plate some distance beyond the object. The hologram was then scaled up optically in the ratio of the light wavelength to the electron wavelength and illuminated with a light wave with the same aberration as the electron wave, scaled in the same ratio. Theoretically, then, the object was visible through the hologram in the original position and magnified by this same ratio of light to electron wavelength, or about 100,000X.

The method did not succeed because of certain technical difficulties, such as mechanical-and electrical stability. Haine and Dyson subsequently suggested an improved arrangement which increased the usable field, thus relaxing the electrical stability requirements [9]. By this method, the "transmission method," lenses were used between the object and the hologram to magnify the diffraction pattern. This increased the effective resolution of the photographic plate. The required apparatus was essentially identical with the classical electron microscope, thus facilitating the location of the object.

Efforts by Haine and Mulvey to implement this method were again frustrated by practical limitations [10]. They managed to obtain diffraction resolutions of about 6Å by eliminating as many of these problems as possible but could not go farther because of difficulties in holding the specimen stationary in the electron beam. It was necessary to hold the object stationary in relation to the objective lens to within a few angstrom units during exposure, a period of several minutes. Further difficulties were encountered in the optical reconstruction stage.

The most serious problem in reconstruction was the disturbance created by the twin wave. Because of the uncertainty of \( x \) in recording the phase in the hologram, there are two possible objects giving rise to the same exposure distribution in the hologram. One of these is the original object, the other a virtual object located symmetrically behind the source. Upon reconstruction, waves from both objects are formed. Therefore, in viewing the image of the real object, one has to look through an out-of-focus background image of the virtual object, a most annoying disturbance. Gabor [1] noted that if a condenser system is used to form a reduced image of the source, the twin image will be severely affected by the aberrations of that system. Thus it will appear blurred, whereas the image of the original object will appear sharply defined, the aberrations having been compensated for in this image during reconstruction. Gabor [11] also suggested a method whereby an obscuring mask is placed at a suitable formed image of the point source during reconstruction. Thus most of the wave containing the information of the real object is passed, but most of the background is
suppressed. Bragg and Rogers [5] suggested that since the reconstruction is really an in-focus image of the object plus a hologram of the conjugate object, a second hologram can be made of the conjugate object and subtracted from the original. The subtraction is performed by placing the second hologram in contact with the original image. The high-transmission portions of the secondary hologram fall on the high-density regions of the unwanted image and vice-versa, so that the background becomes uniform. The great precision required to register the two prevented complete success, but the effects of the twin image were reduced [5, 6].

Paralleling these early attempts to utilize holography for the improvement of electron microscopy were efforts by several workers to produce x-ray holograms. El Sum [6] produced an artificial x-ray hologram of a thin wire by photographing a published picture of the x-ray diffraction fringes of the wire. He managed to obtain a reasonable reconstruction, using light from this hologram, proving at least the feasibility of x-ray holography. Baez [12] did a theoretical study of the problems and also concluded that holograms and reconstructions with x-rays are feasible. He does note, however, that because of film resolution and source size limitations, useful resolution might be achieved with visible light. Further work on x-ray holography is still awaiting a small, monochromatic source of x-rays.

Aside from efforts by Rogers [13, 14] and Kirkpatrick and El Sum [15] to provide more satisfying conceptual explanations, the subject of holography lay dormant for almost a decade. Brief explanations of the principle published in a few optical textbooks [16, 17] represented about all of the published work on the subject for this long period of time. The most serious limitations of the method, which led to interest dying out, were the lack of an intense, coherent source in either the x-ray or optical region of the spectrum and the disturbing presence of the twin wave. Exposure times on the order of one hour were not unusual [6], and resolution in the reconstruction never reached theoretical predictions.

Interest began reviving in the field when Leith and Upatnieks [7] demonstrated a method for the complete elimination of the twin wave by a fairly simple means. Describing the holographic process from a communication-theory viewpoint, they realized that if the signal information (wavefront diffracted from the object) could be put on a carrier frequency (off-axis reference wave), the two reconstructed waves would then represent the sidebands of the process and be physically separated from each other. From an optical viewpoint, if the wave diffracted from the object is made to interfere with a reference wave which is off-axis, rather than in line, the hologram will be a gratinglike structure. Reconstruction will yield two waves representing the two first orders of the grating. One of these waves is the same as the original wave from the object; the other is the unwanted twin wave. Thus a physical separation in space of the two waves is achieved and the disturbing effects of the twin waves are eliminated. Gabor [1] noted that this turn of events would probably occur when he said, "... it is very likely that in light optics, where beam splitters are available, methods can be found for providing the coherent background which will allow better separation of object planes and more effective elimination of the effects of the 'twin wave' than the simple arrangements which have been investigated." Baez [12] came very close to introducing the off-axis concept in 1952 when he noted that "in an analogous way a diffraction grating forms a virtual image of a source," while explaining how a hologram forms an image by diffraction.

Thus Leith and Upatnieks' idea revived interest in holography. Many people began taking note and trying a few simple experiments. The "twin image" would no longer be termed the "unwanted image" but would prove to be a useful adjunct of the holographic process. Although the concept of an off-axis reference wave was a definite advance, there were still problems with dust and imperfections in the optical components. The slightest speck of dust on the lenses or mirrors gave rise to its own hologram, reducing the effective aperture of the hologram and reconstructing itself as noise. El Sum [6] and others took great pains to remove this source of noise by rotating as many of the components as possible, thus smearing out the holograms of the nonstationary dust particles. This method was successful to some degree but extremely impractical. Later developments would eliminate this problem also.

Another advantage of the new off-axis reference beam method is elimination of critical film processing. The original method required that the hologram be processed as closely as possible to a gamma (contrast) of two for linear transfer of exposure to amplitude transmission. In the Gabor technique any nonlinearities in the transmission-exposure transfer resulted in decreased image contrast because of the background light level. In the new technique any nonlinearities of the recording medium result mainly in higher diffraction orders. These higher orders are diffracted at angles larger than the first-order wave; thus nonlinear recording has little effect on the desired image. Curiously enough, processing the film to a high contrast with this new method, thereby increasing the nonlinearity of the recording, is actually somewhat beneficial. A high-contrast hologram results in a brighter image but with little disturbing effect from the higher order terms.

The off-axis reference beam method also results in the elimination of the effects of self-interference between different points of the object. In the earlier method this self-interference resulted in a veiling glare around the image. In the new method this noise term can be avoided completely.
Finally, the new method makes reconstruction possible for objects that do not transmit a large portion of the incident wave and also for continuous tone objects. In the earlier technique it was necessary that the major portion of the light passing the object not be diffracted. This undiffracted light is then only slightly modulated by the light diffracted from the object. In this way the loss of phase information in the recording process is rendered negligible; the resultant phase of the total disturbance at the hologram recording plane is almost that of the background wave. When the hologram is illuminated with the background wave alone the phase of the original total disturbance is approximated and a recognizable reconstruction results. In the new method the phase of the total disturbance at the hologram is recorded as a phase modulation of a gratinglike fringe pattern. This permits reconstructions of a wholly new class of objects which do not transmit a large portion of the incident wave, such as transparent letters on an opaque background and continuous-tone objects.

Thus the new method introduced by Leith and Upatnieks eliminated many of the annoying features of the original method, but the real renaissance of holography had to await two other important advances which were not long in coming.

About the same time (1962) that Leith and Upatnieks were introducing the off-axis reference beam method, people were beginning to make and use a radically different light source that would prove to be eminently suitable for holography. Thus the invention of the gas laser coincided nicely with the revival of interest in holography. The laser is capable of producing very intense monochromatic radiation in regions of the spectrum that can be recorded photographically. Because of the highly coherent nature of the light from a laser it can be focused down to an arbitrarily small spot, hence source size no longer limits the attainable resolution in a holographic image. The monochromaticity of the laser allows for full utilization of the off-axis recording scheme, since now many more interference orders (fringes) can be recorded. This yields much higher resolutions than had previously been obtained. Also there are no longer any restrictions on the size of the object to be used; holograms can now be made of very large objects.

The advent of the gas laser made possible still another important advance, again introduced by Leith and Upatnieks. In 1964 they introduced the concept of diffuse illumination holography [18]. Before this the only holograms that had been made were of thin transparent objects. Holograms of these kinds of objects often consisted of nearly recognizable shadowgrams of the object. Thus a small region of the hologram would bear almost a one-to-one correspondence with a small region of the object. Viewing the images formed with this type of hologram required some additional optical components, since an observer viewing a specularly illuminated transparency will see, for the most part, only that portion of the transparency which lies on a line between the light source and his eye pupil. Hence without optical aids, only a small portion of the image can be viewed at one time. On the other hand, if the transparency is illuminated diffusely, it can be viewed in its entirety with the eye in one location. This, then, is the idea that Leith and Upatnieks introduced into holography in 1964. By placing a diffuser, such as an opal glass, behind the object, a hologram is formed of both the diffuser and object. In this way it is possible to view the image formed from the hologram by merely looking through the hologram as through a window. The ability to view the images formed from this type of hologram in this way is not the only advantage. There are several even more interesting.

Because the object is diffusely illuminated there are no longer any recognizable shadowgrams of the object on the hologram. The light scattered from each point of the object spreads out so as to cover the entire photographic plate. This means that there is no longer a one-to-one correspondence between a portion of the hologram and a region of the object. The information about any single object point is recorded over the whole plate. Therefore only a small portion of the hologram is required to form an image of the whole object. In fact, in viewing the hologram with the unaided eye, only a portion of it roughly the size of the eye pupil is actually used. If the hologram were to be broken, scratched, torn, or damaged, it would still be possible to form an image of the complete object, although some resolution would be lost.

A further advantage of holograms of diffusely illuminated objects is that dirt or scratches on the mirrors, beam splitters, and/or lenses used in making the hologram no longer represent the problem they did in the earlier types. Most of these imperfections are, to a great extent, smeared out over the whole plate and thus have a negligibly small effect on the reconstruction.

Perhaps the single most important aspect of this technique is the ability to record holograms of diffusely reflecting, three-dimensional objects. It had been recognized from the beginning by Gabor that a hologram of a three-dimensional object should be capable of forming a three-dimensional image [1]. El Sum [6] and Rogers [14] managed to make holograms of objects in depth, even with the limited coherence lengths of the light sources at their disposal. The gas laser and the diffuse illumination concept made possible the formation of truly striking three-dimensional images. Since the object is diffusely illuminated or diffusely reflecting, light from a large range of perspectives reaches the photographic plate. An observer viewing the image formed by this hologram can move his head and see around foreground objects, just as if he were viewing the original object; he sees a truly three-dimensional image. He must refocus his eyes, depending on whether he is viewing a near or far object point.

The new diffuse-illumination holograms also lend themselves very nicely...
to the superposition of more than one hologram on a single photographic plate. Rogers [14] had made two holograms on a single plate by double exposure, but they were of two objects suitably situated so that the hologram of one did not obliterate too great a portion of the hologram of the other. Suitable positioning of the objects was necessary with the Gabor-type hologram since the information about an object was fairly well localized in the region surrounding the geometrical shadow of the object. Since this is not the case with diffuse-illumination holograms, it is a simple matter to make multiple holograms on a single plate and still obtain high-quality reconstructions. There is still a limitation on object position, however. To prevent the various images from each hologram from falling on top of one another, the hologram of each object should be recorded with a different angle between the object beam and reference beam. In this way the reconstruction of each object wavefront will be traveling in a different direction and hence will be separated in space. Each reconstruction can thus be viewed separately.

This concept led Leith and Upatnieks to propose a method of multicolor wavefront reconstruction [18]. By illuminating a colored object with coherent light in each of the three primary colors, each with its own reference beam, three holograms will be recorded. Reconstruction is accomplished by reilluminating with the three reference beams; a full-color object wavefront results. The scheme has been demonstrated, but better methods have evolved. They will be discussed in Chapter 7.

At about the same time that Leith and Upatnieks were advancing the field of holography at a great rate, Denisyuk [19] proposed an idea that proved to be of fundamental significance. He suggested that the wavefront from an object traveling in one direction be made to interfere with a coherent reference beam traveling in the opposite direction in a three-dimensional recording medium. In this way a standing wave pattern is set up in the recording medium which is uniquely related to the object wavefront. The medium therefore records a series of surfaces separated by one-half the recording wavelength in the medium. These surfaces, under appropriate conditions, are just the antinodal surfaces of the object wavefront. In this way the actual wavefronts of the object beam are recorded.

The recording medium is processed so that a change in the dielectric constant occurs where the exposure is high. Reillumination with the reference wave alone yields a reflected wave that is an exact replica of the object wave. Denisyuk's idea was thus a fruitful combination of Lippmann's [20] color photographic process and the hologram method of Gabor [1]. This idea was later investigated in detail by van Heerden [21] and has proved to be an important aspect of modern holography.

The holographic method differs significantly from the conventional photographic process in several basic respects and has distinct advantages in many areas. The most obvious advantage of holography is the ability to store enough information about the object in the hologram to produce a true three-dimensional image, complete with parallax and large depth of focus. There has been a great deal of work done in the attempt to produce three-dimensional images using conventional photographic techniques. These methods have been only partially successful because of the limited depth of field and restricted viewing conditions. An observer viewing a stereo pair, for example, cannot move his head from side to side and look behind foreground objects as he can with a hologram. The lenticular-type three-dimensional photograph allows limited parallax but has a rather severe depth limitation. The hologram, on the other hand, has a field of view that is limited in general only by the resolution of the recording medium. The depth of field recorded in a hologram is limited only by source bandwidth. Thus if a hologram is made of a three-dimensional object, it is equivalent to many conventional photographs, each taken from a different point of view and each focused at a different depth. Subsequent viewing of the hologram image at different depths requires only a refocusing of the viewing system. Hence it is fair to say that one hologram is worth a thousand pictures!

The quality of a holographic image is less sensitive to the characteristics of the recording medium than is the quality of a photograph. Holograms made on high-contrast material reproduce tonal variations of the object over a wide range. Nonlinear recording has only a small effect on the final image. Also, imperfections in the emulsion, such as scratches, have very little effect on the final image. Indeed a modern hologram is so redundant that only a small fraction of the holographic record is necessary to form a complete image.

Because of these basic differences between holography and conventional photography, many interesting and novel applications have been proposed. Few of these have been put to commercial use as yet, but the field is still young.

REFERENCES

2 Basic Arrangements for Holography

2.0 INTRODUCTION

Our discussion of holography begins with a description of the general arrangements currently used for recording and reconstructing. The discussion is restricted to the off-axis type of hologram first described by Leith and Upatnieks [1]. If this scheme is used, the object and reference beams are coincident at the recording medium, arriving from substantially different directions. This is achieved in practice by placing the object laterally some distance away from the source of the reference beam. The object is, of course, illuminated with a beam of light from the same source that provides the reference beam. As described in Chapter 1, the recording medium records the two-beam interference pattern; the precise details of the pattern are unique to the object used. This record is now called the hologram of the object and when it is illuminated with a single beam of light similar to the original reference wave, the hologram diffracts the light in such a way as to reconstruct the object wave.

2.1 BASIC DESCRIPTION OF HOLOGRAPHY

Conceptually, the simplest form of an off-axis hologram is one for which the object is just a single, infinitely distant point, so that the object wave at the recording medium is a plane wave (Fig. 2.1a). If the reference wave is also plane and incident on the recording medium at an angle to the object wave, the hologram will consist of a series of Young’s interference fringes. These recorded fringes are equally spaced straight lines running perpendicular to the plane of the figure. Since the hologram con-