Computer generated holograms: an historical review

G. Tricoles

Since Gabor's first hologram, the meanings of that term have grown with increased use of the invention. Computers expanded applications further, and this paper surveys the methods and techniques of computer generated holograms.

I. Introduction

A. Holograms; Holography

In the beginning, a hologram was a tangible record of an intensity pattern that was formed when a wave scattered by an object interfered with a coherent reference wave. The first holograms were formed by Gabor in research on reducing aberrations of electron microscopes.[1] These holograms were utilized in optical experiments to demonstrate a new two-step imaging principle. In the first step, an illuminated object scattered a field which interfered with the reference wave; in Gabor's experiments the reference wave was the wave bypassing the object. The interference pattern exposed a photographic plate; the developed plate was called a hologram. In the second step, the hologram was illuminated to produce an image, which occurred on the optical axis.

A hologram's appearance usually differs from the object's, but holographic images are remarkably realistic. Many holographic images are three-dimensional[2]; this property stimulated interest in holograms for displays. Leith and Upatnieks used a reference field that was from a coherent beam that was distinct from the object wave; this procedure produced a pair of off-axis images.

The term holography in analogy to photography has emerged to describe the expanding applications and techniques of holograms. In addition to imaging, holograms are directly useful for diagnostics without image formation because they are interferograms. Holograms are useful as optical elements to correct aberrations; this kind of application motivated Gabor's original work. Holograms are also useful for data storage. Some of the applications exploit the separation of the process into two steps; the separation of formation and reconstruction provides an interval for analysis and processing.

Optical data processing for synthetic aperture radar stimulated interest in holography during the 1960s. Much of this work was done at the University of Michigan.[3] Formation was done at centimeter wavelengths, which differ from visible reconstruction wavelengths. Gabor's experiments with visible light and their connection to DeBroglie waves further illustrate wavelength diversity between formation and reconstruction.

Since Gabor's 1948 publication, the meaning of the term hologram has grown to include many applications over a wide range of wavelengths in both electromagnetics and acoustics.[4–9] The first half of the word hologram stems from the Greek word for all; the second half suggests a record. The idea is that a hologram records both phase and amplitude in the form of an intensity pattern. Polarization information was not considered in the early days. The potential of holography for applications has stimulated diverse formation experiments, and as a result descriptions of holograms often include several adjectives; some are mentioned later.

The meaning of the word hologram has grown also because holograms can be generated by computers. Computer generated holograms (CGH) have many useful properties. For example, an object need not exist; an ideal wavefront can be computed on the basis of diffraction theory and encoded into a tangible hologram. Such holograms can be optical elements. CGHs can be synthesized for optical filters. The synthesis of holograms is connected to inverse scattering, itself a broad discipline.

This paper describes methods and techniques of CGH from an historical and broad view. Some details are mentioned but only partially; more complete descriptions are in the References and Bibliography.

The author is with General Dynamics, Electronics Division, P.O. Box 85227, San Diego, California 92138.

Received 15 March 1987.
0003-6935/87/204351-10$02.00/0.

B. Computer Generated Optical Filters

Computer generated holograms are similar to computer generated filters which were studied for optical signal processing of synthetic aperture radar data.[10] The detection problem involved filtering noisy 2-D signals or sets of noisy 1-D signals, which were recorded on photographic film. Filtering was done with a coherent optical system consisting of a point source, collimating lens, Fourier transform lens, and imaging lens; a signal was inserted in the front focal plane of the transform lens and a filter in its back focal plane. For noise additive to the signal, the transform and filtering reduced the noise amplitude. The basis of the method was to modulate a spatial carrier, that is, form a deformed grating. Modulated carriers were also described in Refs. 2 and 3. Kozma and Kelly, at the University of Michigan, constructed 1-D filters that operated on phase-only by computing the signal's Fourier transform, drawing a black and white 1-D grating, and photographically reducing the drawing to produce a filter transparency. These filters, which were developed for the detection of signals in noise, were precursors of computer generated holograms.

C. Overall Process of Computer Generated Holography

An overall schema for computer generated holography is shown in Fig. 1. It includes the following entities and processes:

Object The object need not exist physically; it can be imaginary or idealized. The object can radiate or be illuminated by an external source.

Wave Propagation Wave propagation is computed with theories appropriate for Fraunhofer, Fresnel, or near-field diffraction; the theories may be vector or scalar.

Hologram Surface The object-scattered field is evaluated on a hypothetical surface which is usually flat.

Hologram Fabrication The field on the hologram surface is represented by a transparency produced by a computer driven plotter, laser beam, or electron beam on correspondingly diverse materials. Scale reduction is usually necessary for reconstruction with visible light or other radiation of similar wavelength.

Hologram The computer generated hologram is a tangible mask with spatially variable transmittance.

Reconstruction The hologram is illuminated, and diffracted energy propagates to a detector.

Detector Photographic film is a common detector, but other sensors such as charge-coupled devices can be used. For microwaves small antennas and receivers have been utilized.

D. Computer Generated Holograms

Computer generated holograms were described by Brown and Lohmann in 1966.[11] Motivation included optical spatial filtering, which they experimentally demonstrated for 2-D objects. Imaging also was demonstrated. In addition, Ref. 11 pointed out two important general aspects of computer generated holograms.

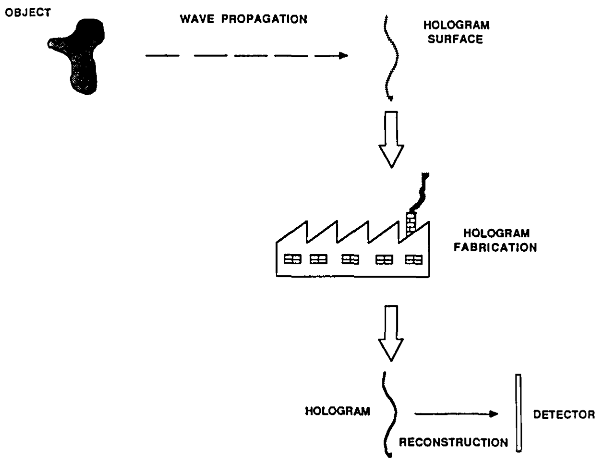

Fig. 1. Schema of computer generated holography.

One aspect is that the object need not exist; therefore, idealized wavefronts can be produced. This attribute underlies application to optical testing with holograms and the generation of optical elements. The second aspect is that hologram synthesis is the opposite from usual diffraction problems in which the diffracting object is given and the diffracted field is sought. Instead, the image is prescribed, and the diffracting object, the hologram, is sought; thus holography is connected to inverse scattering.

The work of Brown and Lohman stimulated interest in applications and research on computer generated holograms.

Research emphasized (1) methods for representing, or encoding, computed wavefront data into tangible holograms and (2) the consequences of approximations in formats and computer aided graphic methods; Sec. II briefly describes this work. Many applications have been made; Sec. III summarizes them and presents a chronological distribution of publications as a means of showing trends.

II. Types of Computer Generated Holograms

A. Detour Phase Holograms

The first computer generated holograms were made by Brown and Lohmann.[11] The holograms, intended for Fraunhofer diffraction, were formed by computing the image's Fourier transform and representing the transform values in a mask that had transparent apertures in an otherwise opaque screen. Because the mask's transmittance had values of zero or unity, the holograms were called binary. The transform plane was subdivided into regions of equal size, or cells. Three ways of representing the data were presented. In two, each cell contained an aperture, whose height or width depended on the transform magnitude at the center of the cell; in the other, each cell contained two apertures whose total width depended on the transform's magnitude. In all three representations, the lateral positions of an aperture was proportional to the transform's phase at the center of its cell. This lateral

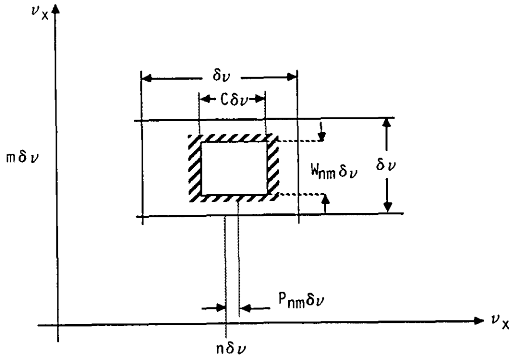

Fig. 2. Typical cell in a binary hologram. Position indices are n and m. Width is $C\delta\nu$, height $W_{nm}\delta\nu$, and lateral displacement from the cell center is $P_{nm}\delta\nu$.

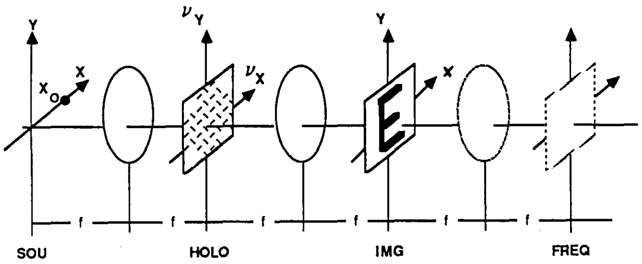

Fig. 3. Optical setup for reconstruction with a point source (SOU) at x_0. HOLO, IMG, and FREQ are, respectively, the hologram, image, and frequency planes.

shift led to the name detour phase, in analogy to diffraction gratings with unequally spaced rulings.

Brown and Lohmann's 1966 paper was significant. It demonstrated imaging and matched filtering; it also stimulated much research and many applications. However, according to Lohmann and co-workers, the paper was intuitive so in 1967 they presented a more analytical connection between reconstructions and hologram configurations and a discussion of approximations.

To describe the approximations we collect and interpret some formulas given by Lohmann and Paris.[12] The reason for the approximations is to describe the amplitude reconstructed from the hologram in the form of a Fourier series, which is equated to a Fourier series representation of the image. In the process, formulas result for dimensions of cells as shown in Fig. 2.

The aim is a hologram with transmittance $H(\nu_x,\nu_y)$, which produces an image amplitude $u(x,y)$ when illuminated by a plane wave $\exp(i2\pi\nu_x x_0) = E(\nu_x x_0)$ in the arrangement of Fig. 3, where $x_H = \lambda f\nu_x$ and $y_h = \lambda f\nu_y$. The requirement is that $h(x,y)$ the amplitude diffracted by the hologram be proportional to that of the image in a finite region; that is,

$$h(x,y) = \text{const.}\, u(x,y). \tag{1}$$

With the image width Δx and height Δy, the diffracted amplitude is

$$h(x,y) = \text{rect}(x/\Delta x)\ \text{rect}(y/\Delta y)$$
$$\int\int H(\nu_x,\nu_y)E[(x+x_0)\nu_x + y\nu_y]d\nu_x d\nu_y, \tag{2}$$

where rect$z$ has a unit value for $|z| \leq \frac{1}{2}$ and zero otherwise.

The hologram is represented by sampled values. The number of samples depends on image size and image resolution. If δx is the resolution element size in the image, the number of resolvable points in the image is, with $\delta_x = (\Delta\nu)^{-1}$,

$$N^2 = (\Delta_x\Delta_y)/(\delta_x)^2 = (\Delta_x\Delta_y)(\Delta\nu)^2. \tag{3}$$

N is called the number of degrees of freedom. The hologram is assumed to have at least N^2 points to preserve information; thus

$$(\Delta\nu/\delta\nu)^2 \geq (\Delta x/\delta x)^2 = (\Delta x\Delta\nu)^2. \tag{4}$$

In fact, $\delta\nu$ was taken to be $(\Delta x)^{-1}$ on the basis of the sampling theorem.

To develop expressions for aperture dimensions, consider Eq. (1) and apply scalar diffraction theory to the left-hand side. The hologram's transmittance is

$$H(\nu_x,\nu_y) = \sum_n\sum_m \text{rect}\{[\nu_x - (n+P_{nm})\delta\nu]/c\delta\nu\}$$
$$\times\ \text{rect}[(\nu_y - m\delta\nu)/W_{nm}\delta\nu]. \tag{5}$$

The rectangle functions describe the aperture dimensions in a cell as in Fig. 3. Illumination by a plane wave produces complex amplitude

$$h = \int\int H(\nu_x,\nu_y)E[(x+x_0)\nu_x + y\nu_y]d\nu_x d\nu y \tag{6}$$

$$= c(\delta\nu)^2\ \text{sinc}[c\delta\nu(x+x_0)] \sum\sum W_{nm}$$
$$\times\ \text{sinc}(yW_{nm}\delta\nu)E\{\delta\nu[(x+x_0)(n+P_{nm}) + ym]\}. \tag{7}$$

For the right-hand side of Eq. (1) the described image $u(x,y)$ in its Fourier representation with the tilde signifying the transform is

$$u(x,y) = \int\int \tilde{u}(\nu_x,\nu_y)E(x\nu_x + y\nu_y)d\nu_x d\nu_y. \tag{8}$$

Consider the Fourier sampling theorem

$$\tilde{u} = \sum_n\sum_m \tilde{u}(n/\Delta x, m/\Delta x)$$
$$\times\ \text{sinc}(\nu_x\Delta x - n)\ \text{sinc}(\nu_y\Delta x - m). \tag{9}$$

With Eq. (9), Eq. (8) gives

$$u(x,y) = \text{rect}(x/\Delta x)$$
$$\times\ \text{rect}(y/\Delta x) \sum_n\sum_m u(n\delta, m\delta)E[\delta\nu(xn + ym)]. \tag{10}$$

To satisfy the condition in Eq. (1), Eq. (7) is approximated to equal Eq. (10) term by term. The result is

$$(\delta\nu)^2 WnmE\,[x_0\delta\nu(n+P_{nm})] \approx \text{const.}\, \tilde{u}(n\delta\nu, m\delta\nu), \tag{11}$$

where const. $\tilde{u}(n\delta\nu,m\delta\nu)$ is defined as $c(\delta\nu)^2A_{nm}E(\phi_{mm}/2\pi)$, and A_{nm} and ϕ_{nm} are the magnitude and phase of the transform $\tilde{u}$. Thus

$$W_{nm} \approx A_{nm};$$
$$P_{nm} + n \approx \phi_{nm}/2\pi x_0 \delta\nu. \tag{13}$$

In words, W_{nm} controls magnitude; P_{nm} controls phase. If x_0 is chosen so that $x_0\delta\nu$ is an integer M,

$$P_{nm} \approx \phi_{nm}/2\pi M. \tag{14}$$

Equation (14) shows that the lateral displacement is proportional to the phase of the transform, an important relation, which has intuitive meaning as the phase difference of waves originating from two separated Huygens sources.

In equating terms on both sides of Eq. (1), in the forms in Eqs. (7) and (10), some approximations are necessary in Eq. (7). These are as follows:

$$\mathrm{sinc}[c\delta\nu(x + x_0)] \approx \text{constant in } |x| \le x/2; \tag{15}$$
$$\mathrm{sinc}(yW_{nm}\delta\nu) \approx 1 \text{ in } y \le (\Delta x/2); \tag{16}$$
$$E(xP_{nm}\delta\nu) \approx 1 \text{ in } |x| \le \Delta x/2. \tag{17}$$

The approximation of Eq. (17) has been called detour phase error; it has been extensively studied.[12,14] It can give inhomogeneous image intensity, which can be more noticeable near image edges. The approximation of Eq. (16) reduces image brightness but does not affect quality. The approximation of Eq. (15) produces inhomogeneous intensity; it has been called aperture error.

Despite the approximations, detour phase holograms are quite useful in imaging and filtering. In practice, distortions can be reduced by reducing cell size. However, as applications progress and requirements become more stringent, the approximations should be recalled. Reference 14 discusses a relationship between image error and the SNR.

Detour phase holograms contain additional approximations that can degrade images. Some of these errors also occur in other types of computer generated holograms, but for convenience the errors are described in this section.

Computer generated holograms represent sampled data. Transform values are computed at discrete points, and plotters often have a finite number of positions. This sampling can introduce aliasing error, which causes intensity in higher-order images to appear in lower orders.[15,16]

An error, called a quantization error, occurs when a finite number of phase or intensity values is plotted in the hologram.[17,18] The consequence is false images. An error known as gap and overlap occurs when phase exceeds ±π. This ambiguity can be eliminated by restricting the spacing between adjacent apertures.[19] Finally, a truncation error occurs when the hologram is smaller than the spatial extent of the transform.[14]

B. More Detour Phase Holograms

Lee developed a method that is based on decomposing the Fourier transform of the object into four quadrature components, which were represented by the real non-negative functions.[20] The phase of each component was coded into sampled functions, and the sum of the four functions represents the sampled function. Lee called this method delayed sampling. It does not require phase quantization.

The four functions are represented in a hologram by apertures at four laterally displaced, or shifted, positions within each cell of the hologram plane. The method does not require phase quantization because the transmittance of each cell varies. In computing the object's transform, the sampling rate along the horizontal ν_x axis in Fig. 2 is 4 times that in the orthogonal ν_y direction. Plotting was quantized because the hologram plotter was quantized. Lee demonstrated images for a binary object and a continuous tone object.

Hsueh and Sawchuck[21] developed binary holograms, called double-phase holograms, by decomposing the hologram transmittance into two-phase quantities. Each cell was divided into two subcells, which were laterally separated. Each subcell contained two transparent slits in an opaque background; slit widths were half of the cell width for diffraction efficiency. The vertical position of each slit was determined by the decomposed phase of the Fourier transform according to the detour phase principle. Consequently, even and odd diffraction orders were vertically displaced. Noise caused by subcell displacement and phase coding were analyzed; noise reduction methods were discussed. Visible images from holograms were presented.

Burckardt developed an approach that utilized three components.[22]

C. Nondetour Phase Holograms

Lee developed binary computer generated holograms that did not utilize the detour phase concept.[23] The holograms were based on considering the computed hologram as an interferogram. Positions and widths of the fringes were determined as the set of points satisfying inequalities involving the phase of the reconstructed wave. Experiments demonstrated reconstruction of spherical, conical, and helical wavefronts.

Lee later applied the view of computer generated holograms as interferograms to two problems of detour phase holograms.[24] One problem was sampling at discrete points in the hologram plane. The other problem was to phase variations exceeding 2π. Methods for encoding both amplitude and phase were described.

The spatial bandwidths of Lee-type nondetour phase holograms have been analyzed with frequency modulation theory; a quantization error model was presented.[25]

Burch developed holograms that encoded the Fourier sine and cosine transforms of real functions that describe objects.[26] Holograms are optically recorded by adding a bias term to the sum of the transforms. The reconstructions give pairs of off-axis holograms, but the on-axis light, corresponding to the object's autocorrelation, is lower than that for holograms formed with an inclined reference beam.

A wavefront reconstruction device that is similar to a lens has been developed.[27,28] This device, called a kinoform, resembles a blazed dielectric transmission grating. Kinoforms are useful as optical filters with noncoherent light. They have high diffraction efficiency compared with many kinds of holograms. They diffract on-axis rather than into off-axis orders as do sampled holograms or holograms formed with an inclined reference beam.

III. Literature Survey

A. Temporal Distribution of Publications

An analysis of the temporal distribution of published papers suggests some trends. (Publications are only one possible measure of activity; significance is another matter.) This analysis included approximately 200 papers published in the 21 years from 1966 through 1986. The bibliography is somewhat arbitrary. It includes papers from the following journals:

Applied Optics;
Chinese Physics;
Electronics and Communications in Japan;
Electronics Letters;
IEEE Transactions on Computers;
Journal of the Optical Society of America, Including A;
Laser Optoelektronika;
Nouvelle Revue d'Optique
Onde Electronique;
Optica Acta;
Optical Engineering;
Optics Communications;
Optica Pura Appl;
Optik
Proceedings of the IEEE;
Review of the Electrical Communication Laboratories (*Tokyo*).

It also includes papers in SPIE Proceedings and Proceedings of the Optical Computing Conference. However, it excludes abstracts of the annual meeting of the Optical Society of America and summaries of papers presented at the International Commission of Optics meetings.

No claim is made that the set of approximately 200 papers includes all on computer generated holography. Some recent books contain useful reviews.[29–31]

As a basis for discussion consider Fig. 4, which shows a plot of the number of papers published in each year from 1966 through 1986. In 1966 one paper by Brown and Lohmann was published. The number per year has an increasing trend until it reaches a maximum of fifteen papers in both 1974 and 1975. A minimum occurs in 1977 and 1978 with an increase to a second maximum of twenty-five papers in 1983. The maximum is followed by a decrease to thirteen papers in 1986; this maximum exceeds the minimum of seven papers in 1977 and 1978. The number of papers may not reliably show interest in and vigor of a field, but it does show activity. Publication is affected in a complicated way by the maturity of a field because applications stimulate research and vice versa. Research funding may also influence activity.

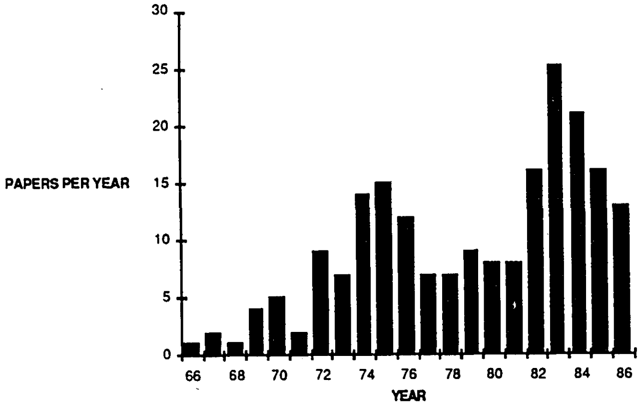

Fig. 4. Number of CGH papers published per year.

As a matter of interest we note that the 1985 Annual Meeting of the Optical Society of America had three papers on CGH; the 1986 meeting had five.

Although Sec. III.B describes applications, we anticipate it and state that the maxima in 1975 and 1983 have an interpretation. In 1975, eight of the fifteen papers were on techniques (methods and theory) of computer generated holography, and four were on optical elements with one or two on other applications. In 1983, ten papers of twenty-five were on optical processing, eight on holographic techniques, and seven on optical elements. This analysis is somewhat approximate because some papers fall into two classifications. Nevertheless, the nature of papers in the two years suggests that in 1975 emphasis was on improving CGH and that in 1983 the application of optical processing emerged.

B. Applications

Computer generated holograms now have several diverse uses which are grouped into broad categories as follows. Specific topics illustrate the nature of the categories.

Diagnostic and Testing acoustic mapping of the earth, determining particle sizes and scattering properties, analyzing fiber optical modes, analyzing vibrations, visualizing aberrations.

Digital and Optical Interconnects sequential optical logic operations, digital optical architecture and computing.

High-Energy Physics character detection, processor alignment.

Imaging and Display map displays, map transformations, hologram scaling, colored displays, 3-D displays, reduced quantization error, electron microscopy, image processing and deblurring, stereoscopic displays.

Improved Holographic Techniques computational efficiency, photographic film and alternate recording materials, detour phase and quantization er-

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | **sales@docketalarm.com** | **1-866-77-FASTCASE**