Published in the Journal of Electronic Imaging, vol. 2, #1, January 1993, pp. 28-34.

Interactive Computation of Holograms Using a Look-up Table

Mark Lucente
MIT Media Laboratory, Spatial Imaging Group
20 Ames St., E15-416, Cambridge, MA 02139 USA

ABSTRACT

Several methods of increasing the speed and simplicity of the computation of off-axis transmission holograms are presented, with applications to the real-time display of holographic images. The bipolar intensity approach allows for the real-valued linear summation of interference fringes, a factor of 2 speed increase, and the elimination of image noise caused by object self-interference. An order of magnitude speed increase is obtained through the use of a precomputed look-up table containing a large array of elemental interference patterns corresponding to point source contributions from each of the possible locations in image space. Results achieved using a data-parallel supercomputer to compute horizontal-parallax-only holographic patterns containing 6 megasamples indicate that an image comprised of 10,000 points with arbitrary brightness (gray scale) can be computed in under 1 second. Implemented on a common workstation, the look-up table approach increases computation speed by a factor of 43.

INTRODUCTION

The real-time display of holographic images has recently become a reality. The Spatial Imaging Group at the MIT Media Laboratory has reported the successful generation of small three-dimensional (3-D) computer-generated holographic images reconstructed in real time using a display system based on acousto-optic modulation of light[1, 2, 3]. In any holographic display, a computer-generated hologram (CGH) must be computed as quickly as possible to provide for dynamic and interactive images. However, the numerical synthesis of a holographic interference pattern requires an enormous amount of computation, making rapid (~1 s) generation of holograms of even limited size impossible with conventional computers.

A holographic fringe pattern is computed by numerically simulating the physical phenomena of light diffraction and interference[4]. Computation of holographic interference patterns often utilizes the fast Fourier transform (FFT) algorithm. Though relatively fast for images composed of discrete depth surfaces[5, 6], this approach becomes slow when applied to images that extend throughout an image volume. A more general ray-tracing approach computes the contribution at the hologram plane from each object point source. This method can produce arbitrary 3-D images, including image-plane holograms (i.e., images that lie in the vicinity of the hologram), a case that is more suitable for various display geometries. However, this method is slow, since it requires one calculation per image point per hologram sample. As presented in this paper, the application of several methods of reducing computation complexity leads to computation times as short as 1 s on a data-parallel-processing supercomputer. First, a *bipolar intensity* representation of the holographic interference pattern is applied and shown to eliminate unwanted image artifacts and simplify calculations without loss of image quality or generality. Second, a look-up table approach is described and shown to provide a further speed increase and to reduce computation to a minimum. Finally, exemplary computation times are presented.

HOLOGRAPHIC IMAGING

This paper focuses on the computation of off-axis transmission holograms possessing horizontal parallax only (HPO). It is possible to represent an HPO hologram with a vertically stacked array of one-dimensional (1-D) holographic lines[6, 7]. The goal of this paper is to compute these 1-D *holo-lines* as quickly as possible. Consider an HPO hologram made optically using a reference beam with a horizontal angle of incidence. Spatial frequencies are large in the horizontal direction (~1000 lp/mm) and increase with the reference beam angle. Since an HPO CGH contains only a single vertical perspective (i.e., the viewing zone is vertically limited to a single location), spatial frequencies are low (~10 lp/mm) in the vertical dimension. The number of holo-lines is, therefore, matched to the vertical image resolution. In the horizontal dimension, the sampling rate (or *pitch*) must be high to accurately represent the holographic information. During reconstruction of this hologram, diffraction occurs predominantly in the horizontal direction. A holo-line diffracts light to a single horizontal plane (*scan plane*) to form image points describing a horizontal slice of the image. Therefore, one holo-line should contain contributions only from points that lie on a single horizontal slice of the object. Essentially, the 2-D holographic pattern representing an HPO 3-D

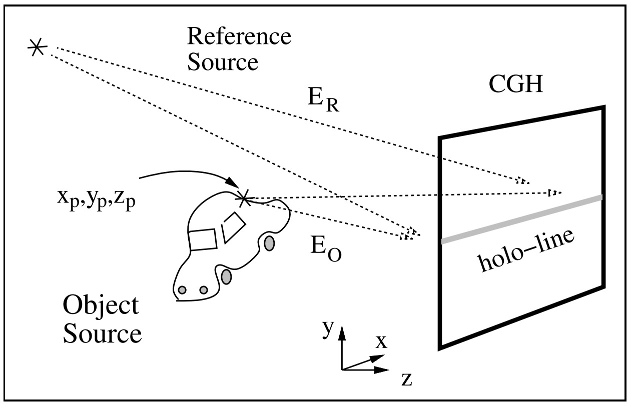

Figure 1: General geometry for HPO CGH.

image can be thought of as a stack of 1-D holograms or holo-lines. The information content of a CGH must be reduced to a size and format that can be manipulated by existing computers. Eliminating vertical parallax reduces CGH information content by at least a factor of 100 by reducing the vertical spatial frequency content from roughly 1000 to roughly 10 lp/mm. Sacrificing the size of the viewing zone (i.e., reducing the required horizontal pitch) and the size of the hologram also reduces the amount of data in the CGH. In the first generation MIT system, the data content (space-bandwidth product) of the CGH was as low as 2 Mbytes[2].

In optical holography[8], the object light and the reference light are incident at the plane of the hologram, represented with spatially varying complex time-harmonic electric field vectors, E_O and E_R. It is assumed that both are mutually coherent sources of monochromatic light. For this analysis, the units of an electric field amplitude are normalized so that the square of magnitude corresponds to optical intensity; the polarizations are assumed to be identical and for simplicity are not specified. The total time-harmonic electric field incident on the hologram is the interference of the light from the entire object and the reference light, $E_O + E_R$. The total intensity pattern,

$$I_{TOTAL} = |E_O|^2 + |E_R|^2 + 2\Re e\{E_O E_R^*\} \tag{1}$$

is a real physical light distribution comprised of three components. The first term is the object self-interference, a spatially varying pattern that is generated when interference occurs between light scattered from two different object points. During image reconstruction, this component of the holographic pattern is unnecessary and often produces unwanted image artifacts. In optical holography, a common solution is to spatially separate the object self-interference artifacts from the reconstructed image by increasing the reference beam angle to at least three times the angle subtended by the object. However, in computational holography, a large reference beam angle is a luxury that one does not have. Instead, the solution is simply to exclude the object self-interference during computation. The second term in I_{TOTAL} is the reference beam intensity and represents a spatially nearly invariant (dc) bias that increases the value of the intensity throughout the hologram. In computational holography, the reference bias can be left out, since the CGH will eventually be normalized. (Normalization is the numerical process that limits the range of the total fringe pattern by introducing an offset and a scaling factor to tailor the fringe pattern to fit the requirements of a display system.) Finally, the third term is the interference between the object wave front and the reference beam. This fringe pattern – called I_B – contains the holographic information that is necessary and sufficient to image reconstruction.

COMPUTATION USING POINT SOURCE SUMMATION

In this paper, the images to be generated are approximated as a collection of self-luminous points of light. The horizontal, vertical and depth locations are specified as x_p, y_p and z_p. (See Figure 1.) Each point has an associated real-valued magnitude (a_p) and phase (ϕ_p). The square of the magnitude (a_p^2) is proportional to the desired brightness of an image point, and ϕ_p is the phase relative to that of the reference beam. The CGH is positioned at $z = 0$. Only the x dependence of E_O and E_R and other physical quantities need to be considered in computing a given holo-line. (Hereafter, functional dependence on x, y or z is explicit, e.g $E_{O(x)}$ and $E_{R(x)}$.) A full CGH is computed simply by generating an array of holo-lines for each value of y. The fan-shaped beams of light from each of the object points contained within a scan plane contribute to a limited width of the particular holo-line that is intersected by the scan plane. Within the region of contribution, the phase of the object wave front, $\Phi_{p(x)}$, is approximated as an angularly truncated spherical wave front centered at the point source location:

$$\Phi_{p(x)} = k\, r_{p(x)} + \phi_p \qquad \text{where} \qquad r_{p(x)} = [(x - x_p)^2 + z_p^2]^{\frac{1}{2}} \tag{2}$$

is the oblique distance to a location on the holo-line and is a function of x. The wave number is $k = 2\pi/\lambda$, where λ is the free-space wavelength of the light. The time-harmonic representation of the total object field for a single holo-line is

$$E_{O(x)} = \sum_{p=1}^{N_{PTS}} \varepsilon_{(x)}\, \frac{a_p}{r_{p(x)}}\, \mathrm{e}^{i\Phi_{p(x)}} \tag{3}$$

where N_y is the number of object points contributing to this particular y-valued holo-line. Finally, to avoid singularities, it is assumed that the magnitude of z_p is never less than some small amount, e.g., 10λ.

In Eq. 3, the expression $\varepsilon_{(x)}$ is an envelope function used to specify the region of contribution for a given object source. In its simplest form, $\varepsilon_{(x)}$ is equal to unity within the region of contribution of point p, and equal to zero for all other values of x. The envelope function $\varepsilon_{(x)}$ provides a means with which to limit the spatial frequencies of the holographic fringe pattern to avoid *aliasing* in the CGH at the extreme angles of object light incidence. Object "light" used to compute the CGH must have a minimum angle of incidence that is greater than that of the reference beam to produce non-overlapping real and virtual reconstructed images. At the other extreme, the total angle subtended by the incident reference and object beams cannot be so large as to give rise to spatial frequencies that cannot be adequately represented given the sampling pitch. For a horizontal sampling pitch of $1/d$, the maximum spatial frequency (f_{max}) that can be represented is $1/2d$, according to the Nyquist Sampling Theorem. Higher spatial frequencies cause aliasing and destroy image quality. In addition to providing *anti-aliasing*, $\varepsilon(x)$ is used for the purposes of image occlusion[9] and advanced image lighting models, resulting in a more realistic looking image.

The reference beam E_R is a point source at some specific location (x_R, z_R) with a horizontal angle of incidence $\theta_R = \arctan(x_R/z_R)$ and curvature in the x dimension only; it is collimated in the y dimension. The time-harmonic representation of the reference beam field at any holo-line is

$$E_R(x) = \frac{a_R}{r_R} \mathrm{e}^{i\Phi_R(x)} \qquad \text{where} \qquad \Phi_R(x) = k[(x - x_R)^2 + z_R^2]^{\frac{1}{2}} \tag{4}$$

and $\frac{a_R}{r_R}$ is the magnitude (approximated as constant versus x) of the reference wave at the center ($x = 0$) of the hologram.

BIPOLAR INTENSITY

The third term of Eq. 1, called $I_B(x)$, contains all of the information needed to reconstruct the image in a given scan plane. Note that I_B is real valued. It represents the combined intensity variations (*fringes*) resulting from each object point interfering with the reference beam. Since it contains negative values as well as nonnegative values, it is a *bipolar* intensity. It exists physically only when superimposed on the first two bias terms in Eq. 1. Computationally, however, I_B can range both positive and negative since it is represented numerically, and is later offset during normalization. The bipolar interference pattern $I_B(x)$ is numerically simpler to compute and has the advantage of containing neither object self-interference nor bias components. To apply $I_B(x)$, the expressions for the object and reference beam wave fronts (Eqs. 3 and 4) are substituted:

$$\begin{aligned} I_B(x) &= 2\Re e \left\{ \left[\sum_{p=1}^{N_{PTS}} \varepsilon(x) \frac{a_p}{r_p(x)} \mathrm{e}^{i\Phi_p(x)} \right] \left[\frac{a_R}{r_R} \mathrm{e}^{i\Phi_R(x)} \right]^* \right\} \\ &= 2\frac{a_R}{r_R} \sum_{p=1}^{N_{PTS}} \Re e \left\{ \varepsilon(x) \frac{a_p}{r_p(x)} \mathrm{e}^{i\Phi_p(x) - i\Phi_R(x)} \right\} \end{aligned}$$

and finally simplified to:

$$I_B(x) = 2\frac{a_R}{r_R} \sum_{p=1}^{N_{PTS}} \varepsilon(x) \frac{a_p}{r_p(x)} \cos[\Phi_p(x) - \Phi_R(x)] \tag{5}$$

The right-hand side of Eq. 5 is simply a scaled sum of the real-valued cosinusoidal fringe pattern resulting from the interference of point source p with the reference beam. The bipolar intensity approach is to compute only I_B (instead of I_{TOTAL}), using only these real-valued cosinusoidal fringes.

The computational advantage of the bipolar intensity approach is readily seen by comparison to computation of the full interference pattern, I_{TOTAL}. The latter method requires the use of complex-valued arithmetic, i.e., manipulating both the real and imaginary parts of the object and reference light. In comparison, the bipolar intensity approach uses only real-valued cosinusoidal fringes; to generate the CGH, these real values are simply summed. A speed-up factor of 2 is expected.

After an intensity pattern has been computed, it must be normalized to satisfy the output device requirements of the CGH display system. Because normalization generally scales the entire pattern, the leading factor of $2\frac{a_R}{r_R}$ on the right side of Eq. 5 is hereafter excluded. The reference beam intensity (the square of $\frac{a_R}{r_R}$) is no longer meaningful. This makes physical sense when considering that in optical holography, the purpose of choosing an optimal reference beam intensity ratio (relative to the object light) is to provide a sufficient offset and scaling of the interference fringes to keep them within the range of sensitivity of the recording medium. Computationally, offset and scaling are provided automatically during normalization. Therefore, the factor of $2\,\frac{a_R}{r_R}$ is set arbitrarily to unity. Substituting the definition of $\Phi_{p(x)}$(Eq. 2), Eq. 5 becomes

$$I_{B(x)} = \sum_{p=1}^{N_{PTS}} \varepsilon(x)\, \frac{a_p}{r_{p(x)}}\ \cos[k\, r_{p(x)} + \phi_p - \Phi_{R(x)}] \tag{6}$$

which is hereafter called the bipolar fringe method of CGH computation. No reference beam ratio needs to be specified during computation, and bias buildup is not an issue. Compare this bipolar intensity method to the physical process occurring in some photorefractive crystals[10] (e.g., lithium niobate), in which uniform (dc) intensity is not recorded due to the material's negligible response to intensity patterns with low spatial frequencies. Researchers exploit this absence of bias build-up in order to sequentially expose multiple holographic intensity patterns.

PRECOMPUTED ELEMENTAL FRINGES: THE LOOK-UP TABLE APPROACH

Continuing with the bipolar intensity summation approach, further improvements in computation speed are gained through the use of a precomputed look-up table containing all possible elemental fringes. A table can contain the precomputed contributions to $I_{B(x)}$ from an image point of unity magnitude for each possible value of (x_p, z_p) in the image volume. However, the image volume is not generally discretized. Therefore, the indices X_i and Z_i are introduced. They are defined as a set of discretized horizontal and depth locations spanning the entire range of the image volume. (The image volume is already discretized vertically to match the the vertical resolution of the HPO CGH.) Before the table is generated, the resolutions of X_i and Z_i must be chosen to discretize the image volume. Since the acuity of the human visual system is limited, to chose resolutions that do not visibly degrade the image is possible. The discretization steps must be sufficiently small such that images points appear to be continuous when viewed from the specified viewing distance. For example, the human visual system generally sees as continuous two points that are separated by 3 milliradians of arc. In the MIT display, since the image is viewed at a distance of approximately 600 mm, and 600 mm$\times$ 0.003 = 180 microns, a horizontal discretization step of 150 microns is chosen.

Eq. 6 must be altered so that $I_{B(x)}$ is expressed in terms of discretized image point locations instead of x_p and z_p. The expressions $\bar{x_p}$ and $\bar{z_p}$ are used to represent x_p and z_p rounded-off to the nearest values of X_i and Z_i. Therefore, the only change made to Eq. 6 is to substitute $\bar{x_p}$ and $\bar{z_p}$ for x_p and z_p in the expression for $r_{p(x)}$ (in Eq. 2). (Note that the variable x is already discretized due to the the sampled nature of the CGH. The horizontal pitch is dictated by the intended display device.) An efficient look-up table should include all of the spatial dependence included in the right-hand side of Eq. 6. Therefore, the elemental fringe look-up table $\mathrm{T}(X_i, Z_i, x)$ is defined as:

$$\mathrm{T}(X_i, Z_i, x) \equiv \frac{\varepsilon(x)}{r_i(x)} \cos[k\, r_i(x) - \Phi_R(x)] \tag{7}$$

$$\text{where} \quad r_i(x) = [(x - X_i)^2 + {Z_i}^2]^{\frac{1}{2}}. \tag{8}$$

The table looks like an array of cosinusoid-like fringes that have an approximately linear chirp in spatial frequency with respect to x. The rate of chirp is a function of the point source depth Z_i, and the horizontal position of the fringe is a function of X_i. To incorporate anti-aliasing, $\varepsilon(x)$ is set to unity within the region of contribution, and zero elsewhere. (This binary representation of $\varepsilon(x)$ is for simplicity. Any further variation in $\varepsilon(x)$ can be dealt with during computation, or with additional look-up tables.) When utilizing the look-up table approach, ϕ_p is set to zero (or randomized) for simplicity.

Once precomputed, the table T is used during the computation of a CGH to represent a specific image. Rather than having to compute a cosinusoidal fringe each time one is needed (as indicated in Eq. 6), the lookup table maps each $(\bar{x_p}, \bar{z_p})$ to the appropriate elemental fringe pattern. Thus, $I_B(x)$, expressed in terms of the precomputed table, is

$$I_B(x) = \sum_{p=1}^{N_{PTS}} a_p\, \mathrm{T}(\bar{x_p}, \bar{z_p}, x) \tag{9}$$

Since each holo-line is computed using the same $E_R(x)$, one precomputed table is used in the computation of every holo-line. Computation for a given holo-line at vertical position y is as follows:

- For every point on scan plane y:
  - Round off x_p and z_p to $\bar{x_p}$ and $\bar{z_p}$ to index the desired elemental fringe.
  - For each x sample in $I_B(x)$:
    - Scale $\mathrm{T}(\bar{x_p}, \bar{z_p}, x)$ by a_p.
    - Accumulate $a_p\, \mathrm{T}(\bar{x_p}, \bar{z_p}, x)$ in $I_B(x)$.

After each holo-line is computed (at each value of y), then normalization and output are performed depending on the specific display system.

The issue of object point phase ϕ_p must be examined to justify setting the phase to zero for all points. Phase is invisible to the viewer. Object phase is important only in the interactions among image points. If the discretization steps of X_i and Z_i (typically 150 microns and 1000 microns) are much larger than the spot size of the display system, no overlap occurs between image points. In this case, as was the case in the MIT holographic display, restricting ϕ_p

does not limit the size, depth, or quality of an image. Therefore, the relative phases ϕ_p are unimportant and can be arbitrarily set to zero. (If nonzero phase values are required, two look-up tables in quadrature are sufficient for this more general case.) In some applications, ϕ_p can be randomized to make more efficient use of the dynamic range. A random phase factor for each point can easily be incorporated into the look-up table.

Computational Complexity

Computational complexity is dramatically reduced through the use of the precomputed look-up table. Consider the full complex computation of I_{TOTAL} that requires a minimum of five additions, five multiplications, one division, one square root, and two cosine (or sine) function calls. In addition, after the real and imaginary field components are summed, they must be squared and added together, and then the square root must be taken at all samples of I_{TOTAL}. In comparison, by using the look-up table approach, only one multiplication and one addition are required at each hologram sample for a given object point. Therefore, an order of magnitude of speedup is expected through use of the precomputed table. Note that the bipolar intensity simplification reduces complexity by allowing the table to contain only real values.

Look-up Table Quantization

The greatest drawback to the precomputed table method is the enormous size of the table. For example, given the rather minimal dimensions of the first generation MIT real-time display system (40-mm wide holo-lines with a pitch of 1 micron^{-1}), X_i typically has 250 discrete values, and Z_i has 50 discrete values; the table requires over 1.6 gigabytes of memory! When using the Connection Machine Model 2 supercomputer (Thinking Machines, Inc., Cambridge, MA), the solution was to quantize the look-up table into only four levels. In this way, only 2 bits (rather than the standard 32-bit floating-point representation) are necessary for each sample point, and the table occupies only 8×10^8 bits of memory, or the equivalent of a manageable 100 Mbytes. (Notice that the r_p^{-1} term cannot be included in the 2-bit look-up table since it requires a more continuous representation. However, for limited viewing zone widths, this term can be approximated by z_p^{-1} and used to scale a_p, resulting in a negligible decrease in speed.) The added quantization noise is acceptable, and by tailoring the exact waveform the theoretical signal-to-noise ratio is 243:1.

Look-up table quantization increases image noise due to the loss of accuracy. In practice, however, table quantization has little effect on the final image quality since $I_{B(x)}$ must be quantized when stored in the output framebuffer of the computing system. For example, in the first generation MIT system, the framebuffer used was capable of storing 6M samples, each represented by 8 bits, giving the output data 256 possible quantized levels. Typically, the object light from at least 64 objects points overlap at a single sample of $I_{B(x)}$. If each object point has an average amplitude of 1.0, then a typical elemental fringe pattern from the look-up table will occupy only about $256/64 = 4$ different levels in the framebuffer. Therefore, a four-level look-up table is well suited to this display system.

Reduction of Table Rank from Three to Two

The precomputed table is a data array of rank three, indexed by X_i and Z_i in the image scan plane and by x (on the hologram). One way to reduce the size of the table is to precompute values for each possible $\Delta_x = (x - X_i)$ rather than for each x *and* each possible X_i. In this way, the table is reduced to rank two.

During the table look-up computation, x_p and z_p are rounded off to $\bar{x_p}$ and $\bar{z_p}$, the nearest values of X_i and Z_i, making $\Delta_x = x - \bar{x_p}$ for a given image point p. In terms of these variables, the expression for r_p (see Eq. 2) becomes a function of $\bar{z_p}$ and Δ_x. Looking at Eq. 6, the envelope function $\varepsilon(x)$ is actually a function of Δ_x/z_p, since ε limits the angular direction of light propagation (equal to $\arctan[\Delta_x/z_p]$). Thus, Eq. 6 becomes

$$I_B(x) = \sum_{p=1}^{N_{PTS}} \varepsilon(\Delta_x)\, \frac{a_p}{r_{p(\bar{z_p},\Delta_x)}} \cos[k\, r_{p(\bar{z_p},\Delta_x)} - \Phi_R(x)] \tag{10}$$

$$\text{where} \quad r_{p(\bar{z_p},\Delta_x)} = ({\Delta_x}^2 + \bar{z_p}^2)^{\frac{1}{2}}. \tag{11}$$

Now only $\Phi_R(x)$ is an explicit function of x (rather than of Δ_x), and must therefore be manipulated to become a function of Δ_x. First, by restricting the reference beam to be a plane-wave (i.e., $r_R \to \infty$), the reference phase is simply $\Phi_R(x) = k_R\, x$, where $k_R = k \sin\theta_R$. The second restriction is that X_i be discretized by $2\pi/k_R$, making every possible value of $k_R\,\Delta_x$ differ by exactly $2\pi m$ for a given value of x, where m is some unimportant integer value. Consider that when computing the cosine of total phase, any integer multiple of 2π is ignored. Therefore, $\Phi_R(x)$ can be expressed as $\Phi_R(x) = k_R\, x + 2\pi m = k_R\,\Delta_x$ for all values of X_i. Finally, Eq. 6 can be expressed as a function of only two variables $\bar{z_p}$ and Δ_x:

$$I_B(x) = \sum_{p=1}^{N_{PTS}} \varepsilon(\Delta_x)\, \frac{a_p}{r_{p(\bar{z_p},\Delta_x)}} \cos[k\, r_{p(\bar{z_p},\Delta_x)} - k_R\Delta_x]. \tag{12}$$

A corresponding look-up table of rank two can be used, with the indices $\bar{z_p}$ and Δ_x.

Essentially, given the restriction of a plane-wave reference beam, the elemental fringe patterns are shift-invariant in x for any $\bar{x_p}$ equal to an integer multiple of $2\pi/k_R$. The first advantage is a reduction in size equal to at least the number of discrete values of $\bar{x_p}$ used in the rank-three table (~100 or more). The second advantage is that X_i is discretized in very small steps ($2\pi/k_R$ ~10 microns) without increasing the size of the rank-two table; the horizontal image discretization is much finer than can be seen by the human visual system. The only drawback to the rank-two table approach is the requirement that the reference beam be a plane wave. However, this is a common case, and is applicable to many display architectures. As demonstrated in Section 6, this contraction of rank is especially useful when implemented on a standard serial-processing computer.

Application to the Computation of Stereograms

For some applications, a stereogram[11] approach may be desirable when presenting 3-D images. In general, a stereogram consists of a series of 2-D object views differing in horizontal point-of-view. The views can be rapidly generated using a computer graphics rendering process, or they can be captured from a real scene by using an array of cameras. The views are presented to the viewer in the correct horizontal location, resulting in the depth cues of binocular disparity (or *stereopsis*) and (horizontal) motion parallax. The 2-D perspective views are generally imaged at a particular depth position and are multiplexed by horizontal angle of view. A given holo-line in this case must contain a holographic pattern that diffracts light to each of the horizontal locations on the image plane. The intensity from a particular horizontal viewing angle should be the image intensity for the correct perspective view. This is accomplished simply by making the amplitude of the fringe contribution a stepwise x-function of the intensity of each image point from each of the views. Calculation involves reading each of the views into the computer, using the view and its corresponding segments of the look-up table to compute the fringe pattern contribution for this view, and accumulating these contributions into the final CGH.

To facilitate rapid computation of stereogram-type CGHs, the precomputed table is indexed by image x-position and view-angle (rather than by X_i and Z_i). Furthermore, the table can be reduced to rank two as described above, making the table much smaller. Stereogram CGHs have been computed and displayed on the MIT real-time holographic display system. A major advantage to the stereogram approach is the integration of sophisticated lighting and shading models in the computer graphics rendering process to produce extremely realistic images possessing important 3-D image characteristics, such as occlusion (overlap) and specular reflections.

RESULTS

Hologram computation has been implemented for use in the MIT real-time display system using the methods of bipolar intensity summation and precomputed elemental fringe patterns. The MIT display required a CGH that contained 192 holo-lines, each of 32K samples, for a total of 6 Msamples. Each sample represented a horizontal physical spacing of 1 micron, or a pitch of 1000 lp/mm. A Connection Machine Model 2 (CM2) employed a data-parallel approach to perform real-time CGH computation. This means that each x location on the hologram was assigned to one of 32K virtual processors. (The 16K physical processors were internally programmed to imitate 32K *virtual* processors.) A Sun 4 workstation was used as a front end for the CM2, and the parallel data programming language C Paris was used to implement holographic computation.

Table 1 contains the computation times (time per image point) using different approaches. "$I_{TOTAL}(complex)$" The common traditional case was "$I_{TOTAL}(complex)$" where Eqs. 1, 2, 3 and 4 are combined to compute I_{TOTAL}. The "Bipolar intensity" approach was simply the computation of I_B instead of I_{TOTAL}. The "Look-up table" method employed the precomputed elemental fringe table, represented in the memory of the CM2 as 2 bits per sample. As each object point was read from an input file, the position (x_p, z_p) was used to index the table in each processor. At each x, conditional logic was performed on each of the 2 bits, and the appropriate fractions of the object point amplitude were accumulated into the register representing I_B. Since this was performed in parallel for all 32K samples of $I_{B(x)}$, rapid computation of images was possible. The time per point listed in Table 1 is the amount of computation time required (on average) to accumulate the fringe pattern contribution of a single object point source. These numbers were obtained by computing holograms of several different test images of varying complexity. For comparison, the different computational approaches were also implemented on a serial computer, a Sun 4 workstation, and the results are included in Table 1.

Computation method	CM2	Sun4
I_{TOTAL} (complex)	2.180 ms	943.4 ms
Bipolar intensity	1.135 ms	486.2 ms
Look-up table	0.084 ms	22.1 ms

In all cases, the computation time per point was simply the total execution time divided by the number of points processed. Computation time scaled linearly with image complexity; despite the different image complexities (from 100 to 50,000 points), the time per point quotient remained within a 2% range. Table 1 shows that, as expected, the bipolar approach was roughly twice as fast as the traditional complex method of computation. This was evident on both the parallel-data and serial machines. By using the look-up table, the CM2 speed increased by a factor of 26; the serial machine was faster by a larger factor of about 43. (The CM2 actually has an array of floating-point math accelerator chips that are no longer utilized in look-up table calculations.) The simpler serial machine gained more from the use of a look-up table.

For practical purposes, additional procedures that are independent of object complexity must be performed, including normalizing the computed holographic pattern and moving it into the frame-buffer. Therefore, a generally fixed overhead time must be added when expressing the total time to compute and output the holographic pattern. For a 6 megabyte holographic pattern, this time is approximately 400 milliseconds (or less) on the CM2. For example, the actual time to compute and display a 10,000-point image using the look-up table approach is $10000 \times 0.084\text{ms} + 400\text{ms} = 1.24$ s, or less.

In the current MIT display system, simple images generated from 3-D computer graphics databases contain only a few thousand points. Using the fastest look-up table computation method, images are computed at a rate of over 1 frame per second. (In many cases, computation time was actually less than the overhead time spent loading the computed CGH into the framebuffer!) To demonstrate interactivity, the viewer can turn any of several dials (interfaced to the computer) in order to translate the image in horizontal, vertical, and depth locations, to change its size, and to spin it along different axes of rotation. In addition, a simple drawing program was written in which the user can move a 3-D cursor to draw a 3-D image that can also be manipulated.

It must also be noted that the holographic patterns computed by these three approaches were equivalent, with the following exceptions. Use of the bipolar intensity eliminated object self-interference and dc terms, making the CGH brighter and less noisy than when using the complex I_{TOTAL} method. The look-up table approach resulted in a pattern that was identical to the bipolar intensity approach, with the addition of some quantization noise if a 2-bit table was used. However, for objects of sufficient complexity, this quantization noise was comparable to that of the more straightforward approaches, due to the quantized nature of the output frame-buffer device used in the system.

CONCLUSION

Experimental results demonstrated that a horizontal-parallax-only off-axis transmission hologram can be computed in under 1 s. Computation was fast enough to provide dynamic holographic images at interactive rates. The interactive display of holographic images added the additional sensations of depth and tangibility that accompany moving 3-D images.

The overall speedup demonstrated here is remarkable. CGH computation that traditionally would require several minutes or hours on a mainframe computer is reduced to 1 s. The look-up table approach, by eliminating the need for all mathematical functions other than simple addition and multiplication, is especially advantageous when only minimal computing power is available for CGH computation. Essentially, the table stores a large array of holographic basis functions, the weighted sum of which is sufficient to describe the CGH corresponding to any arbitrary image. Given adequate memory space to hold the precomputed elemental fringes, it is possible to design a dedicated CGH computer that requires no floating-point mathematics and uses only integer addition (and perhaps bit shifting for normalization purposes). Such a simple machine can be implemented in parallel, e.g., one computer per holo-line, to achieve real-time CGH computation without the need for an expensive supercomputer. It is interesting to note that the 2-bit table approach implemented on the CM2 computes a CGH using only two conditional additions per image point per sample, emphasizing the computational simplicity made possible by the use of a look-up table.

The look-up table is the simplest possible algorithm for computing CGH from the discrete-point object model. To achieve faster speeds, future work involves an approach in which a higher level of object description is processed during CGH computation. Also, to increase speeds on workstation computers, it may be useful to utilize certain specialized hardware such as computer graphics rendering engines or digital signal processing boards used for image coding and decoding.

Analytical simplification of the physical model of light interference made possible this increase in speed. These concepts can be applied to other types of holograms. The bipolar intensity method is applicable to all types, including full-parallax CGHs. The use of the bipolar intensity summation method, whether directly or through a look-up table, eliminates object self-interference noise, eliminates the need to adjust a reference beam ratio, and produces an optimally bright image by eliminating unnecessary dc intensity bias. The look-up table approach can be applied to full-parallax holograms, although by requiring both x and y indices, the precomputed table (data arrays of rank five, reduced to rank three using the method shown) would require enormous amounts of memory space. In the future, as computational power increases, the simplification of computation presented here will continue to provide speed-up for any CGH application.

ACKNOWLEDGMENTS

The author gratefully acknowledges the support of Stephen A. Benton and researchers in the MIT Spatial Imaging Group past and present, especially John Underkoffler, Hiroshi Yoshikawa (Nihon University, Japan), Pierre St. Hilaire, Wendy J. Plesniak, and Michael Halle. The author wishes also to acknowledge enlightening discussions with H. John Caulfield.

This research on computational electronic holography has been supported in part by the Defense Advanced Research Projects Agency (DARPA) through the Rome Air Development Center (RADC) of the Air Force System Command and the Naval Ordnance Station, Indian Head, Maryland, contract #N00174-91-C-0117, and by the TV of Tomorrow consortium at the MIT Media Laboratory.

References

[1] P. St. Hilaire, S. A. Benton, M. Lucente, M. L. Jepsen, J. Kollin, H. Yoshikawa and J. Underkoffler. "Electronic display system for computational holography". In Practical Holography IV*, Proceedings of the SPIE*, volume 1212-20, pages 174–182, Bellingham, WA, 1990.

[2] S. A. Benton. "Experiments in holographic video imaging". In P. Greguss, editor, Holography*, Proceedings of the SPIE*, volume IS#08, pages 247–267, Bellingham, WA, 1991.

[3] P. St. Hilaire, S. A. Benton, M. Lucente, J. Underkoffler and H. Yoshikawa. "Real-time holographic display: Improvements using a multichannel acousto-optic modulator and holographic optical elements". In Practical Holography V*, Proceedings of the SPIE*, volume 1461-37, pages 254–261, Bellingham, WA, 1991.

[4] W. J. Dallas. Topics in applied physics. In B. R. Frieden, editor, The Computer in Optical Research, volume Vol. 41, chapter 6: "Computer-Generated Holograms", pages 291–366. Springer-Verlag, New York, 1980.

[5] D. Leseberg and C. Frere. "Computer-generated holograms of 3-D objects composed of tilted planar segments". Applied Optics, 27(14):3020–3024, July 1988.

[6] D. Leseberg. "Computer-generated holograms: display using one-dimensional transforms". Journal of the Optical Society of America, 3(5):726–730, May 1986.

[7] D. Leseberg and O. Bryngdahl. "Computer-generated rainbow holograms". Applied Optics, 23(14):2441–2447, July 1984.

[8] P. Hariharan. Optical Holography: Principle, Techniques and Applications. Cambridge University Press, Cambridge, 1984.

[9] J. S. Underkoffler. "Toward Accurate Computation of Optically Reconstructed Holograms". Master's thesis, Massachusetts Institute of Technology, June 1991.

[10] A. M. Glass. "The photorefractive effect". Opt. Eng., 17(5):470–479, 1978.

[11] S. A. Benton. "Survey of holographic stereograms". In Processing and Display of Three-Dimensional Data*, Proceedings of the SPIE*, volume 367, pages 15–19, Bellingham, WA, 1983.