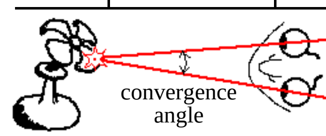
Chapter 1: The Window View Upon Reality

For centuries, our culture has speculated on the future of visual communication, and has imagined that, as a matter of course, the resulting images would be three dimensional—that they would accurately render sensations of depth, locations, and spatial relationships[1]. One can only imagine the collective sense of betrayal when conventional photography turned out to be flat! In only a few years, the public embraced stereoscopic photography, a feeble imitation of the glorious imaging expected from the inventors of their day. Since then, ever better methods for "perfect 3-D" have emerged from decade to decade, each promising more realistic and satisfying imaging than the last. Just when the ultimate limitations of traditional optical methods (such as lenticular photographs) seemed to be all too obvious, a completely new technique emerged in the early 1960's, one that promised an incredibly high quality of depth, detail, and tonal gradation; it was called "holography." Although it was invented in 1947 as a complex solution to a specific problem in electron microscopy, holography actually presented a solution to a fundamental question of wave recording and reconstructing—so fundamental that it eventually won the Nobel Prize in Physics for its inventor, Prof. Dennis Gabor (in 1971, after the advent of the laser had made the impact of holography visually obvious). Most of the applications of holography have been to technical problems since then, but in this course we will emphasize its most important promise for media technology, the three-dimensional "window view upon reality" that Gabriel Lippmann predicted (another Nobel Prize winner in Physics, and the inventor of a 3-D technique called "integral photography")[2].

Provoking Spatial Perceptions

Any discussion of three-dimensional images properly begins with a discussion of human vision, and the mechanisms by which we perceive spatial relationships, including shape, position, distance, and motion through space. These can be roughly grouped into three types, depending on whether they are stimulated by single-eyed (monocular) vision or properly-combined two-eyed (binocular) vision, and by whether they are stimulated by static or moving images (or perhaps the motion of the observer) in various combinations. A thorough discussion goes beyond the scope of this course, although we will revisit the topic in later discussions of the design of holographic images. Course MAS 853, *Spatial Imaging Systems*, explores these issues in more detail, as do several references (e.g., Okoshi[3], Patterson & Martin[4]).

	static	dynamic
monocular	overlap perspective focus, etc.	kinetic depth effect motion parallax
binocular	convergence fusion edge effects	



For our purposes, we will concentrate on the triangulation of point sources by binocular vision as the primary stimulus, or "cue," for spatial vision. Implicit in this are other cues arising from motion of one eye from side to side, which makes a kind of "temporal triangulation" possible, although the sliding of near objects over far objects also seems to be an important cue (time-varying occlusion correlated with observer motion). The eyes separately fixate on an image point (bringing its image onto the retina's *fovea*, the small area of its most acute vision), and the angle of convergence between the eyes is sensed via muscular proprioception. Combined with knowledge (derived from experience) of the inter-ocular or inter-pupillary distance, a fairly accurate estimate of the distance to a point can be generated. A mathematically equivalent approach is to say that the two eyes receive slightly differing 2-D views of a three-dimensional scene, which are fused to produce a single perception (without double vision in most cases), and that it is the "effort of fusion" that produces the impression of distance. As important as these binocular cues are, they are readily outvoted by simple monocular cues, especially by overlap (a.k.a. occlusion, opacity) cues. That one object's image terminates at the boundary of another is very convincing evidence that it is behind the other, and being hidden by it, in spite of possibly conflicting binocular cues. We will see this for ourselves in the study of *pseudoscopic* holographic images to come!

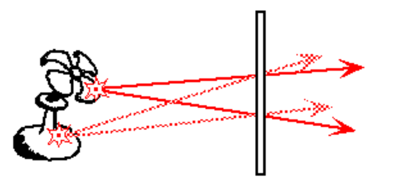
This simplified view makes it possible to say that it is only necessary to reproduce the <u>directions</u> in which light is traveling in order to produce a three-dimensional image. And it is this capability of holography that distinguishes it from other forms of photography. Of course, it must also provide the other depth cues, such as surface shading and occlusion, but those will follow naturally. First we will concentrate on the directions of light waves reaching the eyes through different parts of the hologram.

Optical Information

What do we mean by the "direction of the light" and its reproduction? Well, what do we mean by "light" at all? It has often been said that holography is photography in light so "coherent" (laser light, for example) that it becomes useful to describe it as a *wave* phenomenon. But we are much more familiar with the drawing of *rays*, which are the imaginary trajectories of imaginary particles (photons) traveling through the air. If particles of light, or their corresponding rays, are emitted by a point source of light and reach the pupil of an eye, that eye must rotate so that its optical axis is aligned with the ray in order to focus the light onto the *fovea*, the eye's small region of maximum acuity. Which is to say that an eye's optical axis, or "line of sight," must be rotated to pass through the point source in order to see it clearly. If we think of the point source instead as emitting spherical *waves* of light, the eye must still rotate so that the center of the lens is perpendicular to the wavefront so as to focus the wave on the fovea, which is simply to say again that the "line of sight" must pass through the point source for good vision. So the task is independent of whether we consider light to be represented by rays or by waves; even so, we will worry quite a bit about what representation to use.

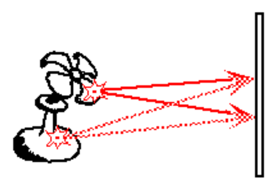
<u>Light as waves and rays</u>

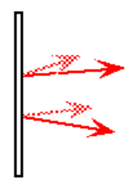
By the time you have gotten this far at MIT, you have probably heard several times of the particle/wave duality properties of matter and radiation. Sometimes, light behaves like a stream of particles; sometimes it behaves like a collection of waves. In fact, it is neither. We are like blind people feeling an elephant for the first time: what we think depends on where we grab it, and we may never quite grasp the entire concept. "Nature is not only stranger than we think, it is stranger than we <u>can</u> think!" (paraphrasing J.B.S. Haldane). Light is neither particles nor waves, and quantum mechanics has proposed a hybrid probabilistic model that is being argued even today. For all of the purposes of this course, it will suffice to adopt a simple wave model of light (that is, we will use a "classical" analysis). It will also suffice in most cases to represent these waves by their perpendiculars or normals at the areas of interest. These normals look a lot like rays! And they should, because the energy of a wave flows perpendicularly to the wavefront (in all but some crystalline materials). Thus we can use ray-like drawings, which are convenient, as long as we understand that we are talking indirectly about waves, or at least the directions of the wavefronts! And it is the directions of the wavefronts received by our two eyes that are compared to give rise to an impression of distance, so these graphical "rays" are enough (for now, at least—we will elaborate on this question in subsequent chapters, especially Ch. 8).



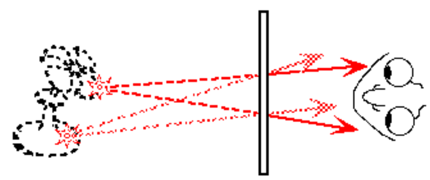
Capturing the directions of rays

We can now consider the basic problem of three-dimensional imaging to be the recording and reproduction of the directions of the light rays that strike some surface between the scene and the viewer. If we can reproduce the directions and relative strengths of all the rays accurately, then looking at this magical surface should be like looking through a window: we should see a three-dimensional image of the scene floating behind it with perfect realism, just as it would have looked if we saw the scene itself. We have created a "window with a memory."





A few other things become clear at this point, by the way. The image is NOT floating in thin air—we can see it only if we look <u>through</u> the window, and not if we look <u>around</u> it. The world's best known "hologram," the Princess Leia projection from *Star Wars*, is pure science fiction and Hollywood special effects: there are no known physical processes that could produce such an image from a projector off to one side—there has to be some optical element in the line



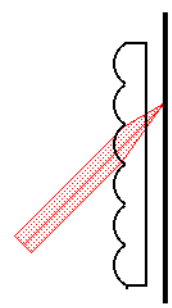
of sight, somewhere. Of course, George Lucas produced this effect by double exposure, but it has come to represent what most people mean by "hologram" (as in "Look out, he's got a *hologram*!" in *Total Recall*, etc.). Likewise, the Haunted Mansion at Disney World and Disneyland employs no true holograms, but a combination of magician's tricks that have been known for almost a century, especially "Pepper's Ghost" in the ballroom scene. It is important to remember that there are really two definitions of holography in our culture: "wavefront recording and reconstruction by interference and diffraction" (the technical field we are about to study) and "the psychologically ultimate three-dimensional imaging medium of the future" (what most people think we are working on at MIT).

Back to reality: the problem with our proposed ray-direction recording and playback scheme is that there is no known material that is sensitive to the direction of light—only its energy (or wave amplitude) matters, which triggers an individual micro-crystal or molecule, can be sensed. This is not to say that no such material could ever exist; we just can't imagine one at the moment. We know that a pane of ordinary glass briefly "traps" light passing through it, and releases it very shortly afterward, which accounts for the delay in propagation that we ascribe to its *index of refraction*. At least we might someday hope for a time-delay window with delay times measured in hours instead of attoseconds!

In the meantime, optical inventors have come up with a succession of techniques for approximating the variation of ray direction between the two eyes, starting with Wheatstone's stereoscope in 1838. Stereoscopes sample and reproduce the ray direction variation very coarsely—only twice! Most users prefer 3-D technologies that do not require them to use viewing aids, such as stereoscopes or spectacles; this has given rise to the class of *autostereoscopic* displays, of which holography is the most recent and the most spectacularly realistic.

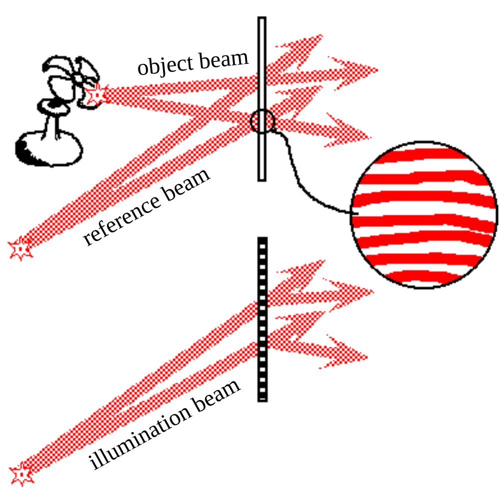
<u>classical optical techniques</u>

This is not the place for a detailed catalog of viewer-aided and autostereoscopic display technologies—Prof. Okoshi's book offers a fairly complete account of that history. The technology that comes closest to anticipating the visual impact of holography is Prof. Lippmann's integral photography, which places an array of small spherical lenses in front of a photographic film layer, the so-called "fly's eye lens array." The smaller the lenslets, the finer the sampling of the variation of light ray direction becomes, but the less accurate the reproduction of that direction, due to diffraction by the small diameter of the lenslets. Lippmann's proposal had some problems: as he first described the method, it produces an image of reversed depth (pseudoscopic); this was overlooked as no experimental tests of the technique were undertaken for several years. In the 1950's, Roger de Montebello perfected a second-generation technique that corrected several of these problems, but he also found severe limits on the image depth that could be provided without blurring.



<u>holographic direction recording</u>

Holography typically uses conventional photographic recording materials, ultra-fine grained versions of the same silver-halide emulsions that we use for black-and-white photography (the volume of the grains is about 1/30,000 of the usual, producing an equivalent ASA rating of about 0.001!). Which is to say that these materials are not sensitive to light direction either—holography records the direction information only indirectly. A second spread-out beam of light also exposes the film, overlapping the first at a carefully pre-arranged angle. That second, or *reference*, beam has to be coherent with the information, or *object*, beam; it has to have the same frequency, and be locked in phase with the object beam. In practice that means that they both have to come from the same laser (ordinary light is nowhere near coherent enough). Where they overlap, a characteristic "picket fence"-like interference pattern is formed which is imprinted on the film. The larger the angular difference between the beams, the finer the pattern becomes (it is very fine indeed, usually more than one thousand dark and light line pairs per millimeter). A 3-D scene consists of many points at different locations, and their waves



impinge on the film at different angles; each of these produces its own interference pattern, creating superimposed picket-fence patterns of different rotations and spacings.

Later, when the exposed and developed film (now the *hologram*) is illuminated with laser light at the same (reference beam) angle, the picket-fence-like pattern diffracts some of the light, with finer patterns deflecting it through greater angles. If everything works out as expected, the diffracted angle will equal the object beam angle, and we will have reconstructed the direction of the object beam at that point. It goes far beyond that simple fact, though. What emerges from the hologram is a perfect replica of the entire wave reflected by the object (plus some other waves). A viewer looking at the hologram does indeed see that "window view" of a 3-D image of the object, just as it looked during the exposure!

Of course we have to prove all these assertions, and wrestle with the limitations on their validity—that is what the rest of the course is about! And we have to understand how we make holograms that we can view in ordinary white light, which is when some of this starts taking on practical utility! But this should give you a general sense of what we are trying to do, and how.

Origins of Holography

Dennis Gabor was a German-trained electrical engineer, born in Budapest, Hungary, and interned in England during World War II. While there, he worked on a three-dimensional movie projection system in London, and later on electron microscope imaging for the British Thomson-Houston company in Rugby, England. The magnetic lenses of electron microscopes are imperfect for fundamental reasons—they distort the shape of the spherical electron waves coming from point-like objects. Gabor hoped to record that wave shape in the electron microscope, and then correct it with optical waves created by specially-ground lenses, but to do this he had to be able to record wavefront <u>shape</u> as well as amplitude/intensity, the wave's <u>phase</u> or <u>local direction</u> in our terms. People had been struggling with this problem for years, and it was considered unsolvable until a key idea came to Gabor while he was waiting for a tennis court one Sunday afternoon. When Gabor published his two-beam recording method in 1948, it was dismissed by most "experts" until they took a close look at his example photographs—something obviously worked! But the requirements that the object and reference beam be coherent limited Gabor's "holography" (inspired by the Greek for "whole" and "message," *holos* and *graphos*) to very small objects. Gabor had not even thought about holography as a three-dimensional imaging technology until he saw the results at the University of Michigan in the early 1960's.

Emmett Leith and Juris Upatnieks were electrical engineers at the University of Michigan's Willow Run Laboratories, near Ann Arbor. During the 1950's, they were working on a highly secret radar technique that allowed images of nearly photographic resolution to be generated by combining data from along a long flight path—the Project Michigan side-looking radar system. The key to the technique was an optical image-processing system that illuminated a long strip of radar data film with light from a mercury arc, focused it through a series of exotic lenses, and produced an incredibly detailed image. Slowly, Leith realized that he had rediscovered Gabor's concepts of holography, but in a much more general context. In 1962, low-power helium-neon lasers began to become commercially available, and Willow Run was one of the first labs to have one to experiment with. After verifying its usefulness for the side-looking radar project, Leith and Upatnieks started extending their ideas to the recording of three-dimensional table-top scenes. First they studied back-lit scenes, and by 1964 they had made holograms of front-lit objects—most notably a brass model of a steam locomotive that one of the machinists at the lab had made. They showed these holograms at the Fall 1964 meeting of the Optical Society of America, where a long line of scientists waited patiently in the hallway to glimpse this amazing sight. This triggered the long and tumultuous history of holographic imaging, which Leith and Upatnieks dubbed "wavefront reconstruction photography." Many artifacts from these early stages of holography research are now at the MIT Museum, joining the collection of the Museum of Holography that is housed there.

application areas

Although it is three-dimensional imaging that jumps into most people's minds when you say "holograms," the fact is that most of the applications of holography have been in other fields. Three-dimensional photography has been the beautiful "love child" of holography (until quite recently, that is), while other applications did the work and earned the money that kept most of the research going. To simplify things, it is useful to categorize the applications of holography into five groups:

1. holographic optical elements (HOEs)

Holograms can deflect and focus light just as prisms, lenses, and mirrors do (for one color at a time, anyway), but they are much lighter and more compact, and usually cheaper to make. Some folks call them "diffractive optical elements," which may be more accurate. Suffice to say that wherever laser light is used, a HOE is now a serious candidate to replace a conventional optical element, such as in supermarket scanners, CD players, aircraft head-up displays, and so forth. More recent work on "binary optics" (pioneered at MIT's Lincoln Labs) has shown how to fabricate these optical elements with VLSI technology.

2. optical computing

There is a small but devoted cult within the computer science community that believes that photons will someday replace electrons for high-speed highly-parallelized processing of data. There are already a few installations where this is beginning to come true. Within that domain, there are several tasks that holograms can do with some unique attributes. Because the thickness of a recording material can be accessed in a particularly efficient way by holographic readout, very high storage densities can be reached (around a gigabyte per square inch of film, for example, or 10^{12} bits in a cubic centimeter of crystal). Also, holographic storage holds the promise of *associative addressing*: illuminating the hologram with a small part of an image that it has seen before can produce a weak image of the rest of the image! A high-volume associative memory (or content-addressable memory) would have important uses in artificial intelligence computing, for example.

3. optical metrology & microscopy

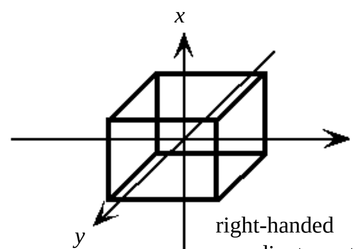
Because a hologram can produce an incredibly accurate replica of a wavefront recorded at another place at another time, the images it produces can be measured with great precision. A room-sized nuclear containment vessel can be recorded in a laser flash, for example, and its image then examined at leisure at a distant and non-hazardous laboratory for cracked metal parts, corrosion, and so forth. The nuclear physics team at MIT built a holographic recording system for a new giant-size bubble chamber (3 meters deep) used in the search for the tau lepton. In ordinary photography, the higher the resolution that is needed, the shallower the depth of field that can be focused. Holography eliminates that tradeoff, allowing 30 micron bubbles to be tracked throughout the depth of the chamber.

4. non-destructive testing (NDT)

Likewise, two optical wavefronts can be compared with high accuracy, even though they were recorded or observed at very different times, and with the object under very different conditions. Because the phase of the wavefront changes very rapidly with very small object motions, the interference pattern formed between two holographic recordings of a scene are very sensitive to small changes. Only five millionths of an inch of object motion will change its image from light to dark—this can be caused by mechanical stress, or by the effect of a defect hidden deep in the structure of the object. Most aircraft tires are retreaded many times, and for many years all these recaps were required to be checked by holographic interferometry (holographic non-destructive testing), which was the only sector of holography making any money at the time! Our first lab demo will show interference between an object (a coffee cup) and its holographic image, so that you can check the sensitivity yourself, and we will have a special lecture on other aspects of this topic.

5. three-dimensional imaging (display holography)

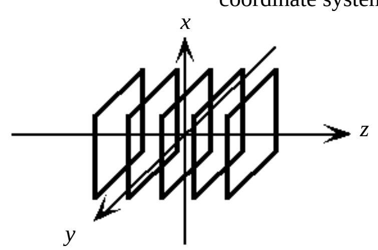
In spite of being concerned mainly with “pretty pictures,” display holography has had a major impact on all the fields mentioned above. Ultimately, they all have similar concerns about making bright and clear holograms, but display holographers attacked these problems first, and in peculiarly inventive and unorthodox ways (most of them didn’t know any better!). Their improvements in manufactured materials and processing chemistry and techniques were taken up by the industrial labs with some reluctance, but worked so well that this “trickle-up technology” has become an important part of the field as a whole. But with their focus on holographic imaging for many reasons (fine art, museum display, security devices, advertising, portraiture, and so forth), the display holography community still seems somewhat separated from the other sectors of the field. Since the development of mass production techniques for white-light holograms, a whole new set of technologies have come into the mixture, and the field is changing rapidly these days. At MIT, our research emphasis is on making synthetic holograms of computational “objects,” to make better understanding of their spatial organization possible, in spite of their complexity. Early application areas include computer-aided design, medical imaging, and scientific visualization, as you will soon see.

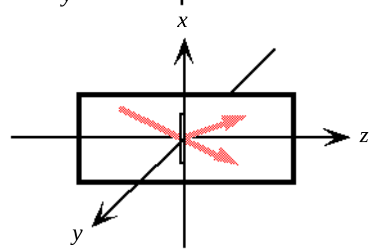


right-handed coordinate system

Styles of Analysis

Just as people use holography for many different purposes, they use many different styles of analysis to understand and control the technique. Physicists tend to use three-dimensional analyses based on Green’s functions, which can be hard to visualize, and don’t hook into optical design thinking very well. Electrical engineers have made many contributions to the field by looking at the volume as a series of flat and parallel planes. The light “signals” on one plane are related to those on another by fairly simple (for them, anyway) integral transforms, or convolutions of impulse functions. For our purposes, it is much easier to concentrate on just one two-dimensional surface, the *x-z* plane, perpendicular to the hologram plane. The sources and rays of interest will be restricted to this plane (mostly) and light will travel in the +*z* direction (mostly). We will find that limiting ourselves to a single plane is what makes “shop math” (algebra, plane geometry, trigonometry) really useful. Things that we learn by limiting the geometry to the *x-z* plane will cultivate many practical insights that can be generalized later on, if we feel so inclined. Actually, we will have to let the rays travel a little ways out of the *x-z* plane to discuss focusing properly, especially to talk about *astigmatism* (forewarned is forearmed!). Those of you who have already had some electrodynamics may well be skeptical of such a simplified approach, but we have many optical components to fold into our story, and I predict that you will be grateful for this point of view. And we will show you how to generalize the approach to the full *x-y-z* space before we are done, I promise. We will also eschew the delights of integral and differential calculus in all but a few cases. This makes some of the proofs and demonstrations a few equations longer than they might otherwise be, but helps us concentrate on the physical phenomena involved instead of the mathematics of the analysis. These two features distinguish this course’s approach from that of any known textbook, but you will be able to find corroborating evidence in a variety of reference volumes, once you see the correspondence between our notation systems.





References:

1. Two of my favorite examples from the “Gulliver’s Travels” school of early science fiction are:

from the *Fables of Fénelon*

Fénelon, Françoise de Salignac de la Mothe (1651–1715; this piece is probably from around 1699).

“Water was placed in great basins of silver or gold, and the object to be painted was placed in front of that basin. After a while the water froze and became a

glass mirror, on which an ineffaceable image remained."
(of course, like a mirror image, it was three dimensional! SAB)
from *Giphantie,*
Tiphaigne de la Roche (Paris, 1760).
The chief of a remote African tribe takes Giphantie into his home, where the sea can be seen through a window. Giphantie, amazed (so far from the shoreline), rushes to the window and bumps his head on something. He reports:
"That window, that vast horizon, those black clouds, that raging sea, all were but a picture…" (again, obviously three dimensional! SAB)
He goes on to describe the picture-making process:
"The elemental spirits have composed a subtle matter, very viscous and quick to dry, by means of which a picture is formed in the twinkling of an eye. They coat a piece of canvas with this material and hold in front of the object that they wish to paint. It is then carried away to some dark place. An hour later, the impression is dry, and you have a picture. The correctness of the drawing, the truth of the expression, the stronger or weaker strokes, the gradation of the shades, the rules of perspective, all this we leave to nature, who with a sure and never-erring hand, draws upon our canvases which deceive the eye."
(change a few words and it sounds a lot like holography itself! SAB)

2. Gabriel Lippmann, "Épreuves réversibles. Photographies intégrales," *Comptes Rendus*, **146**, pp. 446–451 (March 3, 1908).
3. Takanori Okoshi, *Three-Dimensional Imaging Techniques*, (Academic Press, NY, 1972).
4. Robert Patterson and Wayne L. Martin, "Human Stereopsis," Human Factors, **34(6)**, pp. 669–692 (1992).