
DIGITAL HOLOGRAPHY AND THREE-DIMENSIONAL DISPLAY

DIGITAL HOLOGRAPHY AND THREE-DIMENSIONAL DISPLAY

Principles and Applications

Edited by

TING-CHUNG POON

Optical Image Processing Laboratory, Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, Virginia 24061.



Springer

Editor:

Dr. Ting-Chung Poon

Virginia Tech

Bradley Dept. Electrical and Computer Engineering

641 Whittemore Hall

Blacksburg VA 24061

Digital Holography and Three-Dimensional Display: Principles and Applications

Consulting Editor:

D.R. Vij

Library of Congress Control Number: 2005938666

ISBN 0-387-31340-0 e-ISBN 0-387-31397-4

ISBN 978-0-387-31340-5

Printed on acid-free paper.

© 2006 Springer Science+Business Media, Inc.

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, Inc., 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1

springer.com

Dedication

*This book is dedicated to my
parents for their
encouragement and sacrifice
that make this effort possible*

Contents

Dedication	v
Contributing Authors	xii
Preface	xiii
PART 1: DIGITAL HOLOGRAPHY	
Chapter 1: COMPUTER-GENERATED HOLOGRAMS	1
Introduction	1
Mathematical Preliminaries	2
Holography	15
CGH Imperfections	26
The Third Dimension	32
CGH's Grouped by Propagation Region	36
Interferometric Display Holograms	45
Information Content of Scenes	47
Chapter 2: DIGITAL HOLOGRAPHY AND MULTI-WAVELENGTH INTERFERENCE TECHNIQUES	51
Introduction	51
Digital Holography Principles	53
Digital Holographic Microscopy Experiments	58
Multi-Wavelength Phase Imaging Digital Holography	64
Wavelength-Scanning Digital Interference Holography	66

Conclusions	69
Chapter 3: DIGITAL HOLOGRAPHY FOR BULK IMAGE ACOUSTOOPTICAL RECONSTRUCTION	73
Introduction	73
Main Assumptions	74
System Architecture	75
Computer Simulations	77
Modeling Results	88
Color Image Processing	88
Conclusion	91
Chapter 4: DIGITAL HOLOGRAPHY APPLIED TO MICROSCOPY	95
Introduction	96
Diffraction Theory	100
Hologram Formation and Wavefront Reconstruction	108
Reconstruction Algorithms: Numerical Aspects	118
Instruments	123
Applications	130
Conclusions and Perspectives	136
Chapter 5: PHASE-SHIFTING DIGITAL HOLOGRAPHY	145
Introduction	145
Image Formation	147
Measurement of Surface Shape and Deformation	155
Conclusions	168
Chapter 6: CONTROLLING IMAGE RECONSTRUCTION PROCESS IN DIGITAL HOLOGRAPHY	173
Introduction	174
Compensating Aberrations in Digital Holography	175
Controlling Numerical Reconstructions	183
Controlling Reconstructions in MWDH Compensating Chromatic Aberrations	193
Application of DH for Investigation and Testing of MEMS Structures	198
Chapter 7: TELE-METROLOGY BASED ON DIGITAL HOLOGRAPHY	213
Introduction	213
Digital Holography and Related Methods	214
Distant Measurements by Digital Holography	217
Data Compression and Decompression	222

Conclusions	232
PART 2: THREE-DIMENSIONAL DISPLAY TECHNIQUES	
Chapter 8: COMPUTER-GENERATED HOLOGRAMS FOR WHITE LIGHT RECONSTRUCTION	235
Basics of Computer-Generated Holograms	235
Calculation Methods of Simple Holograms	236
Calculations of Holograms for White Light Reconstruction	238
Output Devices of Computer-Generated Holograms	251
Conclusions	253
Chapter 9: WIDE-ANGLE COMPUTER-GENERATED HOLOGRAMS FOR 3-D DISPLAY	257
Introduction	257
Hologram Synthesis by 3-D Fourier Spectrum	259
Hologram Synthesis in Polar Coordinates	265
Conclusion	270
Chapter 10: HORIZONTAL-PARALLAX-ONLY OPTICAL SCANNING HOLOGRAPHY	273
Introduction	273
Optical Scanning Holography	274
Electronic Horizontal-Parallax-Only Holography	282
Concluding Remarks	289
Chapter 11: THREE-DIMENSIONAL PROJECTION DISPLAY SYSTEM	293
Introduction	293
Stereoscopic Projection Displays	302
Time-Sequential Multiview Projection Displays	316
Projection Integral Imaging Displays	319
Fresnel Lens-Based Projection Displays	322
Holographic Projection Display	326
Chapter 12: THREE-DIMENSIONAL DISPLAY AND INFORMATION PROCESSING BASED ON INTEGRAL IMAGING	333
Introduction	333
Fundamentals of Integral Imaging	338
3D Display Based in Integral Imaging	347
Depth Extraction	359

3D Correlator	367
Summary	376
Chapter 13: AUTOSTEREOSCOPIC, PARTIAL PIXEL, SPATIALLY MULTIPLEXED, AND OTHER 3D DISPLAY TECHNOLOGIES	379
Introduction	379
Some Perspectives on 3D display methodologies	381
Spatially Multiplexed Pixelated 3D Message Displays	393
Post-1995 and Futuristic Perspectives	409
Index	427

Contributing Authors

William J. Dallas

University of Arizona

Tucson, Arizona, USA

e-mail: dallas@radiology.arizona.edu

Myung K. Kim, Lingfeng Yu, and Christopher J. Mann

University of South Florida

Tampa, Florida USA

e-mail: mkkim@cas.usf.edu

Vladimir Petrov

Saratov State University

Saratov, Russia

e-mail: petrovvv@sgu.ru

Christian Depeursinge

Swiss Federal Institute of Technology

Lausanne, Switzerland

e-mail: christian.depeursinge@epfl.ch

Ichiro Yamaguchi

Gunma University

Gunma, Japan

e-mail: ichiyama@el.gunma-u.ac.jp

Pietro Ferraro, Sergio De Nicola and Giuseppe Coppola

Istituto Nazionale di Ottica Applicata

Pozzuoli (Na), Italy

e-mail: p.ferraro@cib.na.cnr.it

Thomas Kreis¹, Werner Jüptner¹, Ulf Schnars²¹BIAS - Bremer Institut für Angewandte Strahltechnik²Airbus, Bremen, Germany

e-mail: jueptner@bias.de

Hiroshi Yoshikawa

Nihon University, Japan

e-mail: hiroshi@ecs.cst.nihon-u.ac.jp

Toyohiko Yatagai and Yusuke Sando

Tsukuba University, Japan

e-mail: yatagai@bk.tsukuba.ac.jp

Ting-Chung Poon

Virginia Tech

Blacksburg, Virginia, USA

e-mail: tcpoon@vt.edu

Eun-Soo Kim

Kwangwoon University

Seoul, Korea

e-mail: eskim@daisy.kwangwoon.ac.kr

Byoungho Lee¹, Jae-Hyeung Park¹, and Sung-Wook Min²¹Seoul National University,

Seoul, Korea

²Information and Communications University,

Seoul, Korea

e-mail: byoungho@snu.ac.kr

Monish R. Chatterjee¹ and Shih-Tun Chen²¹University of Dayton

Dayton, Ohio, USA

²Teradyne, Inc.

Agoura Hills, California 91301

e-mail: monish.Chatterjee@notes.udayton.edu

Preface

Digital (or electronic) holography and its application to 3-D display is one of the formidable problems of evolving areas of high technology that has been receiving great attention in recent years. Indeed, the “Holy Grail” for 3-D display is the realization of life-size interactive 3-D displays. Obviously, we are not there yet, but advances in 3-D display allow us to make important steps towards the Holy Grail. The theme of this book is to organize a collection of key chapters that covers digital holography in Part 1 (Chapters 1- 7) and 3-D display techniques in Part 2 (Chapters 8-13) so as to provide the reader with the state-of-the-art developments in these important areas around the world. The book contains a large amount of research material as well as reviews, new ideas and insights that will be useful for graduate students, scientists and engineers working in the field.

Chapter 1

COMPUTER-GENERATED HOLOGRAMS

William J. Dallas

Department of Radiology and College of Optical Science, University of Arizona, Tucson, AZ 85724

Abstract: Computer-generated holograms (CGH's) are diffractive-optical elements that offer the possibility of creating wave-optical display systems that are under complete computer control. Because CGH's can control all aspects of the optical wave, three-dimensional imagery can be presented to an audience. After reviewing some mathematical preliminaries that are essential to the description of CGH's, we move to the description of their construction and operation. We include descriptions of the effects that some common imperfections have on the image presented by a CGH. Although CGH's have been successfully applied in many areas including optical testing and security, we concentrate our attention on 3-D display. We will also examine some basic conditions on the use of CGH's for video display of these images.

Key words: computer-generated hologram, CGH, display hologram, holography, 3-D display, 3-D video

1. INTRODUCTION

We begin this chapter with some mathematical preliminaries. A few good reference books that contain Fourier mathematics are (Gaskill, 1978; Bracewell, 1999; Goodman, 1996). We often use the abbreviation "CGH" for Computer Generated Hologram. Two summary articles on the CGH can be found in (Lee, 1978 ; Bryngdahl, 1990). The Fourier transform is the mathematical tool at the foundation of the CGH. We will define the Fourier transform in one and two dimensions. Most of the CGH mathematics is in two dimensions; one-dimensional examples are used for simplicity. In the

body of the chapter, will also explore Fourier series, the discrete Fourier transform or DFT, and generalized harmonic analysis.

2. MATHEMATICAL PRELIMINARIES

2.1 Fourier transform

We abbreviate the Fourier transform by F.T. and by \mathcal{F} . The inverse Fourier transform is noted by I.F.T and by \mathcal{F}^{-1} . Note that the inverse Fourier transform is an expression of the original function as a "weighted sum" of exponentials with simple arguments. The phase of the exponential is linear in the position coordinate. The exponential is therefore called a linear phase factors. They are also called cisoids, the "cis" coming from the Euler relation for the complex exponential: "cosine + i sine". An example of such a factor is $e^{-2\pi i \xi_0 x}$, where ξ_0 is a constant. The Fourier decomposition is useful in optics because the exponential functions behave in a simple, well understood way as does the re-composition of the functions after they have been acted upon by an optical system.

$$\text{1-D Forward F.T.: } U(\xi) = \int_{-\infty}^{\infty} u(x) e^{-2\pi i \xi x} dx \quad (1-1)$$

$$\text{Inverse F.T. or I.F.T. } u(x) = \int_{-\infty}^{\infty} U(\xi) e^{+2\pi i \xi x} d\xi \quad (1-2)$$

The extension to two dimensions is straight forward. Notice that the exponential functions are separable. Using the following definitions, we can express the arguments as a scalar product of 2-dimensional vectors.

$$\text{2-D F.T. } U(\xi, \eta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x, y) e^{-2\pi i (\xi x + \eta y)} dx dy \quad (1-3)$$

$$\text{I.F.T. } u(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(\xi, \eta) e^{+2\pi i (\xi x + \eta y)} d\xi d\eta \quad (1-4)$$

Expressing the coordinates in vector form makes the expression of these relations somewhat more compact and allows for easy extension to higher dimensions. Vectors are bolded. Unit vectors have a hat. Let

$$\mathbf{r} = x \hat{\mathbf{x}} + y \hat{\mathbf{y}} ; \quad \mathbf{p} = \xi \hat{\mathbf{x}} + \eta \hat{\mathbf{y}} ; \quad \mathbf{p} \cdot \mathbf{r} = \xi x + \eta y \quad (1-5)$$

so that

$$\text{F.T.} \quad U(\rho) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(\mathbf{r}) e^{-2\pi i \rho \cdot \mathbf{r}} d^2 r \quad (1-6)$$

$$\text{I.F.T.} \quad u(\mathbf{r}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(\rho) e^{+2\pi i (\rho \cdot \mathbf{r})} d^2 \rho \quad (1-7)$$

2.2 Fourier series

The Fourier series expresses a periodic function as a sum of linear phase factors. The forward transformation for calculating the Fourier series coefficients and the inverse transformation, the Fourier series which reconstitutes the original function, is

$$\text{Coefficients: } U_{mn} = \Delta\xi\Delta\eta \int_{-\frac{I}{2\Delta\xi}}^{\frac{I}{2\Delta\xi}} \int_{-\frac{I}{2\Delta\eta}}^{\frac{I}{2\Delta\eta}} u(x, y) e^{-2\pi i (m\Delta\xi x + n\Delta\eta y)} dx dy \quad (1-8)$$

$$\text{Fourier Series: } u(x, y) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} U_{mn} e^{2\pi i (m\Delta\xi x + n\Delta\eta y)} \quad (1-9)$$

$$\text{The object is periodic, i.e., } u(x, y) = u\left(x - \frac{m}{\Delta\xi}, y - \frac{n}{\Delta\eta}\right). \quad (1-10)$$

2.3 Discrete Fourier transform and FFT

The correspondence of punctilicity and periodicity can be extended to periodic objects which consist of regular arrays of points. From the previous section, we see that the continuous Fourier transform of such objects will itself be periodic and consist of points. The relation between the point strengths of one period in the object and one period in the Fourier transform is given by the discrete Fourier transform:

$$\text{Forward DFT: } U_{jk} = \frac{1}{\sqrt{M}} \frac{1}{\sqrt{N}} \sum_{m=\frac{-M}{2}}^{\frac{M-1}{2}} \sum_{n=\frac{-N}{2}}^{\frac{N-1}{2}} u_{mn} e^{-2\pi i \left(\frac{jm}{M} + \frac{kn}{N}\right)} \quad (1-11)$$

$$\text{Inverse DFT: } u_{mn} = \frac{1}{\sqrt{M}} \frac{1}{\sqrt{N}} \sum_{j=-\frac{M}{2}}^{\frac{M}{2}-1} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} U_{jk} e^{2\pi i \left(\frac{jm+kn}{M+N} \right)} \quad (1-12)$$

We note that the DFT is a discrete-to-discrete transformation. There are several variations on defining the DFT. One variation is in the normalization factor another in the range covered by the indices. This index variability is of practical importance.

When a Fourier transform is done on a computer, many times people talk about performing a fast Fourier transform or FFT. The FFT is actually an algorithm for performing the discrete Fourier transform or DFT. The algorithm speeds computation of the DFT by a factor $n/\log_2 n$ that can be considerable for large transforms. We will use the term FFT to spotlight a difference in the location of the origin for the transform from our definition of the discrete Fourier transform.

Most libraries of computer algorithms define the indices going from zero to an upper limit whereas we define an almost symmetric index range about zero. These differences require a technique variously known as quadrant swapping or checker-boarding in order to bring consistency into calculations. In one dimension our definition of the DFT is:

$$U_j = \frac{1}{\sqrt{M}} \sum_{m=-\frac{M}{2}}^{\frac{M}{2}-1} u_m e^{-2\pi i \frac{jm}{M}} \quad (1-13)$$

$$\text{The "FFT" definition is: } V_j = \frac{1}{M} \sum_{n=0}^{M-1} v_n e^{-2\pi i \frac{n j}{M}}. \quad (1-14)$$

One can show that $\sqrt{M} (-1)^j V_j = U_j$. Upon performing the discrete Fourier transform, a striping or multiplication by alternating plus and minus ones, in the transform, occurs when the indices are shifted from our definition to the computer algorithm definition. In 2-dimensions, the striping becomes checker boarding. The checker-boarding, i.e. multiplication by the ± 1 checkerboard in the transform, is equivalent to exchanging quadrants in the object. The quadrants that are exchanged are the first with the third and the second with the fourth. The actual correspondence between the two operations is

! Post-transform checker-boarding = pre-transform quadrant-exchange

! Post-transform quadrant-exchange = pre-transform checker-boarding

The usual procedure and order of operations is:

$$\text{DFT} = \text{quadrant-exchange} \rightarrow \text{FFT} \rightarrow \text{quadrant-exchange}.$$

2.4 Generalized harmonic analysis

Generalized harmonic analysis is an intimidating name for a simple procedure that is crucial in explaining certain types of CGH's. In essence, we first expand a function as a Fourier series. We then substitute a function for a variable in the argument. We restrict our attention to one dimension for simplicity and begin with the usual 1-D Fourier series. Consider the Fourier series

$$u(x) = \sum_{m=-\infty}^{\infty} C_m e^{2\pi i m \xi_0 x} \quad (1-15)$$

At this point the key substitution is made and it is important to realize that the Fourier series relation is a point relation, i.e., the "x" in the exponential corresponds point-wise to the "x" in the argument of the function. Therefore, we can substitute for that "x", a function "f of x", giving the expansion

$$u[f(x)] = \sum_{m=-\infty}^{\infty} U_m e^{2\pi i m \xi_0 f(x)} \quad (1-16)$$

2.5 The Dirac delta function

The simple definition is that the Dirac delta function, denoted by $\delta(x)$, has the properties:

$$\delta(0) = \infty, \delta(x_0) = 0 \text{ for } x_0 \neq 0, \int_{-\infty}^{\infty} \delta(x) dx = 1 \quad (1-17)$$

A more careful consideration of its properties comes through applying a family of functions whose properties approach those desired of the Dirac delta function in a limit. One such family is that of normalized Gaussian functions

$$\frac{1}{\sigma \sqrt{2\pi}} e^{\frac{-x^2}{2\sigma^2}} \quad (1-18)$$

The normalization is such that the integral

$$\int_{-\infty}^{\infty} \frac{1}{\sigma \sqrt{2\pi}} e^{\frac{-x^2}{2\sigma^2}} dx = 1 \quad (1-19)$$

for all values of the variance σ . We can think of the delta function as a very tall and narrow Gaussian,

$$\delta(x) = \lim_{\sigma \rightarrow 0} \frac{1}{\sigma \sqrt{2\pi}} e^{\frac{-x^2}{2\sigma^2}} \quad (1-20)$$

2.6 The Comb function

There are several applications that can be simply explained using the delta comb defined as

$$\text{comb}(x) = \sum_{m=-\infty}^{\infty} \delta(x-m) \quad (1-21)$$

It will help to define “comb math” here. This abstraction is useful for dealing with periodic and point-form , i.e. punctile, objects in continuous Fourier transforms. The relations will allow us to write down, in compact form, descriptions that we will later expand upon. It is also straight forward to expand the relations from compact form into their full form.

The 2-D comb function is

$$\begin{aligned} \text{comb}(x, y) &= \text{comb}(x) \text{comb}(y) \\ &= \sum_{m=-\infty}^{\infty} \delta(x-m) \sum_{n=-\infty}^{\infty} \delta(y-n) \end{aligned} \quad (1-22)$$

A comb with unit-spaced “teeth” transforms into itself,

$$F.T. [\text{comb}(x)] = \text{comb}(\xi) \quad (1-23)$$

One important Fourier relation is that a shift transforms into multiplication by a linear phase factor

$$F.T. [f(x \pm x_0)] e^{\pm 2\pi i x_0 \xi} F(\xi) \quad (1-24)$$

this shift theorem gives another useful form for the transform of the comb.

$$F.T. [\text{comb}(x)] = \sum_{m=-\infty}^{\infty} e^{\pm 2\pi i m \xi} = \text{comb}(\xi) \quad (1-25)$$

Another important Fourier relation is scaling: magnification transforms into minification

$$F.T. \left[f\left(\frac{x}{b}\right) \right] = |b| F(b\xi) \quad (1-26)$$

Shifting and scaling the comb function gives us,

$$\text{comb}\left(\frac{x-x_0}{b}\right) = |b| \sum_{n=-\infty}^{\infty} \delta(x - x_0 - nb) \quad (1-27)$$

Using the previous relations we have that

$$F.T. \left[\text{comb}\left(\frac{x-x_0}{b}\right) \right] = |b| \text{comb}(b\xi) e^{-2\pi i x_0 \xi} \quad (1-28)$$

Making a continuous function punctile one is done by the multiplication:

$$\begin{aligned} f(x) \text{comb}(x) &= f(x) \sum_{m=-\infty}^{\infty} \delta(x - m) \\ &= \sum_{m=-\infty}^{\infty} f(x) \delta(x - m) = \sum_{m=-\infty}^{\infty} f_m \delta(x - m) \end{aligned} \quad (1-29)$$

Periodic continuation is usually described for functions of bounded support. It is important that we understand the more general case covering functions of large extent or even unbounded support. The reason for this consideration is the fact that a function cannot be of bounded support both in the object and in the Fourier domains. We will concentrate on objects of bounded support. These functions necessarily have Fourier transforms of unbounded support. Periodically continuing a function is done by convolving with a comb.

$$\begin{aligned} f(x) * \text{comb}(x) &= f(x) * \sum_{m=-\infty}^{\infty} \delta(x-m) \\ &= \sum_{m=-\infty}^{\infty} f(x) * \delta(x-m) = \sum_{m=-\infty}^{\infty} f(x-m) \end{aligned} \quad (1-30)$$

We see that nowhere in this definition is bounded support required. We should be aware of the fact that contributions from each repetition of the function will contribute to the value at each point. This fact is, however, for us not important.

The Fourier transforms of the relations are

$$F.T.[f(x) * \text{comb}(x)] = F(\xi) \text{comb}(\xi) \quad (1-31)$$

and

$$F.T.[f(x) \text{comb}(x)] = F(\xi) * \text{comb}(\xi) \quad (1-32)$$

2.7 Sampling

The Whittaker-Shannon sampling theorem is generally stated for an object which has a band-limited Fourier spectrum. For this case the object can be reconstructed from samples of that object spaced by one over the spectral bandwidth. For computer generated holography, it is more useful to turn this theorem around and consider the sampling to be in Fourier space and the bounded support to be of the object. This form of the sampling theorem becomes

$$\begin{aligned}
 U(\xi) &= \sum_{n=-\infty}^{\infty} U\left(\frac{n}{\Delta x}\right) \text{sinc}\left[\left(\xi - \frac{n}{\Delta x}\right)\Delta x\right] \\
 &= \sum_{n=-\infty}^{\infty} U\left(\frac{n}{\Delta x}\right) \text{sinc}(\xi\Delta x - n)
 \end{aligned} \tag{1-33}$$

$$\text{for } u(x) = 0 \text{ if } |x| > \frac{\Delta x}{2}.$$

The basis of the theorem is replication and truncation. In 1-D the truncation function is uniquely a rect. There is more freedom in 2-D.

The proof of the theorem is straightforward using comb math:

$$\begin{aligned}
 F &\left\{ \left[\text{rect}\left(\frac{x}{\Delta\xi}\right) \right] \left[u(x) * \frac{\text{comb}\left(\frac{x}{\Delta x}\right)}{\Delta x} \right] \right\} \\
 &= [\text{sinc}(\Delta x)] * [U(\xi)\Delta x \text{comb}(\Delta x \xi)]
 \end{aligned} \tag{1-34}$$

Recall that the factor $|\Delta x|$ results from the tooth-thickness of the comb. The end result is

$$U(\xi) = \sum_{n=-\infty}^{\infty} U\left(\frac{n}{\Delta x}\right) \text{sinc}(\xi\Delta x - n) \tag{1-35}$$

There are several variations of this theorem that are useful. They relate to sampling at shifted points, over-sampling, under-sampling, non-rectangular apertures, and sampling on non-rectangular lattices

The Nyquist sampling interval is the minimum separation of samples in the Fourier plane that allows a perfect reconstruction of the image. For the sampling theorem, this rate is $1/\Delta x$.

2.8 Carrier modulation

Generalized harmonic analysis allows us to apply Fourier series to the analysis of non-linear processes. We will look at instances where the

generalized harmonics have arguments consisting of a simple carrier and a modulation term. We will illustrate the principle of carrier modulation by two examples. First, we begin with a simple cosine carrier and then move to the more complicated square-wave carrier. In both cases, we will see that the resulting complex amplitude transmittances can be expressed in terms of a superposition of images. For most applications we impress one desired image on the carrier. This modulation image is termed the true image. As a consequence of the modulation, other images are generated. These undesired images are termed false images.

2.8.1 Cosine-Carrier Modulation

We look at the encoding of complex information on a real cosine carrier. We begin by Hermitian symmetrizing the object about the origin. The object is

$$u(x) = 0 \text{ for } |x| > \frac{\Delta x}{2} \quad (1-36)$$

and the Hermitian symmetric version is

$$u_h(x) = \frac{1}{2} \left[u\left(x + \frac{\Delta x}{2}\right) + u^*\left(-x + \frac{\Delta x}{2}\right) \right] \quad (1-37)$$

We then Fourier transform and find that the symmetrization has led to displacements which transform into linear phase factors.

$$U_h(\xi) = \frac{1}{2} \left[U(\xi) e^{+2\pi i \frac{\Delta x \xi}{2}} + U^*(\xi) e^{-2\pi i \frac{\Delta x \xi}{2}} \right] \quad (1-38)$$

Converting the object Fourier transform into polar form,

$$U(\xi) = A(\xi) e^{i\phi(\xi)} \quad (1-39)$$

We can write the total transform as a modulation in amplitude and phase of a cosine carrier.

$$U_h(\xi) = \frac{1}{2} \left\{ A(\xi) e^{2\pi i \left[\frac{\Delta x \xi}{2} + \frac{\phi(\xi)}{2\pi} \right]} + A(\xi) e^{-2\pi i \left[\frac{\Delta x \xi}{2} + \frac{\phi(\xi)}{2\pi} \right]} \right\} \\ = \cos \left\{ 2\pi \left[\frac{\Delta x \xi}{2} + \frac{\phi(\xi)}{2\pi} \right] \right\} \quad (1-40)$$

We also see that the end result is equivalent to simply taking the original object a distance $\frac{\Delta x}{2}$ off axis, Fourier transforming and then taking the real part of the Fourier transform. For the practical generation of cosine CGH's, this procedure has an advantage in simplicity over doing the actual symmetrization.

2.8.2 Square-Wave Carrier Modulation

The basic formula is the Fourier series for a square-wave of duty cycle a and unit period,

$$Sq(a, x) = \sum_{n=-\infty}^{\infty} a \operatorname{sinc}(na) e^{2\pi i n x} \quad (1-41)$$

In order to pave the way for some meaningful substitutions, instead of x , we will use the variable p so that

$$Sq(a, p) = \sum_{n=-\infty}^{\infty} a \operatorname{sinc}(na) e^{2\pi i np} \quad (1-42)$$

In pulse-code modulation (PCM), a square-wave replaces the cosine as the carrier in the calculations. Impressing the modulation on the carrier is done by generalized harmonic analysis, i.e., first expanding the square-wave as a Fourier series and then substituting more complicated functions for the simple arguments.

The more complicated arguments are

$$a(\xi) = \frac{1}{\pi} \arcsin [A(\xi)] ; \quad p(\xi) = \left[\Delta x \xi + \frac{\phi(\xi)}{2\pi} \right] \quad (1-43)$$

The leading factor of $\frac{1}{\pi}$ is inserted so that when $A(\xi) = 1$ the duty cycle is

$$a(\xi) = \frac{1}{\pi} \arcsin[1] = \frac{1}{\pi} \frac{\pi}{2} = \frac{1}{2}.$$

The substitutions result in

$$\begin{aligned} V(\xi) &= \text{Sq} \left\{ \frac{1}{\pi} \arcsin \left[A(\xi); \Delta x \xi + \frac{\phi(\xi)}{2\pi} \right] \right\} \\ &= \sum_{m=-\infty}^{\infty} \frac{1}{\pi} \arcsin[A(\xi)] \text{sinc} \left\{ m \frac{1}{\pi} \arcsin[A(\xi)] \right\} \cdot \\ &\quad \bullet e^{2\pi im \left[\Delta x \xi + \frac{\phi(\xi)}{2\pi} \right]} \end{aligned} \quad (1-44)$$

which simplifies to

$$\begin{aligned} V(\xi) &= \sum_{m=-\infty}^{\infty} \frac{1}{\pi} \arcsin[A(\xi)] \\ &\quad \bullet \text{sinc} \left\{ \frac{m}{\pi} \arcsin[A(\xi)] \right\} e^{im\phi(\xi)} e^{2\pi im\Delta x \xi} \end{aligned} \quad (1-45)$$

With the definition

$$\begin{aligned} U_m(\xi) &= \frac{1}{\pi} \arcsin[A(\xi)] \text{sinc} \left\{ \frac{m}{\pi} \arcsin[A(\xi)] \right\} e^{im\phi(\xi)} \\ &= \frac{1}{\pi} \arcsin[A(\xi)] \frac{\sin \left\{ \pi \frac{m}{\pi} \arcsin[A(\xi)] \right\}}{\pi \frac{m}{\pi} \arcsin[A(\xi)]} e^{im\phi(\xi)} \\ &= \frac{1}{m\pi} \sin \left\{ m \arcsin[A(\xi)] \right\} e^{im\phi(\xi)} \end{aligned} \quad (1-46)$$

we have that

$$V(\xi) = \sum_{m=-\infty}^{\infty} U_m(\xi) e^{2\pi im\Delta x \xi} \quad (1-47)$$

The important consideration is the reconstruction at the first diffraction order, i.e., the structure of the $m = 1$ term. We see that its structure is really quite simple,

$$\begin{aligned} U_1(\xi) e^{2\pi i \Delta x \xi} &= \frac{e^{2\pi i \Delta x \xi}}{\pi} \sin \left\{ \arcsin [A(\xi)] \right\} e^{im\phi(\xi)} \\ &= \frac{e^{2\pi i \Delta x \xi}}{\pi} A(\xi) e^{i\phi(\xi)} = \frac{e^{2\pi i \Delta x \xi}}{\pi} U(\xi) \end{aligned} \quad (1-48)$$

The inverse Fourier transform of this term is

$$\frac{1}{\pi} u(x + \Delta x) \quad (1-49)$$

We have purposefully made the displacement for this case twice as large as the displacement we used for the cosine modulation case. There is a modulation image due to the $m=0$ term that is located on the optical axis. It is to move away from this disturbing influence that the displacement is increased.

2.9 Multiplexing

The term "multiplexing" refers to encoding two or more images in a single CGH transmittance. There are several techniques to accomplish this multiplexing. Two of the most common are space-division multiplexing and frequency-division multiplexing. In space-division multiplexing, the components containing images in the hologram are spatially separated, effectively taking two different holograms and recording them side-by-side on the transparency. In frequency-division multiplexing, the two hologram transmittances are modulated onto two different carrier frequencies. These carrier frequencies may differ in magnitude, direction or both. The particular implementation we will discuss is two-fold multiplexing (duplexing) where we use frequency-division multiplexing and the two modulation images appear in two distinct diffraction orders. Multiplexing is particularly useful for including color and even polarization in the reconstructed image.

2.9.1 Crossed Square Wave Carrier and Modulation

The carrier we refer to as crossed square-wave is

$$\begin{aligned} Sq(p; \alpha) Sq(q; \beta) &= \\ = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \alpha \beta \operatorname{sinc}(m\alpha) \operatorname{sinc}(n\beta) e^{2\pi i(m\phi + n\theta)} \end{aligned} \quad (1-50)$$

The two image transforms that we will be encoding on the carrier are

$$Ae^{i\phi} \quad Be^{i\theta} \quad (1-51)$$

We will make the substitutions in two stages. First we look at the phases to be encoded and make the substitutions

$$p = \Delta x \xi + \frac{\phi}{2\pi} \quad q = \Delta y \eta + \frac{\theta}{2\pi} \quad (1-52)$$

This substitution gives the modulation images

$$U_{mn}(\xi, \eta) = \alpha \beta \operatorname{sinc}(m\alpha) \operatorname{sinc}(n\beta) e^{im\phi} e^{in\theta} \quad (1-53)$$

We then select two particular diffraction orders that should contain the reconstructions. First the (1,0) diffraction order

$$U_{1,0}(\xi, \eta) = \alpha \beta \operatorname{sinc}(m\alpha) e^{im\phi} \quad (1-54)$$

and then the (0,1) diffraction order

$$U_{0,1}(\xi, \eta) = \alpha \beta \operatorname{sinc}(n\beta) e^{in\theta} \quad (1-55)$$

In order that the proper magnitude be obtained simultaneously in the two diffraction orders, we have two simultaneous non-linear equations to solve.

$$\begin{aligned} A &= \alpha \beta \operatorname{sinc}(m\alpha) = \frac{\beta}{\pi} \sin(\pi\alpha) \\ B &= \alpha \beta \operatorname{sinc}(n\beta) = \frac{\alpha}{\pi} \sin(\pi\beta) \end{aligned} \quad (1-56)$$

We can either solve these equations through variable substitution or we can form a look-up table to be used in plotting the duplex hologram.

2.10 Modulation Image Superposition

Although the appearance is of a very neat air-tight calculation, there is an Achilles heel; the higher orders are not necessarily of bounded support even when the original object is. The higher orders may leak over into the desired reconstruction. The second consideration is that pulse-code modulation calculation done here is in continuous mathematics. When the hologram transmittance is sampled, this sampling will lead to a periodic replication of the entire reconstruction plane thereby bringing higher diffraction orders into registration with the desired reconstruction region.

It is convenient here to introduce the term "modulation image". The modulation image is the Fourier transform of a modulated amplitude associated with a carrier for a given order. The total image is a sum of the modulation images. The modulation images are spatially translated due to their association with the various carriers but may overlap. We will term the modulation image, which is the desired reconstruction, to be the true image. Other images in this super position we will call "false images". As an example, in the square wave decomposition above, the modulation image associated with the $m = 1$ carrier, is the true image. The images associated with all other carriers, are false images.

3. HOLOGRAPHY

The word "hologram" is assembled from Greek words approximating the meaning "entire recording". Entire in this context is meant to distinguish the hologram from the photograph. A photograph records only the irradiance or strength of an incident wave whereas the hologram records not only the strength of the wave at each point on its surface but also the direction that the wave is propagating at that point. The direction of propagation is equivalent of knowledge of the phase of the wave across the recording medium. The gradient of the phase gives the propagation direction.

3.1 Some historical events related to computer-generated holography

1. Bragg (1939). In this experiment, an x-ray diffractogram was recorded of a crystal. The x-ray diffractogram is in essence a Fourier transform of the crystal structure. Bragg reasoned that for a center symmetric structure, the diffractogram would be real so that recording the x-rays with film would not lose the diffractogram

phase. He then synthesized a hologram by drilling holes at the diffraction peak locations in a sheet of brass. Illuminating with a filter mercury arc lamp, the wave propagation implemented an inverse Fourier transform and so the atoms in the crystal structure were reconstructed. Because of the wave length difference between x-rays and optical waves, a strong magnification was achieved.

2. Gabor (1948). Gabor initially called holography "wave-front reconstruction". He was attempting to develop a method for recording and reconstructing electron micrographs. In the process, what we now call the "on axis interferometric hologram" was born. He imaged a low contrast object that was well described as the sum of two waves: the background and modulation. The hologram was formed by the interference of these waves.
3. Rogers (1949). Rogers reasoned that for simple objects, for instance a wire, the interference pattern for an on-axis hologram could be analytically calculated and the synthetic hologram drawn by hand. He actually fabricated the holograms and reconstructed the images.
4. Lohmann (1956). With "single-sideband" holography Lohmann combined communications theoretical and physical views of optics to address the twin-image problem of Gabor holograms.
5. Leith and Upatnieks (1962) invented the off-axis hologram. With the advent of the laser and the invention of this technique, very high quality holograms began to capture the imagination of the scientific and popular world.
6. Lohmann and Brown (1966) invented the binary computer-generated hologram. Here the newly increased power of computers and the recently invented fast Fourier transform algorithm were combined with a calculation method utilizing detour phase to create holograms that had transmittance of only zero and one across their surfaces.
7. L.B. Lesem, P.M. Hirsch, and J.A. Hordan Jr., (1969). The Kinoform, a phase-only phase CGH uses dielectric phase retardation implemented through index of refraction change, on thickness change.

3.2 Classifying CGH's

We now compare and contrast CGH's that fall into two broad classes. These classes are cell-oriented and point-oriented. Holograms in these two classes differ in the way they utilize resources. These resources are computation and plotting or printing. The most significant computational resource is the size of the discrete Fourier transform used in the wave propagation calculation. The second resource is the recording device: the printer or plotter together with the final medium in which the CGH information is encoded.

In point-oriented CGH's, each pixel is uniform. There is no sub-structure. Even though a hologram pixel may consist of many output device points, all of these points are set to the same value. Point-orientation derives its name from translating each point in the calculated hologram transmittance to a pixel value in the actual hologram. Cell-oriented holograms, on the other hand, manipulate the internal structure of each hologram pixel so that each pixel in some way controls amplitude and phase. The two types of holograms have different advantages and different limitations. In order to explore these differences, we first take two specific examples, one from each class. These examples will be of binary, amplitude and phase encoded, Fourier CGH's. The resources available to the two holograms will be identical. We assume an output device capable of 600 dpi nominal resolution and 300 dpi practically obtainable resolution. We will assume 6.8 x 6.8 inches of plotting area which will give 2048 x 2048 addressable points. The computational resources will be limited by a 128 x 128 point DFT. We will assume that the hologram will be used for display purposes so that the object will be overlaid with a random diffuser. The point-oriented hologram will be generalized-harmonic encoded in crossed square wave carriers. The cell-oriented hologram will be a Lohmann-type 3 detour phase encoded hologram. The following table summarizes the operations in constructing the two types of holograms.

Point-Oriented	Cell-Oriented
Symmetrize the object	
Discrete Fourier transform	Discrete Fourier transform
Binarize	Print Cells
Reconstruct	Reconstruct

3.3 Point-oriented CGH's

The point oriented hologram is characterized by a resolution cell that has no internal structure. These holograms are generated by calculating the total transmittance of the hologram after the effective interference of the object wave with the reference wave. This total transmittance is then sampled and the sample value is used to determine the transmission of each resolution cell. The most commonly used carriers are the cosine wave and the square-wave.

3.3.1 The Perfect CGH

The perfect CGH is an idealization that intended to reduce calculations to a bare minimum yet retain some of the properties we will be dealing with throughout the course. This CGH will consist of an array of point apertures. Each aperture has a complex-amplitude transmittance, i.e., it modifies both the amplitude and phase of the light passing through it.

Using Comb math, we can write the transmission of an idealized CGH in very compact form. The idealized CGH is one that has a periodic point structure. The points have complex amplitude transmittances and the reconstruction is simply by Fourier transform. Such a hologram transmittance in one dimension is

$$\begin{aligned} T_{\text{PERFECT_CGH}}(\xi) &= \mathcal{F}^{-1} \left\{ \left[\text{comb}\left(\frac{x}{\delta x}\right) \right] \left[u(x) * \text{comb}\left(\frac{x}{M\delta x}\right) \right] \right\} \\ &= [\delta x \text{comb}(\delta x \xi)] * [U(\xi) M\delta x \xi] \end{aligned} \quad (1-57)$$

The reconstruction is done by a simple inverse Fourier transform. The result is immediate using Comb math

$$\begin{aligned} v(x) &= \mathcal{F}^{-1} [T_{\text{PERFECT_CGH}}(\xi)] \\ &= \left\{ \text{comb}\left(\frac{x}{\delta x}\right) \right\} \left[u(x) * \text{comb}\left(\frac{x}{M\delta x}\right) \right] \end{aligned} \quad (1-58)$$

For this type of hologram, simulation is straight-forward and we follow the steps:

- (1) load the object
- (2) apply a random-phase diffuser
- (3) Fourier transform
- (4) inverse Fourier transform
- (5) display

One variation of this procedure can bring us closer to simulating the point nature of the ideal CGH, imbed the object in one quarter of the available field, apply a random phase, Fourier transform, set all of the array elements to zero except those at every other point, inverse Fourier transform, display.

3.3.2 The Cosine CGH

The representation for the on-axis hologram can be extended to media that control only the amplitude and not the phase. This extension is done by choosing a slightly more complicated object. The desired object is moved off-axis and to it a twin image is added. This twin image has hermitian symmetry to the original object. The Fourier transform of this compound object, because it has hermitian symmetry, is real. By seeking the minimum of this real transmittance and adding a constant to the transform value, the transmittance can be made real and positive. Such a transmittance can be realized using monochrome film. The symmetrization is expressed as follows.

$$u_s(x, y) = u_0(x - x_0, y - y_0) + u_0^*(-x - x_0, -y - y_0) \quad (1-59)$$

Adding the bias modifies this expression to

$$u(x, y) = u_s(x, y) + \text{constant} \quad (1-60)$$

the mathematics of reconstruction are then exactly the same as presented for the on-axis hologram.

3.3.3 The Square-wave CGH

The carrier for this hologram is a square-wave. The corresponding optical element is a Ronchi ruling. The local duty cycle of this square wave is modulated to encode the amplitude and the lateral displacement modulated to encode the phase

$$Sq(a; p) = \sum_{m=-\infty}^{\infty} a \operatorname{sinc}(ma) e^{2\pi i m p} \quad (1-61)$$

The appropriate substitution for the general variables is

$$p = \Delta x \xi + \frac{\phi}{2\pi} \quad a = \frac{I}{\pi} \arcsin(A) \quad (1-62)$$

where

$$\begin{aligned} T_{CGH}(\xi, \eta) &= \sum_{m=-\infty}^{\infty} \frac{1}{\pi} \arcsin(A) \\ &\cdot \operatorname{sinc}\left[\frac{m}{\pi} \arcsin(A)\right] e^{2\pi i m \Delta x \xi} e^{im\phi(\xi)} \end{aligned} \quad (1-63)$$

The modulation image interpretation of this sum is

$$T_{CGH}(\xi, \eta) = \sum_{m=-\infty}^{\infty} C_m U_m(\xi, \eta) e^{2\pi i m \Delta x \xi} \quad (1-64)$$

These modulation-image transforms are defined by

$$U_m(\xi, \eta) = \frac{I}{\pi} \arcsin(A) \operatorname{sinc}\left[\frac{m}{\pi} \arcsin(A)\right] e^{im\phi} \quad (1-65)$$

Performing an inverse Fourier transform on these functions gives the actual false images,

$$u_m(x, y) = IFT[U_m(\xi, \eta)] \quad (1-66)$$

It is important to note that the image of interest which falls in the (1, 0) diffraction order is exactly the desired reconstruction,

$$u_1(x, y) = IFT[A(\xi, \eta) e^{i\phi(\xi, \eta)}] \quad (1-67)$$

The coefficients $C_m = 1$ determine the irradiances and the factors $e^{2\pi i m \Delta x \xi}$ determine the positions of the modulation images.

3.3.4 The Crossed-Square-wave CGH

$$\text{Sq}(a; p)\text{Sq}(b; q) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} ab \text{sinc}(ma) \text{sinc}(nb) e^{2\pi i(m p + n q)} \quad (1-68)$$

This hologram will be generated by modulating crossed square wave gratings. Recall that the unmodulated crossed square-wave grating transmittance is

$$\begin{aligned} & \text{Sq}\left(\frac{1}{2}; 2\pi\Delta x \xi\right) \text{Sq}\left(\frac{1}{2}; 2\pi\Delta y \eta\right) \\ &= \frac{1}{4} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \text{sinc}\left(\frac{m}{2}\right) \text{sinc}\left(\frac{n}{2}\right) e^{2\pi i(m\Delta x \xi + n\Delta y \eta)} \end{aligned} \quad (1-69)$$

Modulating one grating with the amplitude and the other grating with the phase requires the substitutions

$$a = \frac{1}{2} \quad b = A \quad p = \Delta x \xi + \frac{\phi}{2\pi} \quad q = \Delta y \eta \quad (1-70)$$

These substitutions give the generalized harmonic decomposition

$$\begin{aligned} T_{CGH}(\xi, \eta) &= \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{1}{2} \text{sinc}\left(\frac{m}{2}\right) \\ &\cdot A \text{sinc}(nA) e^{2\pi i(m\Delta x \xi + n\Delta y \eta)} e^{im\phi} \end{aligned} \quad (1-71)$$

This decomposition can be interpreted as a sum of modulation image Fourier transforms

$$T_{CGH}(\xi, \eta) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} C_{mn} U_{mn}(\xi, \eta) e^{2\pi i(m\Delta x \xi + n\Delta y \eta)} \quad (1-72)$$

These false-image transforms are defined by

$$U_{mn}(\xi, \eta) = A \operatorname{sinc}(nA) e^{im\phi} \quad (1-73)$$

Performing an inverse Fourier transform on these functions gives the actual false images,

$$u_{mn}(x, y) = IFT [U_{mn}(\xi, \eta)] \quad (1-74)$$

It is important to note that the image of interest which falls in the (1, 0) diffraction order is exactly the desired reconstruction,

$$u_{1,0}(x, y) = IFT [A(\xi, \eta) e^{i\phi(\xi, \eta)}] \quad (1-75)$$

The coefficients

$$C_{mn} = \frac{1}{2} \operatorname{sinc}\left(\frac{m}{2}\right) \quad (1-76)$$

control the strength of the false images.

The factors $e^{2\pi i(m\Delta x\xi + n\Delta y\eta)}$ determine the positions of the false images through shifts.

3.4 Cell-oriented CGH's

We consider two closely related hologram types: the Lohmann Type III, binary CGH and its circular overflow corrected version. We will show how a handy mathematical solution can transform the troublesome linear phase error into the familiar false image decomposition.

3.4.1 The Lohmann Type-III Binary CGH

We first do the calculation for an ideal Lohmann type III hologram. In the next section we will expand the calculation to include circular overflow correction.

The resolution cell structure for ideal type-III CGH is that a single rectangular aperture is placed in the cell. The aperture is centered vertically. Its height is proportional to the desired amplitude transmittance. Its lateral position is shifted proportionally to the phase of the complex amplitude

transmittance for that cell. We express the hologram transmittance as a sum of rectangle functions

$$T_{CGH} = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \text{rect}\left(\frac{\xi - \frac{m}{\Delta x} - \frac{\phi_{mn}}{2\pi\Delta x}}{\frac{1}{2\Delta x}}\right) \text{rect}\left(\frac{\eta - \frac{n}{\Delta y} - \frac{A_{mn}}{\Delta y}}{\frac{A_{mn}}{\Delta y}}\right) \quad (1-77)$$

On inverse Fourier transforming this array of rectangles assumes an interesting but somewhat difficult to interpret form,

$$t_{CGH}(x, y) \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{1}{2\Delta x \Delta y} \text{sinc}\left(\frac{x}{2\Delta x}\right) A_{mn} \text{sinc}\left(\frac{A_{mn}y}{\Delta y}\right) e^{\frac{x}{\Delta x} \phi_{mn}} e^{2\pi i \left(\frac{mx}{\Delta x} + \frac{ny}{\Delta y}\right)} \quad (1-78)$$

The traditional approach is to analyze the reconstruction by picking

$x = \Delta x$ $y = 0$. The reconstruction is perfect at this point. One then assumes small deviations from perfection for the points in the neighborhood. A more analytically tractable approach makes use of the identity

$$e^{i\alpha\phi} = \sum_{n=-\infty}^{\infty} \text{sinc}(\alpha - n) e^{in\phi} \quad (1-79)$$

This identity allows us to expand the phase terms. The x-phase term requires only a simple substitution.

$$e^{\frac{i}{\Delta x} \phi_{mn}} = \sum_{j=-\infty}^{\infty} \text{sinc}\left(\frac{x}{\Delta x} - j\right) e^{ij\phi_{mn}} \quad (1-80)$$

The y-phase term requires more caution. The following is a useful identity.

$$A \operatorname{sinc}(A\beta) = \sum_{k=-\infty}^{\infty} \operatorname{sinc}(\beta - k) A \operatorname{sinc}(kA) \quad (1-81)$$

The amplitude term of interest is

$$A_{mn} \operatorname{sinc}\left(\frac{A_{mn} y}{\Delta y}\right) = \sum_{k=-\infty}^{\infty} \operatorname{sinc}\left(\frac{y}{\Delta y} - k\right) A_{mn} \operatorname{sinc}(kA_{mn}) \quad (1-82)$$

Substituting for both the amplitude and the phase terms we have

$$\begin{aligned} t_{CGH}(x, y) &= \frac{1}{2\Delta x \Delta y} \operatorname{sinc}\left(\frac{x}{2\Delta x}\right) \\ &\cdot \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \operatorname{sinc}\left(\frac{y}{\Delta y} - k\right) \operatorname{sinc}\left(\frac{x}{\Delta x} - j\right) \\ &= \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} A_{mn} \operatorname{sinc}(kA_{mn}) e^{i\phi_{mn}} e^{2\pi i \left(\frac{mx}{\Delta x} + \frac{ny}{\Delta y}\right)} \end{aligned} \quad (1-83)$$

defining the false-images to be

$$u_{jk}(x, y) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} A_{mn} \operatorname{sinc}(kA_{mn}) e^{i(j\phi_{mn})} e^{2\pi i \left(\frac{mx}{\Delta x} + \frac{ny}{\Delta y}\right)} \quad (1-84)$$

and the coefficients to be

$$C_{jk}(x, y) = \operatorname{sinc}\left(\frac{y}{\Delta y} - k\right) \operatorname{sinc}\left(\frac{x}{\Delta x} - j\right) \quad (1-85)$$

we have that

$$t_{CGH}(x, y) = \frac{1}{2\Delta x \Delta y} \operatorname{sinc}\left(\frac{x}{2\Delta x}\right) \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} C_{jk} u_{jk} \quad (1-86)$$

3.4.2 The Gap and Overlap Problem: Circular Overflow Solution

One simplification we have made in this analysis for the cell-oriented hologram is to ignore the gap and overlap problem. Rectangles from adjacent cells will, at times overlap. When they overlap, because we are using a linear model, the CGH transmission will no longer be binary. This problem can be eliminated by using a slightly modified cell structure called circular overflow. Briefly put, when the rectangle flows out of its own resolution cell, it is truncated at the edge. The overflowed portion is then placed at the opposite edge of the cell. For a single cell, this procedure can be mathematically described by

$$\begin{aligned} & \operatorname{rect}\left(\frac{\xi - \frac{\phi_{mn}}{2\pi\Delta x}}{\frac{1}{2\pi\Delta x}}\right) \\ & \rightarrow \operatorname{rect}(\Delta x \xi) \left[\operatorname{rect}\left(\frac{\xi - \frac{\phi_{mn}}{2\pi\Delta x}}{\frac{1}{2\pi\Delta x}}\right) * \Delta x \operatorname{comb}(\Delta x \xi) \right] \end{aligned} \quad (1-87)$$

The transform of the phase-related aperture function then becomes

$$\begin{aligned} & \frac{1}{\Delta x} \operatorname{sinc}\left(\frac{x}{\Delta x}\right) * \left[\frac{1}{2\Delta x} \operatorname{sinc}\left(\frac{x}{2\Delta x}\right) e^{i \frac{x}{\Delta x} \phi_{mn}} \operatorname{comb}\left(\frac{x}{\Delta x}\right) \right] = \\ & \frac{1}{\Delta x} \operatorname{sinc}\left(\frac{x}{\Delta x}\right) * \left[\frac{1}{2} \operatorname{sinc}\left(\frac{x}{2\Delta x}\right) e^{i \frac{x}{\Delta x} \phi_{mn}} \frac{1}{\Delta x} \sum_{j=-\infty}^{\infty} \delta\left(\frac{x}{\Delta x} - j\right) \right] = \\ & \sum_{j=-\infty}^{\infty} \operatorname{sinc}\left(\frac{x}{\Delta x} - j\right) \operatorname{sinc}\left(\frac{j}{2}\right) e^{ij\phi_{mn}} \end{aligned} \quad (1-88)$$

Notice that the conversion from a linear multiplier to integral multiples of the phase is included automatically in the calculation. We now include the factor from the amplitude so that the coefficients in the false-image decomposition become

$$C_{jk}(x, y) = \text{sinc}\left(\frac{x}{\Delta x} - j\right) \text{sinc}\left(\frac{j}{2}\right) \text{sinc}\left(\frac{y}{\Delta y} - k\right) \quad (1-89)$$

The structure of the images as well as their die-off remain unchanged, only their strengths change.

4. CGH IMPERFECTIONS

In this section we deal with the five sources of image degradation: phase quantization, phase clipping, phase matching, the linear phase error in detour-phase holograms, and the constant modulus error in kinoforms. Each of these distortions is conveniently described as a point-nonlinearity.

4.1 Complex Amplitude Degradations

- Degradation of the complex amplitude vs. degradation of irradiance. For off-axis CGH's the reconstructed images, both those from the various diffraction orders and those coming from imperfections are spatially separated. We can therefore calculate complex amplitudes and absolute-square the result to get irradiance because there is no interaction among the images.
- Phase nonlinearities. We will be dealing with imperfections in storing the amplitude and the phase.
- Display vs. diffractive optical elements. Many times DOE's use an axial carrier that corresponds to a longitudinal displacement.

4.2 Sampling and Quantization

Consider a function of position, x , that has a value $v(x)$. Sampling is a discretization of the signal on the spatial axis. Quantization is a discretization of the signal on the value axis. For CGH's, the value is typically amplitude or phase. For sampling alone, we have an integer number of sample values. Each value can attain a value from continuous set.

For quantization alone, there are values on all of the spatial axis. However, there are only an integer number of attainable values.

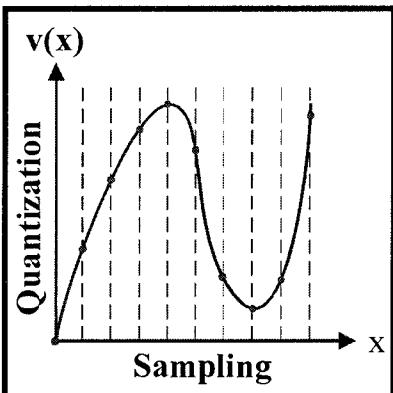


Figure 1-1: Sampling alone

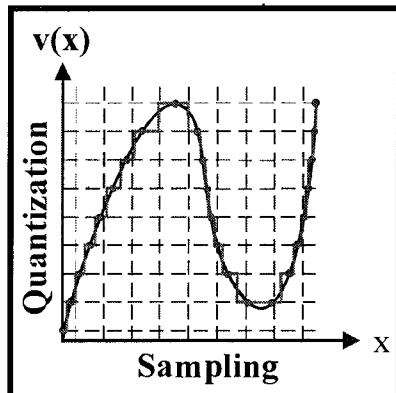


Figure 1-2: Quantization alone

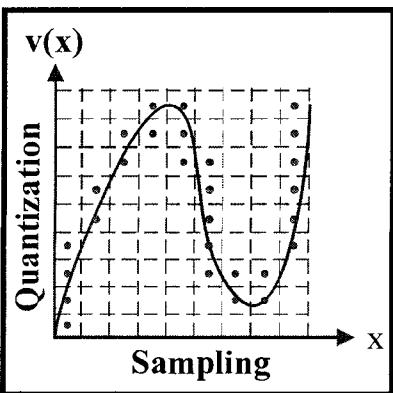


Figure 1-3: Sampling and Quantization

4.3 Sampling and Point-Oriented CGH's

Until now we have considered continuous transmittances in our discussion of binarization. We have, however, ignored sampling. At this point we include sampling in our consideration. It is important to note that for this type of hologram, the order of binarization and sampling is interchangeable. Sampling can easily be described by multiplication with a two-dimensional comb with the hologram transmittance. The most difficult problem is using the intervals properly. Recall that the extent of the Fourier CGH is equal to the reciprocal of the sampling interval in the object. We begin with a continuous CGH of extent $\Delta\xi$ by $\Delta\eta$ and a sampling intervals of $\delta\xi$ and $\delta\eta$. The sampled CGH transmittance is

$$T_{CGH} \rightarrow T_{CGH} \cdot \text{comb}\left(\frac{\xi}{\delta\xi}\right) \text{comb}\left(\frac{\eta}{\delta\eta}\right) \quad (1-90)$$

This multiplication in the Fourier domain translates to a convolution in the spatial domain. This convolution is a replication over the entire reconstruction plane which is,

$$t_{CGH} \rightarrow \delta\xi\delta\eta \cdot \{t_{CGH} \ast \ast [\text{comb}(\delta\xi x) \text{comb}(\delta\eta y)]\} \quad (1-91)$$

This equation simplifies for particular choices of the replication distance which are determined by the sampling intervals. Let's take two examples. The first example is for the CGH sampling interval being equal to the reciprocal of the object extent, $\delta\xi = \frac{1}{\Delta x}$ and $\delta\eta = \frac{1}{\Delta y}$, and the object sampling interval

equal to the reciprocal of the CGH extent $\delta x = \frac{1}{\Delta\xi}$ and $\delta y = \frac{1}{\Delta\eta}$. For this example we have the modulation-image decomposition

$$\sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} c_{mn} u_{mn}(x - m\Delta x, y - n\Delta y) \quad (1-92)$$

The strengths of the images in the desired reconstruction region are given by the coefficients

$$c_{mn} = \frac{1}{2} \text{sinc}\left(\frac{m}{2}\right) \quad (1-93)$$

We see that the strength of the modulation images dies off as a sinc-function.

For the second example, we choose the replication distance to be twice the desired object extent. The false image decomposition is

$$\sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} C_{mn} u_{mn}(x - 2m\Delta x, y - 2n\Delta y) \quad (1-94)$$

the strengths of the images in the desired reconstruction region are given by the coefficients

$$c_{mn} = \frac{1}{2} \operatorname{sinc}\left(\frac{2m+1}{2}\right) \quad (1-95)$$

except for the strength of the true image which remains

$$c_{1,0} = \frac{1}{2} \operatorname{sinc}\left(\frac{1}{2}\right) \quad (1-96)$$

We see that the false image die-off again varies as a sinc function, but is accelerated by a factor of 2 in the sinc function argument. This behavior corresponds with our intuition. The finer we sample the hologram, the more exact our reconstruction and the more nearly it approaches the best attainable value, i.e. that for a continuous CGH.

4.4 Phase Nonlinearities

The complex-amplitude transmittance of the CGH can be decomposed into a linear superposition in a way that is meaningful for describing imperfections. When we are dealing with a phase-nonlinearity we have an ideal phase that we are attempting to realize. However, for some reason the ideal is not attainable. If we call the ideal phase ϕ and the realized phase $\hat{\phi}$ we are assuming a point-wise functional relation $\hat{\phi} = \hat{\phi}(\phi)$. We can display this relation as a graph such as the one shown in Figure 4. A perfect realization would correspond to the 45° dashed line in that figure. The solid line represents an imperfect realization.

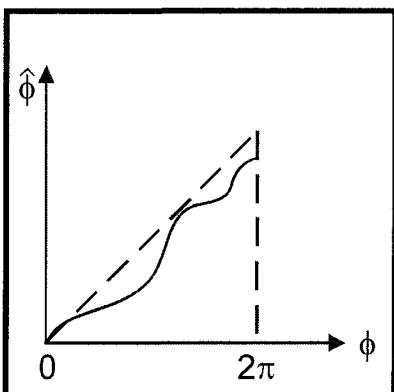


Figure 1-4: Phase Nonlinearity

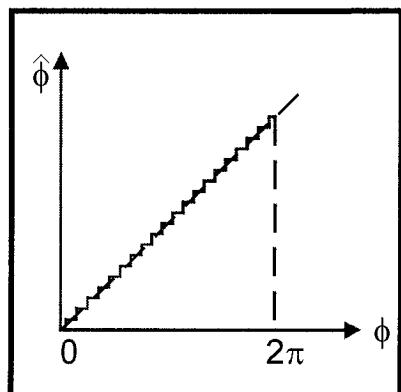


Figure 1-5: Phase Quantization

The nonlinearity gives rise to a complex amplitude $\hat{u} = Ae^{i\hat{\phi}}$ in place of the ideal $u = Ae^{i\phi}$. We can expand the function $e^{i\hat{\phi}} = e^{i\hat{\phi}(\phi)}$ in a Fourier series such that $e^{i\hat{\phi}} = \sum_{m=-\infty}^{\infty} c_m e^{im\phi}$. The Fourier-series coefficients are calculated by

$$c_m = \frac{1}{2\pi} \int_0^{2\pi} e^{i\hat{\phi}(\phi)} e^{-im\phi} d\phi.$$

The utility of this decomposition is similar to that

of the modulation-image decomposition that we've encountered in the encoding process. For an ideal complex-amplitude, the action of the phase nonlinearity generates a linear superposition of complex amplitudes. The superposition is $Ae^{i\hat{\phi}} = \sum_{m=-\infty}^{\infty} c_m (Ae^{im\phi})$. Or, $\hat{u} = \sum_{m=-\infty}^{\infty} c_m u_m$ where $u_m = Ae^{im\phi}$. We see that the sum consist of

$$\hat{u} = c_1 u_1 + \sum_{m \neq 1} c_m u_m = c_1 u + \sum_{m \neq 1} c_m u_m \quad (1-97)$$

including the ideal or true term and a collection of false terms.

4.4.1 Fourier Domain Phase Quantization

When fabricating CGH's, the phase can become quantized in ways that depend on the fabrication process. If a deposition or etching process is used to manipulate the dielectric phase, the quantization can come from the limited number of layers. For carrier-wave based CGH's, geometrical limitations such a pixelization can cause the phase quantization. The object has complex amplitude $u(x)$ and Fourier spectrum $U(\xi) = A(\xi)e^{i\phi(\xi)}$. Quantizing $\phi(\xi)$ into N bins gives $\hat{U}(\xi) = A(\xi)\exp[i\hat{\phi}(\xi)]$. The superposition becomes

$$U = c_0 U_0 + \sum_{n \neq 0} c_n U_n = \text{sinc}\left(\frac{I}{N}\right) Ae^{i\phi} + \sum_{n \neq 0} \text{sinc}\left(n + \frac{I}{N}\right) Ae^{i(nN+1)\phi} \quad (1-98)$$

On inverse transforming, the image becomes

$$\hat{u}(x) = \text{sinc}\left(\frac{1}{N}\right)u(x) + \sum_{n \neq 0} \text{sinc}\left(n + \frac{1}{N}\right)u_n(x) \quad (1-99)$$

with

$$u_n(x) = \int_{-\infty}^{\infty} A(\xi) e^{i(nN+1)\phi(\xi)} e^{2\pi i \xi x} d\xi \quad (1-100)$$

For phase-only CGH's, the amplitude is constant and the false images become multiple convolutions of the original image.

4.4.2 Phase Clipping

Another important phase-nonlinearity is phase clipping. It happens when the material depth is not sufficient for a 2π -etch depth. The graph of phase clipping is shown in Figure 6. The output phase is equal to the input phase up to a clipping phase ϕ_c . The output phase remains constant at this level to the remainder of the input range.

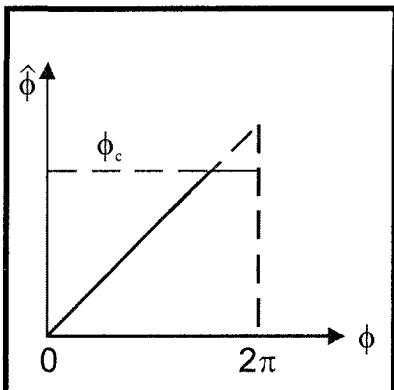


Figure 1-6: Phase clipping

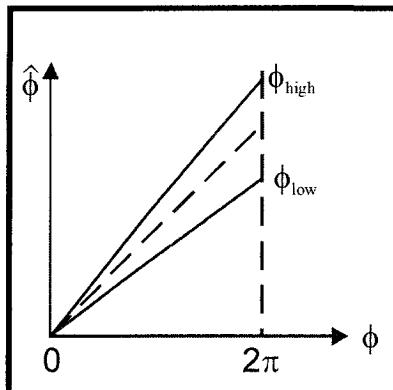


Figure 1-7: Phase mismatching

4.4.3 Phase Mismatching

Phase mismatching occurs when the process realizes an improper slope for the phase mapping. Fig. 1-7 illustrates two cases: too low a slope and too high a slope.

5. THE THIRD DIMENSION

Until this point, we have restricted our attention to two-dimensional objects and wave propagation in the Fraunhofer regime. Holograms are, however, more interesting for their ability to display three-dimensional objects. In this chapter we will shift our attention to the theoretical elements necessary to produce such holograms. These elements focus on illumination, transmission, reflection and propagation.

A more general approach will also describe diffractive optical elements or DOE's. We will see that exactly the same tools that we apply to describing production and defect analysis of display holograms will be directly applicable to DOE's.

Our 2-D calculations have concentrated on the illumination being a plane wave, transmission being diffused scattering, and propagation being a Fourier transform. We will continue considering propagation between parallel planes which is one of way of decomposing the most general case. The advantage to this choice is that we will continue to find the Fourier transform a most useful tool.

5.1 Illumination

The reconstruction produced by hologram is, of course, dependent on the illumination. Most often the illumination we assume is a plane wave. For diffractive optical elements we also consider spherical waves and pinhole waves. We consider the computer-generated hologram to be a thin planar element though appendices in this chapter show how to extend the generality of the CGH to include thin CGH's coded onto surfaces of arbitrary shape. The illumination will be monochromatic. Sources of more complicated spectral content can be decomposed into a superposition of monochromatic sources.

5.2 Transmission

We consider the CGH as modifying the complex amplitude of the incident wave by a simple multiplication. Again we concentrate on thin optical elements and optical elements where there is no complicated internal behavior. We also assume the usual isotropic behavior.

5.3 Reflection

Although we will concentrate on transmission CGH's, the mathematics we develop are easily applicable to reflection. The amplitude attenuation is the same. The phase generated by a reflecting surface can be generated either by index changes or by profile changes. A caution on calculating reflection phase changes: Remember to take into account the fact that the waves travel in both directions, that is, the phase will be doubled.

5.4 Propagation

The elementary function used for propagation between parallel planes is the pinhole wave. The pinhole wave has the form:

$$\begin{aligned} u_{ph}(x, y, z) &= -\frac{1}{2\pi} \frac{\partial}{\partial z} \left(\frac{e^{ikr}}{r} \right) \\ &= -\frac{1}{2\pi} \frac{\partial}{\partial z} \left[\frac{e^{ik\sqrt{x^2+y^2+z^2}}}{\sqrt{x^2+y^2+z^2}} \right] = -\frac{1}{2\pi} \left(ik - \frac{1}{r} \right) \frac{e^{ikr}}{r} \frac{z}{r} \end{aligned} \quad (1-101)$$

The final factor, z/r , is the obliquity factor $\cos(\gamma)$ where γ is the angle to the z-axis. The pinhole wave is proportional to the z-derivative of the spherical wave. Its form, as you will remember, is a point in the plane $z = 0$, i.e.,

$$u_{ph}(x, y, 0^+) = \delta(x)\delta(y) \quad (1-102)$$

and satisfies the Helmholtz equation in the open right half space. Propagation between the planes $z = 0$ and the plane $z = z_0$ is in the form of a convolution

$$\begin{aligned} u(x, y, z_1) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x_0, y_0, z_0) \\ &\quad \cdot u_{ph}(x - x_0, y - y_0, z_1 - z_0) dx_0 dy_0 \end{aligned} \quad (1-103)$$

that we write in the shorthand notation

$$u(x, y, z) = u(\cdot, \cdot, 0) * * u_{ph}(\cdot, \cdot, z) \quad (1-104)$$

It is convenient to think of the pinhole wave as the propagation mechanism between the two planes. In fact, the propagation operation \mathbf{P} between two planes can explicitly be expressed as the convolution

$$u(x, y, z) = \mathbf{P}[u(x, y, 0)] = u(\cdot, \cdot, 0) * * u_{ph}(\cdot, \cdot, z) \quad (1-105)$$

The operation of propagation is a linear operator so we can easily demonstrate its effect explicitly. First we use the sifting property of the delta function to express the object in the plane $z = 0$.

$$u(x, y, 0) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x_0, y_0, 0) \delta(x - x_0) \delta(y - y_0) dx_0 dy_0 \quad (1-106)$$

Next we apply the propagation operator which, because it is linear, penetrates the integral, passes through the function of dummy variables, and operates on the delta functions. The operation is simply conversion of the delta functions to the appropriate plane of the pinhole wave and so we have demonstrated the original equation. Direct space propagation is described by:

$$\begin{aligned} \mathbf{P}[u(x, y, 0)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x_0, y_0, 0) \mathbf{P}[\delta(x - x_0) \delta(y - y_0)] dx_0 dy_0 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x_0, y_0, 0) u_{ph}(x - x_0, y - y_0, z) dx_0 dy_0 \end{aligned} \quad (1-107)$$

Because the propagation operation is a 2-D convolution, Fourier transforming the equation leads to a simplification of the 2-D convolution to a product. And, in fact, we recognize the equation as a spatial filtering operation. The filter is the complex-amplitude transfer function of free space.

$$U(\xi, \eta, z) = U(\xi, \eta, 0) \bullet U_{ph}(\xi, \eta, z) \quad (1-108)$$

We can look at the propagation either in direct space or in Fourier space. In direct space, we already have an explicit expression for the propagation kernel. It remains to transform that kernel to obtain an expression for the transfer function. We could directly transform the kernel. However it is somewhat easier to use the following indirect approach. We will use the fact that the pinhole wave is proportional to the z-partial derivative of a spherical wave. The calculation is broken into three steps:

1. Find the three-dimensional Fourier transform of the spherical wave by using the fact that it is the solution of the three-dimensional wave equation for a three-dimensional point in free space.
2. Multiply by the linear factor that is equivalent to taking a derivative in the z-direction.
3. Inverse transform it in one-dimension, the z-direction, in order to obtain the two-dimensional Fourier transform of the pinhole wave.

5.5 The Born Approximation

When we look at 3-dimensional wave propagation, there are many effects that have to be considered. There are also many simplifications possible. One simplification is to assume that we have a non-absorbing self-luminous object. This assumption corresponds to the Born approximation. The wave propagation will actually give us a quite nice reconstruction if there is not a high degree of overlap, that is hidden surfaces in the object we are trying to display.

5.6 The Lumped-Element Model of Propagation

When we begin considering hidden surfaces, we can make a simplification that allows us to reduce the amount of calculation to a reasonable level. This simplification is to assume that the absorption and transmission of the object is lumped into planes so that our propagation has three distinct phases:

- 1) a propagation through free space

$$u(x, y, \Delta z^-) = u(\cdot, \cdot, 0^+) * u_{ph}(\cdot, \cdot, \Delta z) \quad (1-109)$$

- 2) a transmission by a plane of the object

$$u_0(x, y, \Delta z^+) = t(x, y, \Delta z) u(x, y, \Delta z^-) \quad (1-110)$$

3) the addition of a wave generated by sources at the surface.

$$u(x, y, \Delta z^+) = u_0(x, y, \Delta z^+) + u_1(x, y, \Delta z^+) \quad (1-111)$$

It is generally sufficient to consider only the wave components that continue on to the right. In principle, there are reflected waves that travel to the left. Even worse, there are waves that multiply reflected. Having noted this difficulty, we ignore it.

5.7 Ping-Pong Propagation

Although we can calculate the propagation through the lumped model, strictly as convolutions and multiplications, it is computationally more advantageous to perform the propagation in Fourier space and the transmissions in direct space. We can then simply multiply our corresponding waves by the appropriate transfer functions or transmittances. The largest computational effort is then in the Fourier transforms going from direct space to Fourier space and back, and so forth. Because we are jumping between spaces, this type of propagation is sometimes called ping-pong propagation.

$$U(\xi, \eta, \Delta z^-) = U(\xi, \eta, \theta^+) U_{ph}(\xi, \eta, \Delta z) \quad (1-112)$$

6. CGH'S GROUPED BY PROPAGATION REGION

A hologram is the recording of a complex amplitude. This complex amplitude may occur anywhere in the propagation path of the wave. We have concentrated most of our attention on Fourier CGH=s because they conveniently demonstrate many of the interactions between CGH=s and wavefields. Let us briefly consider other alternatives.

6.1 Image Plane CGH

The image-plane CGH records the wavefield at the object or image plane in the propagation path. A strictly image-plane CGH is of course two-dimensional. The interesting variation should be called a near-image-plane CGH. The object may be two-dimensional and slightly removed from the image plane, or it may be three-dimensional but with its entire extent located

near, or embedded in, the image plane. We generally do not consider the evanescent waves, that consideration is reserved for the next type of CGH.

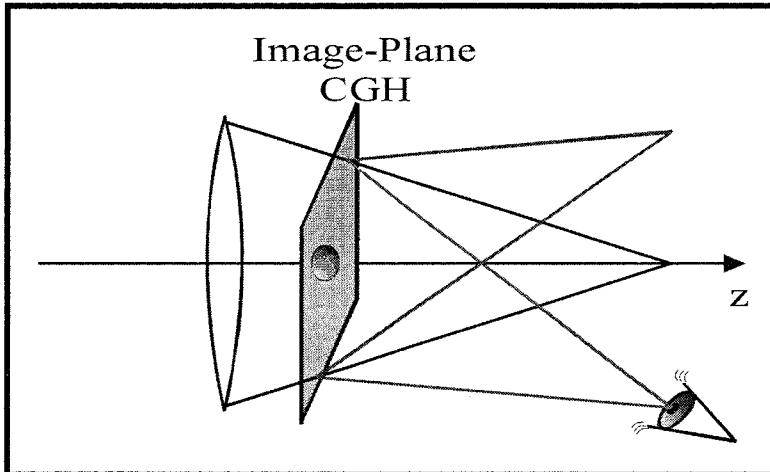


Figure 1-8: Reconstructing an Image-Plane CGH

There is a temptation to use the HFK propagation formula to calculate the wavefield at the CGH. There are however potential numerical difficulties with this approach. The paraxial form of the HFK propagation formula is

$$u(x, y, z_1) = \frac{e^{\frac{2\pi i(z_1 - z_0)}{\lambda}}}{i\lambda |z_1 - z_0|} \cdot \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x_0, y_0, z_0) e^{\frac{i\pi[(x-x_0)^2 + (y-y_0)^2]}{\lambda(z_1 - z_0)}} dx_0 dy_0 \quad (113)$$

We generally multiply out the quadratic exponent and use the resulting version of pre-multiplication by a quadratic phase factor, Fourier transform, post-multiplication by a quadratic phase factor.

$$u(x, y, z_1) = \frac{e^{\frac{2\pi i(z_1 - z_0)}{\lambda}}}{i\lambda |z_1 - z_0|} e^{\frac{i\pi(x^2 + y^2)}{\lambda(z_1 - z_0)}} \cdot \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x_0, y_0, z_0) e^{\frac{i\pi(x_0^2 + y_0^2)}{\lambda(z_1 - z_0)}} e^{\frac{-2\pi i(xx_0 + yy_0)}{\lambda(z_1 - z_0)}} dx_0 dy_0 \quad (114)$$

The difficulties are with the quadratic phase factors. When $|z_1 - z_0|$ is small, the exponential varies too rapidly for stable calculation. Let's look at an example. Suppose that we have a CGH that is 1cm x 1cm. We wish to embed an object that has 11 depth planes separated by 0.5mm. Let's use 500nm wavelength light for this example. The argument of the exponential is

$$\phi(x, y) = \frac{\pi(x^2 + y^2)}{\lambda(z_1 - z_0)} \quad (115)$$

Let's now look at the local spatial frequency at the edge of the CGH. $\xi_{edge} = \frac{\partial}{\partial x} \left[\frac{\pi x^2}{\lambda(z_1 - z_0)} \right] = \frac{2\pi x}{\lambda(z_1 - z_0)}$. Inserting the chosen numerical values gives us $\xi_{edge} = \frac{(6.28)(5mm)}{(5 \times 10^{-4} mm)(0.5mm)} = 1.26 \times 10^5 mm^{-1}$. This

would give us a CGH of pixel dimensions $N_{pixelsOnEdge} = 1.26 \times 10^6$. A better way is to use the paraxial version of the RSD propagation formula.

$$u(x, y, z_1) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(\xi, \eta, z_0) e^{i\pi\lambda(z_1 - z_0)(\xi^2 + \eta^2)} e^{2\pi i(\xi x + \eta y)} d\xi d\eta \quad (116)$$

The propagation is generally most easily done by transforming each of the object planes to Fourier space and there applying the proper quadratic phase factor. Ping-pong propagation can also be applied to incorporate hidden surface considerations. The output of the propagation is the wavefield just after exiting the object. It is this wavefield that is encoded into the CGH. The CGH itself looks like a coherently defocussed picture. When illuminated, the diffraction pattern separates into diffraction orders. The

proper order is then selected for viewing. This selection is commonly done by correctly locating the pupil of the eye in preparation for viewing.

6.2 Near-Field CGH

Although not relevant to display, we describe the this type of CGH for completeness. The near-field CGH records an object wave that is so near to the object that the evanescent waves make a significant contribution. The transfer function of free space is used in its entirety including both the square root and the evanescent shoulder.

An effect that is associated with applications of the near-field CGH is total internal reflection. The well-known formula, Snell=s Law, for bending of a plane wave at an interface between two media is

$$n_0 \sin(\theta_0) = n_1 \sin(\theta_1) \quad (117)$$

The slight rearrangement

$$\sin(\theta_1) = \frac{n_0}{n_1} \sin(\theta_0) \quad (118)$$

makes clear the fact that there will be a difficulty if $n_0 > n_1$ and $\sin(\theta_0) > \frac{n_1}{n_0}$. This difficulty is total internal reflection. The wave doesn=t

pass through the interface but is rather reflected. Geometrically, a plane wave passing from medium zero to one makes a larger angle to the axis in medium one until the wave, at the critical angle of incidence, is traveling parallel to the interface. The angle is determined by matching the wavefronts in the two media. When the incident angle is made even steeper, total internal reflection comes into effect.

The complement to this effect is demonstrated in frustrated total internal reflection. A second surface is brought very near to the first one. If the index is larger than that of medium one, then some of the evanescent waves are converted back to propagating waves. The classic example is the finger print on the water glass. For us, the variation that is important is diffractive rather than refractive. The CGH is encoded on a carrier that has a period shorter than the wavelength of the light being used. A second grating, with a

similar period, is placed very near to the CGH to frustrate the total internal reflection.

6.3 Fresnel CGH

The Fresnel CGH is located far enough from the reconstructed image that the parabolic approximation can be used. One view of the operation of this type of CGH is that it contains a superposition of Fresnel zone plates, FZP's. Each FZP has a later location corresponds to the lateral position of a points in the image. The depth of the image point corresponds the focal power of its associated FZP. A significant amount of focal power is built into the CGH itself, though focal power requires bandwidth and forces the CGH to contain very fine structure. It is often wise to combine a Fresnel CGH with a refractive optical element. The bulk of the focal power is supplied by the conventional element while the CGH supplies the flexibility of image reconstruction.

6.4 Near-Fourier CGH

The near-Fourier CGH is similar to the Fresnel-CGH, except that no significant focal power is built into the CGH. Consider the 4-F imaging system of Fig. 1-9 with a Fourier CGH at the mid-plane (Fourier plane). The reconstructed image will lie in the back focal plane of the second lens. If a quadratic phase factor is incorporated into the CGH, it will draw the reconstruction slightly out of the reconstruction plane. It is this quadratic phase factor that is at the heart of the near-Fourier CGH. Several planes can be incorporated into such a reconstruction.

6.5 Fourier CGH

The Fourier CGH contains no average focal power, it reconstructs a flat image in the back focal-plane of the second lens of the 4-F imaging system shown in Fig. 1-9. It gets its name from the fact that the encoded transmittance is the Fourier transform of the desired image. The 4-F system is also commonly referred to as the telecentric system. Telecentricity is the proper of having a pupil at infinity. As the system aperture is in the mid-plane, both the entrance and exit pupils are located at infinity so it is more accurate, though less common, to label it as a doubly-telecentric system.

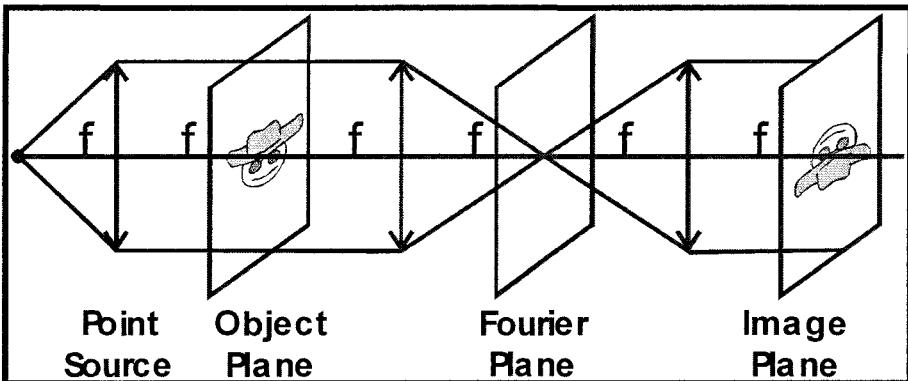


Figure 1-9: The 4-F imaging system

6.6 Fraunhofer CGH

For Fraunhofer propagation, no lens is used, propagation is simply to a large distance. To see why the Fourier transform results from this simple geometry, the method of stationary phase is well suited. Considerations specific to display CGH's. Holograms for display purposes are generally made of diffusely reflecting surfaces or transmitting objects with a superimposed diffuser. The reason for applying a diffusing element is to spread light about the hologram. Diffusers can be engineered to optimize any of several criteria. Among these is uniformity of the Fourier transform, band-limiting of the Fourier transform and spreading information redundantly about the Fourier plane. Incidentally spreading light about the Fourier plane is not the same as spreading information. It is possible to design a diffuser that spreads the light almost uniformly, but still allows the information associated with local areas in the object to remain clustered in the Fourier transform.

6.7 Desirable Properties of the Fourier Spectrum

- Uniformity
- Band Limitation
- Redundancy

6.7.1 Uniformity

The reason for spreading the light about the Fourier plane is to minimize the range of amplitude information that the hologram must capture. The relevant parameter in this consideration is the so-called dynamic range. The dynamic range is the ratio of the maximum range to the step-size. It turns out to be then the number of steps that can be reproduced in the amplitude. Typical

values for a good grey tone laser film plotter is about 2000 steps. For a laser printer, on the other hand, the plotted values are either dark or white, that is a binary plotting range.

The second reason for having uniform amplitude is to maximize the diffraction efficiency. The amplitude is normalized to a range (0-1). Performing this normalization means that if there is a great variation in the amplitude, especially if a large portion of the amplitudes are small and a very number are large, as is generally the case, for an object without a diffuser, then a small portion of the light will be diffracted. Another way of stating the fact that the spectrum should be uniform is that the object itself would have a delta-like auto-correlation. This relation arises from the Fourier transform connection between the Fourier-power spectrum and the auto-correlation in an object.

6.7.2 Band Limitation

Speckle is a grainy-appearance of the reconstructed image that arises from high frequencies of the original object being lost through truncation of the Fourier spectrum. Because diffusers spread light about the Fourier plane, it is an optimization criteria that is quite important that this spreading not be too strong, if it is too strong, light will be spread outside of the area that can be captured by the hologram and therefore speckle will be caused. The speckle cab also be reduced by reconstruction techniques in addition to the careful design of the original diffuser. The first technique is to replicate the hologram. Replicating the hologram on a regular grid will cause the reconstruction to be point-like. This point-like object will lack speckle because speckle occurs in the regions between the sample points. The image can then be resmoothed using illumination that does not spread the sample valued to great a distance retaining a sort of lumpy appearance. The other way is to use partially coherent illumination in the reconstruction because this will minimize the amount of interfacing between the samples. This will give a smooth reconstruction.

6.7.3 Redundancy

Redundancy in information storage can have two significant values. One is that the hologram may be damaged and still reconstruct an image. In fact, interferometric holograms have been demonstrated that can be broken into small pieces and when the small pieces reconstructed; the image is within slightly decreased resolution reconstruction caused by loss of small compact areas in the hologram. By spreading the information around, these burst errors have a reduced effect on the reconstruction quality of the hologram.

6.7.4 Diffusers for Interferometric Holograms

Methods for applying diffusers to smooth objects in interferometric holography use Scotch Magic Tape which is a good diffuser or a ground glass in contact with the object. The second method is to use a diffuser that is a bleached speckle pattern where the speckle pattern was formed with band-limited illumination. For reflecting objects, the surfaces can be rough or painted with diffusely reflecting paint. There are two classes of diffusers: the so-called general purpose diffusers and the objects specified diffusers. The general purpose diffusers are designed to be applicable to a large class of objects. The complex amplitude transmission of these diffusers is independent of the object being used. The object's specific diffusers, on the other hand, have values that are dependent on the object's amplitude.

6.7.5 RADAR Codes

There are a class of pulse codes that were developed for ranging applications in TADAR that were used as diffusers for computer generated holograms. The range of a target is determined by the transit time of a pulse sent from the transmitter and detected at the receiver. This pulse must have two properties that at first appear to be contradictory. First, the pulse must be narrow, more precisely; the autocorrelation of the pulse must approach that of a delta function. Secondly, the pulse power cannot be arbitrarily high. This contradiction is resolved by using pulse codes that are extended and near uniform, but whose autocorrelations are delta like.

The conversion for use as diffusers is immediate. A delta autocorrelation means that the power spectral density is uniform. The biggest difference is that the diffusers are two dimensional while RADAR codes are one dimensional.

6.7.6 Modern Methods

The Schroeder, Frank-Heimiller and Colabro-Wolf codes are examples of traditional radar-based diffusers. The more modern approach is one of optimization. For an optimization approach, it is generally necessary to quantize the diffuser phase so that it can be changed in a certain step-size during the optimization procedure and to choose an optimization criterion in the discrete Fourier transform. For example

- Simulated Annealing
- Iterative (Gerchberg-Saxton)

6.7.7 Simulated Annealing

A simple optimization algorithm is as follows:

1. Choose step size = $\Delta\phi$.

2. Choose uniformity criterion for the DFT, e.g., :
$$\frac{\langle |U|^2 \rangle - \langle |U| \rangle^2}{\langle |U| \rangle^2},$$

$\langle \rangle$ = spatial average.

- Start with some selection
- Randomly choose point
- Randomly choose the sign for a phase change $\Delta\phi$.
- DFT
- Check for uniformity using the criterion
- If the present DFT is more uniform than the previous iteration, accept the change; otherwise reject
- Loop to “Randomly choose point”
- Repeat for many iterations.

A slightly more sophisticated approach is to use an annealing schedule. At the following steps:

- Accept change with a certain probability even if the change is in the wrong direction.
- Decrease that probability with it iteration number $e^{-n/kT}$
- Guarantees global optimum at $T \rightarrow \infty$

These additional steps guarantee that a global optimum is attained as the temperature approaches infinity. Unfortunately this also means the number of operations approach infinity. Among the best methods of determining a diffuser for a particular object is the Lesem-Jordan-Hirsh (Gerchberg-Saxton) procedure.

6.7.8 Gerchberg-Saxton Procedure

Gerchberg and Saxton came from electron microscopy and applied the method to the problem of phase retrieval. That problem is to calculate the complex amplitude on the basis of information of the modulus in the object and the modulus of its Fourier transform. The algorithm starts by assigning a

random phase distribution to the object. It proceeds by ping ponging between direct space and Fourier space and in each space setting the magnitudes to the proper values while retaining the phase from the iteration cycle. There are important modifications to this procedure. The Fienup modification is one of the most important. Instead of violently returning the amplitudes to their values, a relaxation procedure is used. This relaxation procedure allows the amplitude values to slowly approach their proper values through never, of course, never attaining the values. As in any iterative algorithm, it is important to look for oscillations. Improvements can be obtained by averaging values of oscillating cycles. An algorithm due to Papoulis aims at obtaining a Fourier spectrum of bounded support. It can be combined with the Gerchberg-Saxton algorithm to suppress speckle. Phase quantization can also be included.

7. INTERFEROMETRIC DISPLAY HOLOGRAMS

Many of the examples we have used in describing computer generated holograms have been for display. The applications and examples have been rather simple and because of this simplicity, the amount of computation has been manageable. However, for real world display tasks, the "straight forward" method lead to unmanageably large computational loads. For this reason, various forms of information reduction must be introduced. We will begin by discussing limitations on the resolution and viewing angles. We proceed from there to more efficiently modeling reflection and transmission by use of shading models. We will then look at CGH multiplexing and the use of such multiplex holograms for color display. Finally we will look at real time display and video applications. From the time that holograms were invented, they have been used for the display of images. Three-dimensional objects are easily and naturally recorded in holograms. A common object was the model train for small 4"x5" holograms recorded on a glass photographic plate .

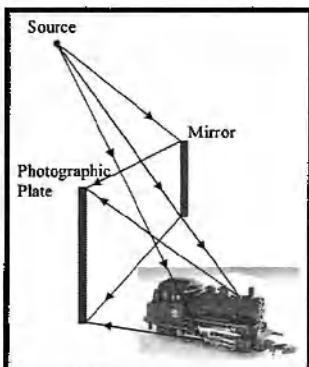


Figure 1-10: A common recording setup.

The diffusely reflecting object is illuminated by the same laser beam as is the photographic plate. After being developed, the hologram is viewed on illumination by reconstruction beam that is identical to the reference beam as shown in Figure 1-10. A Fresnel hologram is illustrated in the two figures. For CGH real-time display, holovideo, or 3-D TV, an image-plane hologram would be a more appropriate analogy. As is shown in Figure 1-8, the recording is done by imaging the object at the photographic plate and combining it there with a reference wave. On reconstruction, the object appears to be embedded in the hologram, part of it behind the hologram, and part in front of the hologram.

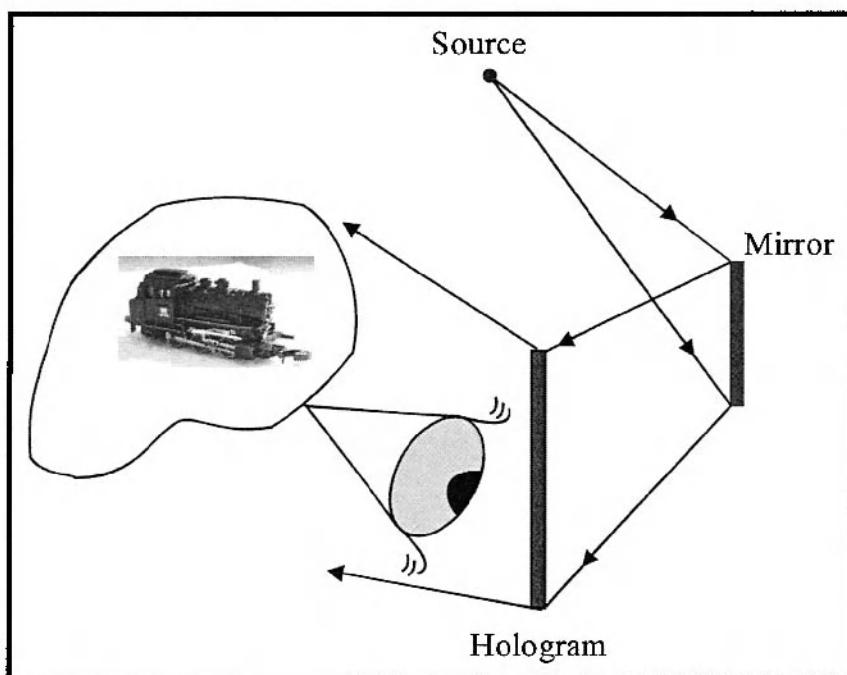


Figure 1-11: Reconstruction setup.

8. INFORMATION CONTENT OF SCENES

Computer generating a 3-D display hologram requires different considerations than those for interferometric holograms. Let's start with the display device. The resolution capability of photographic materials used for holography is in excess of 2000 lp/mm. In terms of a digital display, a 4"x5", 100mm x 125mm, plate, at 2000 lp/mm, has the equivalent of 400,000 x 500,000 pixels. In order to compare these numbers with a digital display, we need to take into account wasted space. These pixels accommodate the true image, twin image, D.C. spike, and intermodulation noise, that total about four times the wide of the true image. If the digital display can control complex amplitude directly, it need only a half the number of pixels. Considering a 5:4 aspect ratio we are down to about 250,000x200,000 pixels for an image quality comparable with a good interferometric hologram. Light modulators suitable for CGH's have more modest capabilities.

Let's approach this calculation in another way. Suppose that a scene can be described by NxN lateral pixels, and in addition, N vertical viewing perspectives, and N horizontal viewing perspectives. For a television (not HDTV) quality picture the lateral dimension would be N=512 we see that the total amount of information incorporated into CGH is going to be about $N^4 = 2^{36}$ samples. In decimal numbers this becomes approximately 10^{11} or 100 Giga-samples. That would be approximately a 300,000 x 300,000 sample CGH. This number is quite close to the number we estimated from the analogy interferometric hologram recording.

Yet another estimate can be gained from considering the 3-D image being N^2 lateral pixels and having a number of depth planes, say N. This calculation gives a much smaller number of N^3 , or for our example $512^3 = 2^{27}$ or about 10^9 samples corresponding to a 3×10^4 by 3×10^4 pixel CGH. The difference is that we have reduced the effective number of perspective views.

8.1 Information Reduction

One of the easiest ways to reduce the amount of information contained in a CGH is to eliminate the vertical perspective. That is, to perform the calculations in such a way that vertical motion of the head does not change the view scene whereas moving from side-to-side or using stereo vision does change the perspective.

If we eliminate the vertical perspective, for instance by simply replicating a single row of CGH's in the vertical direction, we eliminate a factor of N

thereby reducing the number to 2^{27} or approximately 10^8 which could be done today through it would require specialized resources. With a further reduction of the number of lateral perspectives to 32 we come to 2^{23} or 10^7 or a $3,000 \times 3,000$ pixel hologram a strenuous but realistic number today even on a well-equipped personal computer with high-quality output peripherals.

8.2 Shading Models

In the exact calculations models of wave propagation we have discussed, the surface detail is what determines properties of reflection from the objects. This reliance on microstructure to determine microscopic properties of the object needs to unnecessarily large computational complexity. For instance, we may use the statistical distribution of the random phase on a diffusely reflecting surface and yet be only interested in the surface having a particular, for instance, lambertian reflecting property. We may also regulate the amount of light transmission and scattering pixel-by-pixel, or we may use other methods to impart these reflection and transmittance properties. One way of doing the calculations is to use computer graphics shaded perspective display calculations to determine the view from several perspectives and then to encode these different perspectives as parts of these computer generate holograms.

Two shading models are in 3-D perspective display (Goudraud shading and Phong shading.) A common technique in displaying these images is to break the surface into facets or tiles. Each tile is then shaded according to the type and location of the illumination source, surface normal function and the viewing position. For Goudraud shading, luminance values at the vertices are calculated using the surface normals assigned to the vertices of the facets. These luminance values are then interpolated across the facet. Phong shading performs a different type of interpolation in that it interpolates the (vector) surface normals form the vertices across the facet. The surface-normal function later used to determine the luminance.

Chromatic dependence can be added by assigning color separation values to the vertices.

8.3 Multiplexing and Color Display

Displaying color images reduces to separating the images we wish to see into various color components, usually three: red, green, blue, and then displaying these separated images in registration that is properly aligned to one another. Multiplex computer generated holograms provide a convenient way to bring the color separated images into registration is by having the

different illumination waves incident on the hologram at different angles. Each of the illumination waves incident on the hologram at different angles. Each of the illumination waves will reconstruct all three of the color separation images, but as long as the support of each image is bounded, the illumination can be adjusted so that the proper images overlap.

8.4 Realtime Display

The bar is raised considerably when video applications are considered. Going back to our television quality static display example, we calculated a necessity of 10^7 samples per frame. For progressive scan display, we would desire at 60 frames per second. This gives 6×10^8 samples per second. Without considering possible compression schemes and assuming that we will need 16 bits per sample to represent the (monochrome) complex amplitude transmittance. We have arrived at a 10^{10} bit/second data stream or a 20 GHz bandwidth. Many economics would be necessary in order to reduce this number to a manageable value.

REFERENCES

- Bracewell, R. N., 1999, *The Fourier Transform and its Applications* (3rd edition, revised), McGraw-Hill, New York.
- Bragg, W. L., 1939, "A new type of x-ray microscope", *Nature* **143**: 678-678.
- Brown, B. R., Lohmann, A.W., 1966, "Complex Spatial Filtering with Binary Masks", *Appl. Opt.* **5**: 967-969.
- Bryngdahl, O., Wyrowski, F., 1990, "Digital holography/computer-generated holograms", *Progress In Optics*, E. Wolf, ed., North-Holland, Amsterdam, Vol. 28, pp. 1-86.
- Gabor, D., 1948, "A new microscopic principle", *Nature* **161**: 777-778.
- Gaskill, J. D., 1978, *Linear Systems, Fourier Transforms, and Optics*, Wiley, New York.
- Goodman J. W., 1996, *Introduction to Fourier Optics* (2nd edition), McGraw-Hill, New York.
- Lee, W.-H., 1978, "Computer-generated holograms: Techniques and applications in *Progress Optics XVI*", E. Wolf (ed.), Elsevier, Amsterdam", pp. 121-231.
- Leith, E. N., Upatnieks, J. , 1962, "Reconstructed wavefronts and communication theory", *J. Opt. Soc. Am.* **52**: 1123-1130.
- Lesem, L. B., Hirsch, P. M. , Jordan Jr., J. A., 1969, "The Kinoform: A New Wavefront Reconstruction Device", *IBM J Res. Dev.* **13**: 150-155.
- Lohmann, A., 1956, „Optische Einseitenbandübertragung angewandt auf das Gabor-Mikroskop (Single-sideband transmission applied to the Gabor Microscope)", *Optica Acta* **3**: 97-99
- Rogers, G. L., 1950, "Gabor diffraction microscopy. The hologram as a generalized zone plate", *Nature* **166**: 237.

Chapter 2

DIGITAL HOLOGRAPHY AND MULTI-WAVELENGTH INTERFERENCE TECHNIQUES

Myung K. Kim, Lingfeng Yu, and Christopher J. Mann

Dept. of Physics, University of South Florida, Tampa, FL 33620

Abstract: Recent developments in digital holography techniques are presented. The Fourier transform and convolution methods to calculate the diffraction fields are compared with the angular spectrum method, which allows short reconstruction distances down to zero with flexibility in the control of image noise. Examples of applications of digital holography in biological microscopy are presented with submicron lateral resolution. The phase imaging digital holography using two wavelengths is described that achieves unambiguous phase unwrapping, even for noisy images common in biological imaging. The wavelength scanning digital interference holography (WSDIH) is described that allows tomographic imaging by numerical superposition of many holograms using a series of different wavelengths to synthesize short coherence length. The WSDIH technique is applied to image sub-surface structures of an animal tissue down to 1.5 mm depth with 10 μm axial resolution.

Key words: Computer holography; Tomography; Three-dimensional image acquisition; Phase contrast microscopy; Holographic interferometry.

1. INTRODUCTION

Dennis Gabor invented holography in 1948 while attempting to improve the resolution of electron microscopy. At the time however his invention could not be made practical, as there were no sources available with the required coherence. The invention of laser and the introduction of off-axis holography provided the critical elements to make holography a practical and powerful tool for large areas of applications from metrology, data storage, optical processing, device fabrication, and even fine arts. On the

other hand, the conventional process of holography using photographic plates is time-consuming and cumbersome. Real time process is not feasible unless one uses photorefractives and other nonlinear optical materials. Recently, the field of holography has been undergoing another paradigm shift by electronic image capture using CCD array cameras and digital processing of the holographic images[1,2,3].

By recording the phase as well as intensity of light wave, holography allows reconstruction of the images of three-dimensional objects, and gives rise to a host of metrological and optical processing techniques. With the advance of computer and electronic imaging technology, it is now very practical and often advantageous to replace portions of the holographic procedures with electronic processes. In digital holography, the holographic interference pattern is digitally sampled by CCD camera and the image is numerically reconstructed by applying the results from the diffraction theory. It offers a number of significant advantages such as the ability to acquire the images rapidly, the availability of both the amplitude and the phase information of the optical field and the versatility of the processing techniques that can be applied to the complex field data. The advances in digital imaging devices such as CCD and CMOS cameras and in computational and data storage capacities have been central to the widening applications of digital holography. In 1967, J. Goodman demonstrated the feasibility of numerical reconstruction of holographic images from a digitized vidicon image[4]. Schnars and Jueptner, in 1994, were the first to use a CCD camera connected to a computer as the input[1]. Since then, developments of digital holographic techniques and applications have been gaining pace ever more rapidly. Microscopic imaging by digital holography has been applied to imaging of microstructures[5,6] and biological systems[7,8,9,10]. In digital holography, the phase of the optical field, as well as the amplitude, results directly from the numerical diffraction of the optically recorded holographic interference pattern and leads to images of optical thickness profile with a precision of a fraction of a wavelength[11,12,13,14]. This can be utilized for numerical corrections of various aberrations of the optical system such as field curvature and anamorphism[15]. In microscopy applications[15], the reconstructed image can be numerically focused to any plane in the object. Although the common preconception is that most coherent imaging techniques such as holography has the degrading effect of coherent noise, through careful control of laser beam and other optical quality, remarkably clean images can be obtained. This is especially true with phase imaging digital holography because of its relative immunity to amplitude or phase noise of the laser profile. Numerous techniques have been developed to exploit the unique advantages of digital holography, such as the ability to reconstruct along any

inclined planes[16,17], to transmit holographic data over communication channels[18], and to even generate holograms of fluorescence patterns[19].

We present recent developments in digital holography techniques carried out in our laboratory. In Sec. 2, various methods to calculate the holographic diffraction fields are described and compared, and the effectiveness of the angular spectrum method is particularly emphasized. In Sec. 3, several examples of applications of digital holography in biological microscopy are presented. In Sec. 4, the phase imaging digital holography using two or more wavelengths is described that achieves unambiguous phase unwrapping by optical method. Section 5 describes the wavelength scanning digital interference holography that allows tomographic or optical section imaging by numerical superposition of many holograms using a series of different wavelengths to synthesize short coherence length. Section 6 gives concluding discussions and remarks.

2. DIGITAL HOLOGRAPHY PRINCIPLES

A common method for calculating the propagation of the optical field is based on the Fresnel diffraction theory. Let $E_0(x_0, y_0)$ be the two-dimensional pattern of the optical field on a plane. Then the field $E(x, y)$ at another plane a distance z away is given by

$$E(x, y) = E_0 \oplus S(x, y; z) \quad (2-1)$$

where the point spread function (PSF) is

$$S(x, y; z) = -\frac{ik}{2\pi z} \exp\left[ik\sqrt{x^2 + y^2 + z^2}\right]. \quad (2-2)$$

The convolution can be numerically calculated by three fast Fourier transforms, or the PSF can be simplified by the Fresnel approximation as

$$S(x, y; z) = -\frac{ik}{2\pi z} \exp\left[ikz + \frac{ik}{2z}(x^2 + y^2)\right], \quad (2-3)$$

which allows the calculation to be carried out as a single Fourier transform.

Suppose the object consists of a point source located at (X_o, Y_o) on the object plane (x_o, y_o) a distance z_o from the hologram plane H :

$$E_o(x_o, y_o) = \varepsilon_o \delta(x_o - X_o, y_o - Y_o). \quad (2-4)$$

The object field at the H plane is a spherical wave:

$$E_{ho}(x_h, y_h) = \varepsilon_o \exp \left\{ \frac{ik}{2z_o} \left[(x_h - X_o)^2 + (y_h - Y_o)^2 \right] \right\}. \quad (2-5)$$

The reference field at the H plane is a plane wave, incident at an angle from the z-axis:

$$E_{hr}(x_h, y_h) = \varepsilon_r \exp \left[i(k_x x_h + k_y y_h) \right] \quad (2-6)$$

where $k_x = \mathbf{k} \cdot \hat{\mathbf{x}}$ and $k_y = \mathbf{k} \cdot \hat{\mathbf{y}}$. The total field at the H plane is $E_h(x_h, y_h) = E_{ho} + E_{hr}$ and the intensity is:

$$\begin{aligned} I_h(x_h, y_h) &= |E_h|^2 \\ &= |\varepsilon_r|^2 + |\varepsilon_o|^2 \\ &\quad + \varepsilon_r \varepsilon_o^* \exp \left\{ -\frac{ik}{2z_o} \left[(x_h - X_o)^2 + (y_h - Y_o)^2 \right] + i(k_x x_h + k_y y_h) \right\} \\ &\quad + \varepsilon_r^* \varepsilon_o \exp \left\{ \frac{ik}{2z_o} \left[(x_h - X_o)^2 + (y_h - Y_o)^2 \right] - i(k_x x_h + k_y y_h) \right\} \end{aligned} \quad (2-7)$$

In digital holography, the intensity pattern is sampled at $(x_h, y_h) = \{(\alpha\Delta, \beta\Delta); \alpha, \beta = 0, 1, 2, \dots, N_x - 1\}$ by the CCD array of $N_x \times N_x$ pixels and effective size $a_x \times a_x$ with $a_x = N_x \Delta$:

$$\begin{aligned} I_h(\alpha, \beta) &= |\varepsilon_r|^2 + |\varepsilon_o|^2 \\ &\quad + \varepsilon_r \varepsilon_o^* \exp \left\{ -\frac{ik}{2z_o} \left[(\alpha\Delta - X_o)^2 + (\beta\Delta - Y_o)^2 \right] + i(\alpha k_x + \beta k_y) \Delta \right\} \\ &\quad + \varepsilon_r^* \varepsilon_o \exp \left\{ \frac{ik}{2z_o} \left[(\alpha\Delta - X_o)^2 + (\beta\Delta - Y_o)^2 \right] - i(\alpha k_x + \beta k_y) \Delta \right\} \end{aligned} \quad (2-8)$$

The first two terms are the zero-order terms due to the reference and object beams, which can be eliminated by a number of different methods. The use of phase-shifting digital holography removes the zero order and twin images through multi-exposure holographic recording while shifting the phase of reference field by an integer fraction of 2π [20,21]. Off-axis holography spatially separates the holographic images away from the undiffracted zero-order. One can also take separate exposures of reference and object waves and subtract these from the original holographic exposure, which is effective in reducing any noise from these terms. On the other hand, the off-axis configuration reduces the available spatial frequency bandwidth by half. The last two terms in (2-8) are the holographic twin images, and these can be separated if the off-axis angle of the reference beam is large enough. We take the third term as the holographic image term. Numerical reconstruction of the holographic image starts with a multiplication by a conjugate reference field:

$$\begin{aligned} H(\alpha, \beta) &\equiv I_h(\alpha, \beta) \cdot E_{hr}^*(\alpha, \beta) \\ &= |\varepsilon_r|^2 \varepsilon_o^* \exp \left\{ -\frac{ik}{2z_o} \left[(\alpha\Delta - X_o)^2 + (\beta\Delta - Y_o)^2 \right] \right\} \end{aligned} \quad (2-9)$$

Numerical Fresnel diffraction is applied to this expression over a distance of z_i to obtain the holographic image over a grid of the same size and resolution as the CCD array $(x_i, y_i) = \{(\gamma\Delta, \delta\Delta); \gamma, \delta = 0, 1, 2, \dots, N_x - 1\}$:

$$\begin{aligned} E_i(\gamma, \delta) &= H \oplus S(\gamma, \delta; z_i) \\ &= -\frac{ik}{2\pi z_i} \exp(ikz_i) \\ &\times \sum_{\alpha, \beta=0}^{N_x-1} \Delta^2 H(\alpha, \beta) \exp \left\{ \frac{ik\Delta^2}{2z_i} \left[(\alpha - \gamma)^2 + (\beta - \delta)^2 \right] \right\} \end{aligned} \quad (2-10)$$

The image at $z_i = z_o$ is:

$$\begin{aligned}
E_i(\gamma, \delta) = & -\frac{ik\Delta^2}{2\pi z_i} |\varepsilon_r|^2 \varepsilon_o^* \exp(ikz_i) \\
& \times \exp \left\{ \frac{ik}{2z_o} \left[(\gamma^2 \Delta^2 - X_o^2) + (\delta^2 \Delta^2 - Y_o^2) \right] \right\} \\
& \times \sum_{\alpha, \beta=0}^{N_x-1} \exp \left\{ -\frac{ik\Delta}{z_o} [\alpha(\gamma\Delta - X_o) + \beta(\delta\Delta - Y_o)] \right\}
\end{aligned} \tag{2-11}$$

$$\begin{aligned}
= & -\frac{ik\Delta^2}{2\pi z_i} |\varepsilon_r|^2 \varepsilon_o^* \exp(ikz_i) \\
& \times \exp \left\{ \frac{ik}{2z_o} \left[(\gamma^2 \Delta^2 - X_o^2) + (\delta^2 \Delta^2 - Y_o^2) \right] \right\} \\
& \times \exp \left\{ -\frac{ika_x}{2z_o} [(\gamma\Delta - X_o) + (\delta\Delta - Y_o)] \right\} \\
& \times \frac{\sin \frac{ka_x}{2z_o} (\gamma\Delta - X_o)}{\sin \frac{k\Delta}{2z_o} (\gamma\Delta - X_o)} \cdot \frac{\sin \frac{ka_x}{2z_o} (\delta\Delta - Y_o)}{\sin \frac{k\Delta}{2z_o} (\delta\Delta - Y_o)} \\
\approx & -\frac{ika_x^2}{2\pi z_i} |\varepsilon_r|^2 \varepsilon_o^* \exp(ikz_i) \delta_{\gamma, X_o/\Delta} \delta_{\delta, Y_o/\Delta}
\end{aligned}$$

The Kronecker delta in the last line follows from the $\sin N\theta / \sin \theta$ factors, familiar from the theory of diffraction by a grating. The ‘width’ of the Kronecker delta is $d_\gamma = \frac{2\lambda z_o}{N_x \Delta^2}$ pixels: i.e., d_γ pixels for $\frac{ka_x}{2z_o} (\gamma\Delta - X_o)$ to span $-\pi$ to $+\pi$. The discrete summation is in fact periodic and can lead to aliasing unless $\frac{k\Delta}{2z_o} (\gamma\Delta - X_o) < \pi$ for all γ , which is satisfied if

$$z_o > \frac{a_x^2}{N_x \lambda} = z_{\min} \tag{2-12}$$

This sets the minimum object-to-hologram distance, to avoid aliasing in the Fourier transforms.

The pixel resolution of the reconstructed images determined from the Fresnel diffraction formula as a single Fourier transform, varies as a function of the reconstruction distance z as $\Delta_i = \lambda z / (N_x \Delta_0)$, where Δ_0 and Δ_i are the resolutions of the hologram and the reconstructed image, respectively. In order to get consistent resolution, the Fresnel diffraction can also be implemented as a convolution, as described above. However, both of the above approaches assume the Fresnel approximation, which limits a minimum reconstruction distance z . By using the angular spectrum algorithm, the problems associated with the Fresnel diffraction formula are avoided. From Fourier optics[23], if $E(x, y; 0)$ is the object wave field at plane $z = 0$, the corresponding angular spectrum of the object wave at this plane is the Fourier transform:

$$F(k_x, k_y; 0) = \iint E(x, y; 0) \exp[-i(k_x x + k_y y)] dx dy, \quad (2-13)$$

where k_x and k_y are corresponding spatial frequencies of x and y . The object angular spectrum $F(k_x, k_y; 0)$ can be separated from other spectral components of the hologram with a numerical band-pass filter if the off-axis angle θ of the incident beam is properly adjusted. The object field $E(x, y; 0)$ can be rewritten as the inverse Fourier transform of its angular spectrum, properly filtered,

$$E(x, y; 0) = \iint F(k_x, k_y; 0) \exp[i(k_x x + k_y y)] dk_x dk_y. \quad (2-14)$$

The complex-exponential function $\exp[i(k_x x + k_y y)]$ may be regarded as a projection, on to the plane $z = 0$, of a plane-wave propagating with a wave vector (k_x, k_y, k_z) , where $k_z = [k^2 - k_x^2 - k_y^2]^{1/2}$ and $k = 2\pi/\lambda$. After propagating along the z axis to a new plane, the field distribution is

$$E(x, y; z) = \iint F(k_x, k_y; 0) \exp[i(k_x x + k_y y + k_z z)] dk_x dk_y. \quad (2-15)$$

This is the angular spectrum method and the resolution of the reconstructed images is the same as that in the hologram plane. In spite of the apparent differences, the angular spectrum method will yield identical predictions of diffracted field as the first Rayleigh-Sommerfeld solution[23]. However, as an approximate Rayleigh-Sommerfeld solution, Fresnel diffraction is not capable of reconstructing the wave field near to the hologram plane.

3. DIGITAL HOLOGRAPHIC MICROSCOPY EXPERIMENTS

The experiments are performed with modified Mach-Zhender interferometers, illustrated in Fig. 2-1. In reflective geometry (Fig. 2-1a), a collimated beam from the laser is split into the object and reference beams by the beam splitter BS1. The object beam is focused by the lens L2 on to the point F2, which is also the front focal point of the objective lens L3. Thus the object is illuminated by a collimated beam. An aperture A is placed at the conjugate point of the object with respect to L3, so that the aperture is imaged on the object and the illumination is confined to an area of the object that is being imaged by the holographic system. This is necessary to avoid light scattered from the surrounding area of the object entering the camera contributing to the noise of the imaging system. The laser light is reflected by the object and travels toward the camera, which is placed at the conjugate plane of the plane H with respect to L3. The reference beam's focus F1 is equidistant from BS3 as F2, so it is optically equivalent to a collimated beam incident from the left of objective L3. The object and the reference beams are tilted with respect to each other in an off-axis hologram arrangement. Figure 2-1b) illustrates another apparatus for transmissive geometry using a pulsed laser. A magnified image of an object specimen is projected on the CCD camera, as well as the reference beam. A pair of similar microscope objectives (L1 and L2) in the two optical branches are used to match the curvature of the two wavefronts. A digital delay generator (DDG) triggers both the laser and the camera.

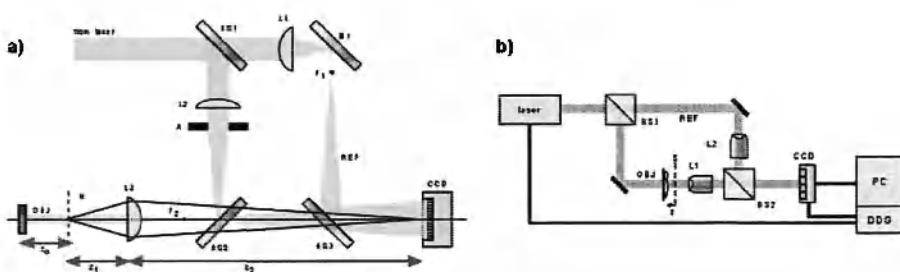


Figure 2-1. a) Apparatus for reflective digital holography. BS's: beam splitters and combiner; L's: lenses; F's: focal points; M1: mirror; A: aperture; OBJ: object; REF: reference; H: hologram plane, imaged to CCD plane; CCD: camera. b) Apparatus for transmissive digital holography. DDG: digital delay generator; PC: computer.

The digital holography experiment proceeds as follows, described using an example set of parameters. The wavelength is $\lambda = 0.532 \mu\text{m}$. An area of the object is chosen for imaging with $a_x = 88 \mu\text{m}$ and $N_x = 360$, so that $\Delta = 0.24 \mu\text{m}$. This gives $z_{\min} = 40 \mu\text{m}$, and z_o is set at this value. This area of H-plane is imaged by a microscope objective on to the CCD with lateral magnification $M = 37$, to correspond to the physical pixel size of $M\Delta = 9 \mu\text{m}$ and $Ma_x = 3.3 \text{ mm}$. The resolution of the imaging system given by Rayleigh criterion is $d = 1.22\lambda / 2N.A. = 0.82 \mu\text{m}$, using an $f = 8.55 \text{ mm}$ microscope objective lens with $N.A. = 0.4$. This corresponds to $Md = 30 \mu\text{m}$ at the CCD. On the other hand, the width of the PSF of the digital reconstruction is $d_r\Delta = 0.48 \mu\text{m}$, which again is smaller than the optical PSF d , and therefore the resolutions of the CCD array and the numerical reconstruction are all sufficient for the given optical system. The optical system described here provides a straightforward means for high resolution holographic microscopic imaging. There is no need for elaborate processing such as magnification by using reconstruction wavelength long compared to recording wavelength[3], which inevitably introduces aberration, or using an aperture array in front of the camera and scanning it to artificially increase the CCD resolution[22].

Figure 2-2 shows an $88 \times 88 \mu\text{m}^2$ area of a USAF resolution target. The element 6 of group 7 has $2.2 \mu\text{m}$ bar width, and the resolution of the image appears somewhat better than that, say $1 \mu\text{m}$. The holographic amplitude image, Fig. 2-2c), is quite indistinguishable from the direct image, Fig. 2-2a). Quantitative phase imaging is particularly effective in digital holography – one only needs to plot the phase of the calculated complex optical field, as shown in Fig. 2-2d). The thickness of the chrome coating on the glass plate of the resolution target is easily measured from the phase image to be about 75 nm.

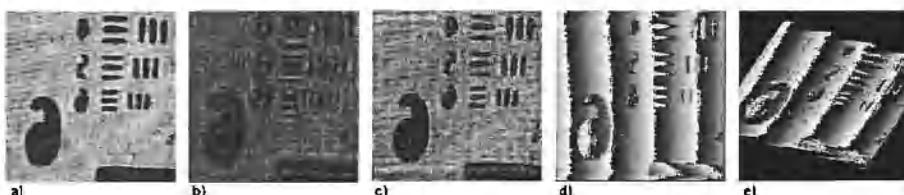


Figure 2-2. High resolution holographic image reconstruction of resolution target of area $88 \times 88 \mu\text{m}^2$ with 360×360 pixels. a) direct image when the object is on the H-plane, illuminated by laser; b) the hologram recorded by the camera; c) reconstructed amplitude image; d) phase image; e) phase image in perspective view.

The Fresnel transformation[7] has been commonly used to calculate the holographic diffraction, where the approximation of spherical Huygens

wavelet by a parabolic surface allows the calculation of diffraction integral using a single Fourier transform. In the Huygens convolution method[2], the diffraction integral can be calculated without such approximation using three Fourier transforms through the convolution theorem. On the other hand, the angular spectrum method[16] involves two Fourier transforms, plus simple filtering of the angular spectrum. We have applied all three methods to the reconstruction of the image of a resolution target, as shown in Fig. 2-3. A $25 \times 25 \mu\text{m}^2$ area of USAF resolution target is imaged containing the group 7 element 6 using 452×452 pixels. Figure 2-3a) is the holographic interference pattern recorded by the CCD camera, and its Fourier transform in Fig. 2-3b) is the angular spectrum. It contains three main peaks that correspond to the spectra of the zero-order and the two twin-images. One of these peaks – the highlighted rectangular area – is selected, a propagation phase factor is multiplied, and finally inverse-Fourier transformed to obtain the amplitude image in Fig. 2-3c) and the phase image in Fig. 2-3d). The individual bars are $2.2 \mu\text{m}$ wide, which are clearly resolved and consistent with diffraction-limited resolution of $0.5 \mu\text{m}$ for 0.65 NA lens. The physical thickness of film is given by $d = \lambda(\Delta\phi/2\pi)/(n - n_0)$, where λ is the wavelength, $\Delta\phi$ is the phase step, and $(n - n_0)$ is the index difference between the film and air. The thickness-equivalent noise level of the flat area is $\sim 30 \text{ nm}$. (On the film-coated opaque bar areas, the lack of light causes larger uncertainty in phase.) The phase map is rendered in pseudo-colored 3D perspective in Fig. 2-3i). Especially notable in the phase map is the lack of the coherent noise conspicuous in the amplitude image and prevalent in most other holographic imaging methods. The amplitude and phase images obtained from the Huygens convolution method are shown in Fig. 2-3e) and 2-3f), as well as those obtained from the Fresnel method in Fig. 2-3g) and 2-3h). The main reason for the obvious degradation of these images, calculated from the same hologram, is the insufficient off-axis angle to separate out the zero-order component. The effect is most detrimental in the Fresnel images, where part of the holographic image is buried in the zero-order background, and its phase image is completely scrambled. The effect shows up as the spurious interference patterns in the convolution images. While the minimum off-axis angle must be strictly satisfied in order to avoid the zero-order intrusion in Fresnel or convolution methods, the control and removal of zero-order component is straightforward and flexible in the angular spectrum method. Another potential problem of the Fresnel and convolution methods is that they require a minimum hologram distance to avoid aliasing, whereas the angular spectrum method does not have such minimum and the image can be calculated even at zero distance. The minimum distance is given by $z_{\min} = a^2 / N\lambda$ where a is the size of the hologram and N is the number of pixels.

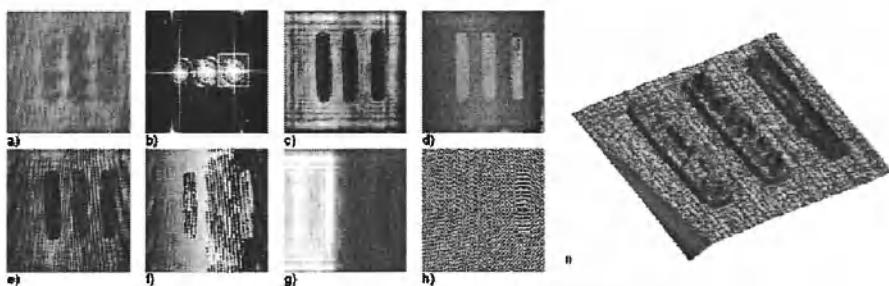


Figure 2-3. Holography of a resolution target.

Some examples of digital holography in biological microscopy are presented below. In Fig. 2-4, a $100 \times 100 \mu\text{m}^2$ area of a layer of onion cells is shown. The cell walls are sharply focused in the amplitude image and the phase image shows accurate representation of optical thickness, modulo wavelength, of the cell bodies. The phase-unwrapped image is generated using a software algorithm[24]. Figure 2-5 displays digital holography of a paramecium. Slightly different directions of reconstruction reference waves are used in the phase images c) and d), which may be useful for emphasizing different aspects of the microscopic images. Numerical focusing is another unique capability of digital holography, where a single hologram is used to calculate the optical field at any number of image planes, emulating the focusing control of a conventional microscope. In Fig. 2-6, a paramecium as well as a euglena (the thin rod shape in the upper right of the field) are within field of view but separated by about $400 \mu\text{m}$ of axial distance.

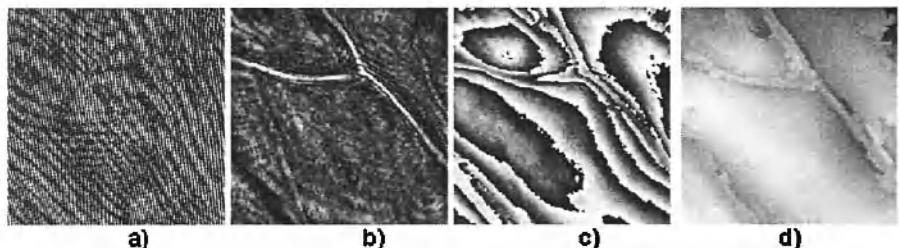


Figure 2-4. Holography of a layer of onion cells. Images, $100 \times 100 \mu\text{m}^2$, of a) hologram, b) amplitude, c) wrapped phase, and d) unwrapped phase.

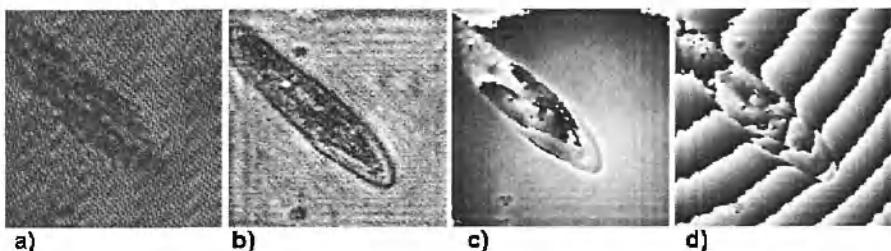


Figure 2-5. Holography of a paramecium, $250 \times 250 \mu\text{m}^2$, showing a) hologram, b) amplitude, c) and d) phase images.



Figure 2-6. Numerical focusing in digital holography of paramecia.

We present two examples of high-resolution quantitative phase-contrast microscopy studies by angular spectrum digital holography. Figures 2-7 and 2-8 show images of SKOV-3 ovarian cancer cells: confluent group of cells in Fig. 2-7 and two isolated cells in Fig. 2-8. In each of the two cases, the panels display a) Zernike phase contrast image, b) holographic amplitude and c) phase images, and d) phase image unwrapped by a software algorithm. Pseudocolor 3D rendering of d) is shown in e). The image size is $60 \times 60 \mu\text{m}^2$ with 404×404 pixels. Figure 2-8 is a particularly good demonstration of the level of image resolution and fidelity that can be obtained by the present technique, displaying the nuclear membranes and chromosomes. The overall height of the cell is calculated to be about $2.8 \mu\text{m}$, with the assumption of the average index of the cell to be 1.375. Thickness of the lamelipodium around the edge of the cell is determined to be about 320 nm. The noise level in the substrate area is 60 nm, which may be partly due to the residues from fixing of the cells. These images represent a new level of capabilities in quantitative phase contrast imaging that is not available in current technology, in terms of the simplicity of principle and apparatus, quantitative accuracy of the representation, and versatility of processing techniques that can be applied.

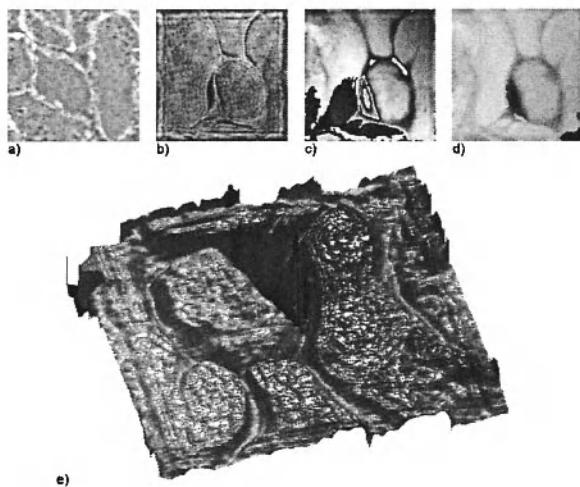


Figure 2-7. Holography of confluent SKOV-3 ovarian cancer cells. a) Zernike phase contrast image; b) holographic amplitude and c) phase images; d) unwrapped phase image; e) 3D pseudocolor rendering of d).

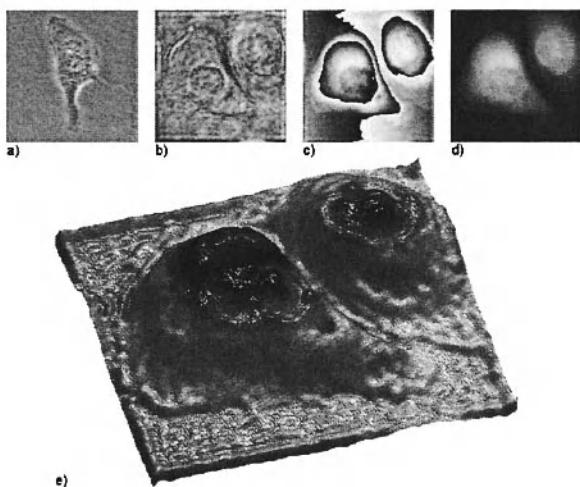


Figure 2-8. Holography of non-confluent SKOV-3 cells. a) Zernike phase contrast image; b) holographic amplitude and c) phase images; d) unwrapped phase image; e) 3D pseudocolor rendering of d).

4. MULTI-WAVELENGTH PHASE IMAGING DIGITAL HOLOGRAPHY

The digital holography is thus capable of high resolution microscopic imaging. In particular, the availability of the phase image can be of significant advantage with high precision of optical thickness profiles. On the other hand, the phase images contain 2π -discontinuities for objects of optical depth greater than the wavelength. Most of the phase unwrapping algorithms for removal of the discontinuities require subjective assumptions for detection of phase discontinuity when there are height steps on the object that are more than a wavelength[25,26]. The computation load is substantial at several minutes per image and the results are often unsatisfactory for images with complex phase structures. We have developed a holographic phase imaging method to remove these phase discontinuities by the generation and combination of two phase maps using separate wavelengths[27]. The basic principle of multi-wavelength phase imaging is described by referring to Fig. 2-9, with numerical values that were used in generating the simulation plots. Suppose the object is a tilted plane of height $h = 5.0 \mu\text{m}$. A single-wavelength phase image has 2π discontinuity wherever the height is a multiple of the wavelength. Using wavelengths $\lambda_1 = 0.532 \mu\text{m}$ or $\lambda_2 = 0.633 \mu\text{m}$ the phase map $\varphi_m(x)$, $m = 1$ or 2 , converted to surface profile $z_m(x) = \lambda_m \varphi_m / 2\pi$ will consist of a number of ramps of height equal to the wavelength, Figs. 2-9a) and 2-9b). Subtraction of the two phase maps $\varphi'_{12} = \varphi_1 - \varphi_2$ has numerous discontinuities of 2π , Fig. 2-9c), but adding 2π wherever $\varphi'_{12} < 0$ yields a new phase map $\varphi_{12}(x) = \varphi'_{12} + 2\pi \cdot (\varphi'_{12} < 0)$, with a longer range free of discontinuities. (The parentheses in the last expression is either 1 or 0 depending on whether the inequality is true or false.) In fact the new phase map is equivalent to that of a longer ‘beat wavelength’ $\Lambda_{12} = \lambda_1 \lambda_2 / |\lambda_1 - \lambda_2| = 3.33 \mu\text{m}$, and the corresponding surface profile is the ‘coarse map’ $z'_{12}(x) = \Lambda_{12} \varphi_{12}(x) / 2\pi$, Fig. 2-9d). By proper choice of the two wavelengths, the axial range Λ_{12} can be adjusted to any value that would fit the axial size of the object being imaged[28,29,30]. This technique provides a straightforward and efficient phase imaging method in a wide range of applications. A limitation is that any phase noise in each single-wavelength phase map is amplified by a factor equal to the magnification of the wavelengths. The other half of the phase imaging method consists of an algorithm to reduce the noise back to the level of the single-wavelength phase maps. This is accomplished by using the coarse map to determine z within an integer n -multiple of a wavelength, say λ_1 , and pasting the single-wavelength phase map φ_1 to $n\lambda_1$, to obtain the ‘fine map’. Figure 2-10 shows perspective views of phase maps of the surface of a resolution target. Figures 2-10a) and 2-10b) are

single-wavelength phase maps using the two wavelengths $\lambda_1 = 0.532 \mu\text{m}$ and $\lambda_2 = 0.633 \mu\text{m}$. Figure 2-10c) is the coarse map and Fig. 2-10d) is the fine map.

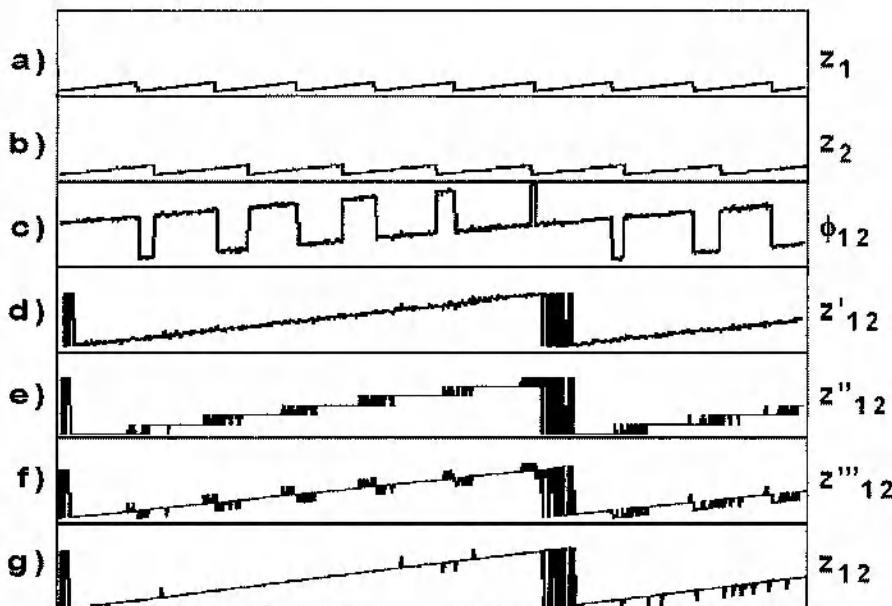


Figure 2-9. Simulation of two-wavelength phase imaging digital holography. a) height profile of a $10 \mu\text{m}$ high incline, derived from phase of $0.532 \mu\text{m}$; b) derived from phase of $0.633 \mu\text{m}$; c) difference phase map; d) coarse map, , with beat wavelength $3.33 \mu\text{m}$; e) , where is divided into integer multiples of ; f) , where is pasted on ; g) fine map, , where most of the spikes in are removed by comparison with . The vertical axis is $5.0 \mu\text{m}$ full scale in every panel, except for c) where the vertical range is -2π to $+2\pi$.

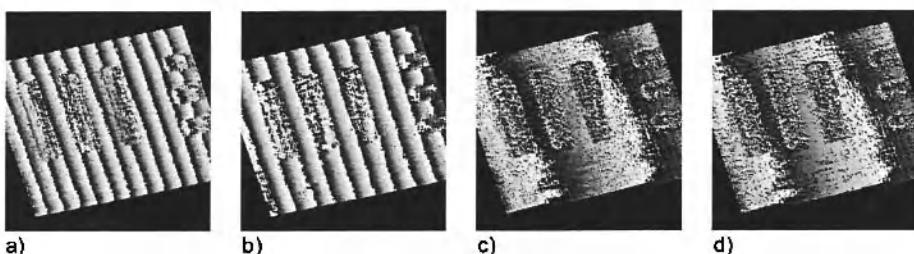


Figure 2-10. Two-dimensional profiles of a resolution target, $157 \times 157 \mu\text{m}^2$ with 360×360 pixels, from two-wavelength phase imaging digital holography. a) phase map $z_1(x, y)$ derived from phase $\phi_1(x, y)$ of $\lambda_1 = 0.532 \mu\text{m}$; b) phase map $z_2(x, y)$ derived from phase $\phi_2(x, y)$ of $\lambda_2 = 0.633 \mu\text{m}$; c) coarse map $z'_{12}(x, y)$; d) fine map $z_{12}(x, y)$.

The method of phase imaging and phase unwrapping works equally well on biological microscopy. In Fig. 2-11, a $193 \times 193 \mu\text{m}^2$ area of a layer of onion cells is imaged. The single wavelength phase images of Figs. 2-11a) and 2-11b) contain numerous 2π discontinuities, making it difficult to discern the cell body and walls. These discontinuities are completely removed in the phase-unwrapped image of Fig. 2-11c), and one can clearly observe the cell bodies delineated by the cell walls. Note, in particular, that conventional phase unwrapping technique, Fig. 2-11d), has difficulty with the quite irregular image of the onion cells shown. Furthermore, the method can be extended to three- or more-wavelength phase imaging to increase the beat wavelength further without introducing amplification of noise.

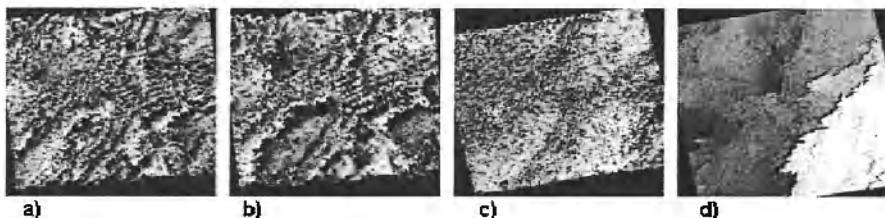


Figure 2-11. Two-dimensional profiles of onion cells, $193 \times 193 \mu\text{m}^2$ with 360×360 pixels, from two-wavelength phase imaging digital holography. a) phase map derived from phase of $0.532 \mu\text{m}$; b) phase map derived from phase of $0.633 \mu\text{m}$; c) fine map; d) unwrapped phase map using numerical algorithm.

5. WAVELENGTH-SCANNING DIGITAL INTERFERENCE HOLOGRAPHY

Although the hologram produces a 3D image of the optical field, this does not by itself yield the topographic or tomographic distance information, other than by focusing and defocusing of the image points. The distance information can be obtained by counting the number of wavelengths or some multiples of it, which is the basis of various interference techniques. A well-known technique is the interference of two holograms recorded at two different wavelengths, resulting in a contour interferogram with the axial distance between the contour planes inversely proportional to the difference in wavelengths, $\Lambda \sim \lambda^2 / \Delta\lambda$. In digital holography, it is possible to extend the process to recording and reconstruction of many holograms without introducing any wavelength mismatch. If a number, N , of regularly spaced wavelengths are used for recording and reconstruction, then the peaks of the

cosine squared intensity variation of the two-wavelength interference become sharper and narrower, $\delta = \Lambda / N$, as when a number of cosines with regularly spaced frequencies are added. We are developing wavelength scanning digital interference holography (WSDIH) as a novel method of microscopic three-dimensional imaging by numerical superposition of a number of holographic fields taken with varying wavelengths[31,32,33]. Unlike some of the other 3D microscopic methods such as confocal microscopy and optical coherence tomography[34], the digital interference holography does not involve pixel-by-pixel mechanical scanning of three-dimensional volume and yet achieves comparable resolutions. Wide field optical coherence tomography is another 3D imaging method using incoherent light interference imaging[35,36].

Suppose an object is illuminated by a laser beam of wavelength λ . A point P (at \mathbf{r}_p) on the object scatters the illumination beam into a Huygens wavelet, $A(\mathbf{r}_p)\exp(ik|\mathbf{r} - \mathbf{r}_p|)$, where $A(\mathbf{r}_p)$ is proportional to the amplitude and phase of the scattered wavelet. For an extended object, the field at \mathbf{r} is

$$E(\mathbf{r}) \sim \int A(\mathbf{r}_p)\exp(ik|\mathbf{r} - \mathbf{r}_p|) d^3\mathbf{r}_p, \quad (2-16)$$

where the integral is over the object volume. The amplitude and phase of this field at the hologram plane $z = 0$ is recorded by the hologram. If the holographic process is repeated using N different wavelengths, and the reconstructed fields are all superposed together, then the resultant field is

$$\begin{aligned} E(\mathbf{r}) &\sim \sum_k \int A(\mathbf{r}_p)\exp(ik|\mathbf{r} - \mathbf{r}_p|) d^3\mathbf{r}_p \sim \int A(\mathbf{r}_p)\delta(\mathbf{r} - \mathbf{r}_p) d^3\mathbf{r}_p \\ &\sim A(\mathbf{r}). \end{aligned} \quad (2-17)$$

That is, for a large enough number of wavelengths, the resultant field is proportional to the field at the object and is nonzero only at the object points. In practice, if one uses a finite number N of wavelengths at regular intervals of $\Delta(1/\lambda)$, then the object image $A(\mathbf{r})$ repeats itself (other than the diffraction/defocusing effect of propagation) at a beat wavelength $\Lambda = [\Delta(1/\lambda)]^{-1}$, with axial resolution $\delta = \Lambda/N$. By use of appropriate values of $\Delta(1/\lambda)$ and N , the beat wavelength Λ can be matched to the axial extent of the object, and δ to the desired level of axial resolution.

In a WSDIH experiment, the wavelength of a dye laser is scanned in the range of $575.0 \sim 605.0$ nm in 20 steps, taking the exposure of a hologram at each step. The optical field of a volume around the image location is calculated by numerical diffraction from each hologram. At this point, the

field patterns in the individual 3D arrays show little variation along a few millimeters of z-direction. Now the twenty 3D arrays are numerically superposed together, by adding the arrays elementwise, resulting in the accumulated field array of the same size. This new array then has field distribution that represents the three-dimensional object structure, with axial range of $\Lambda = 220 \mu\text{m}$ and axial resolution of $\delta = 11 \mu\text{m}$, as described above. Figure 2-12 illustrates the building up of axial resolution as a series of holographic images are superposed using a range of wavelengths. The five frames shown are with one, two, four, eight, and twenty wavelengths superposed, and one notices the narrowing of the contour widths as the synthesized coherence length shortens. Figure 2-13 shows a few contour images at $60 \mu\text{m}$ axial distance intervals.



Figure 2-12. Build up of axial resolution by superposition of holographic images of a penny using a range of wavelengths with $N = 1, 2, 4, 8$, and 20.

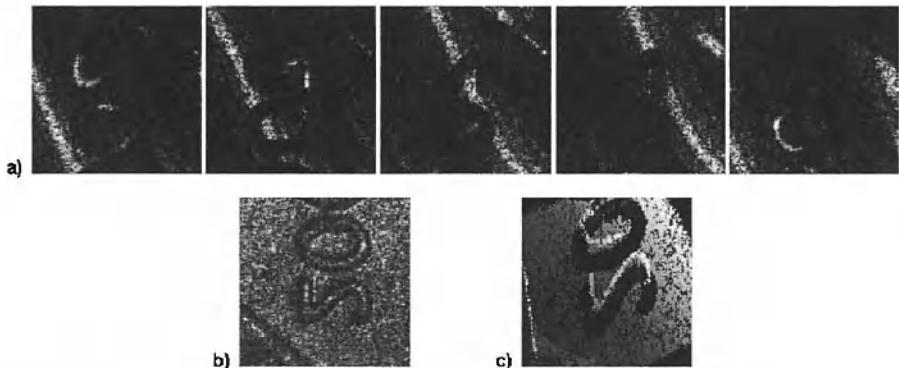


Figure 2-13. a) Contour images at various axial distances in the WSDIH image of a penny. The image volume is $2.62 \text{ mm} \times 2.62 \text{ mm} \times 200 \mu\text{m}$, and $N = 20$, so that . b) Flat view of DIH by adding all the contour images. c) Perspective view of the 3D reconstruction of the penny by WSDIH.

The digital interference holography is a coherent imaging method and is capable of tomographic imaging through a thickness of scattering medium. Figure 2-14 is the result of a DIH imaging experiment on a $2.62 \times 2.62 \text{ mm}^2$ area of a piece of beef tissue. Here we have used wavelengths in the range of $585.0 \sim 591.9 \text{ nm}$ at 41 steps so that the axial range is $\Lambda = 2.00 \text{ mm}$ and the axial resolution $\delta = 50 \mu\text{m}$. The specimen was a thin layer of beef tissue pressed to about 1.5 mm thickness on a slide glass and otherwise exposed to air. The images in Fig. 2-14(a) show tissue layers at several depths up to about 1.5 mm below surface. Much of the reflection signals are apparently from the tissue-air and tissue-glass interface. One can discern the striation of muscle fiber bundles. Figure 2-14(b) shows variations of the tissue layers in a few x-z cross-sectional images.

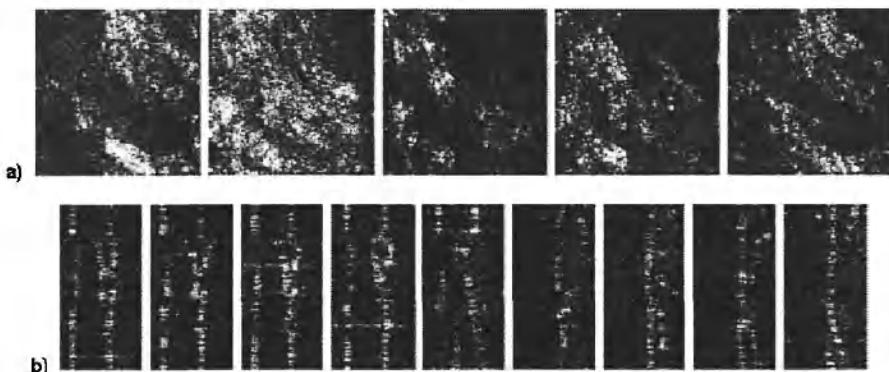


Figure 2-14. WSDIH tomography of a beef tissue. The image volume is $2.62 \text{ mm} \times 2.62 \text{ mm} \times 2.0 \text{ mm}$, and $N = 30$, so that $\Delta z = 66.7 \mu\text{m}$. a) x-y transverse images at several depths up to 1.5 mm. b) x-z cross-sectional images displaying variations of tissue layers across the field.

6. CONCLUSIONS

In conclusion, we have presented the recent developments of digital holographic techniques for biological microscopy applications. Different numerical reconstruction algorithms are reviewed and compared. As the most frequently used method, the Fresnel theory offers quality reconstruction without aliasing if the reconstruction distance is larger than the minimum object-to-hologram distance. Although the Fresnel diffraction formula can still give an accurate reconstruction for smooth and slowly varying objects where the Fresnel approximation is not strictly satisfied, it cannot correctly reconstruct near wave fields for more diffractive objects,

where the higher-order terms in the expansion of the Fresnel approximation are more significant. The angular spectrum method in contrast, demonstrates flexibility and effectiveness in image reconstruction without the limitation of having to satisfy a minimum object-to-hologram distance as in Fresnel diffraction.

In this paper we have also presented experimental results that demonstrate the capabilities of digital holography for biological microscopy. In holographic movies a series of holograms are recorded and the images are reconstructed with numerically adjustable focus so that the rapidly moving microbes can be accurately tracked. Real-time imaging is possible because only one exposure per time frame is required. The lateral and longitudinal resolutions obtained are consistent with diffraction limited imaging. Digital holography and movies of digital holography therefore are seen to be a useful new tool for biological microscopy.

Although numerical phase unwrapping methods are an established technique, they have difficulty correctly interpreting the phase images of rough and irregular objects. In this paper, we have demonstrated the use of an optical method of phase unwrapping by use of two or more wavelengths in digital holography. Through the application of this technique it is possible to remove 2π ambiguities and increase the axial range. The noise in the final phase profile is that of the single wavelength phase profile. The technique can be applied to incoherent interference imaging as well as holographic imaging.

Finally, we have described the principle of multi-wavelength scanning digital interference holography. By superposition of a number of numerically reconstructed optical diffraction fields of digital holograms that are optically recorded with varying wavelengths, it results in a synthesized short coherence length and corresponding axial resolution. Experimental results are presented to verify the principles.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the financial support for this work by the National Science Foundation.

REFERENCES

- [1] U. Schnars and W.P. Jueptner, "Direct recording of holograms by a CCD target and numerical reconstruction", *Appl. Opt.* **33**, 179-81 (1994).

- [2] S. Grilli, P. Ferraro, S. De Nicola, A. Finizio, G. Pierattini and R. Meucci, "Whole optical wavefields reconstruction by digital holography", *Opt. Express* **9**, 294-302 (2001).
- [3] U. Schnars and W.P.O. Jueptner, "Digital recording and numerical reconstruction of holograms", *Meas. Sci. Technol.* **13**, R85-R101 (2002).
- [4] J.W. Goodman and R.W. Lawrence, "Digital image formation from electronically detected holograms", *Appl. Phys. Lett.* **11**, 77-79 (1967).
- [5] S. Seebacher, W. Osten and W. Jueptner, "Measuring shape and deformation of small objects using digital holography", *Proc. SPIE* **3479**, 104-15 (1998).
- [6] L. Xu, X. Peng, J. Miao and A.K. Asundi, "Studies of digital microscopic holography with applications to microstructure testing", *Appl. Opt.* **40**, 5046-5051 (2001).
- [7] W.S. Haddad, D. Cullen, J.C. Solem, J.W. Longworth, A. McPherson, K. Boyer and C.K. Rhodes, "Fourier-transform holographic microscope", *Appl. Opt.* **31**, 4973-8 (1992).
- [8] D. Dirksen, H. Droste, B. Kemper, H. Delere, M. Deiwick, H.H. Scheld and G. von Bally, "Lensless Fourier holography for digital holographic interferometry on biological samples", *Opt. Lasers Eng.* **36**, 241-249 (2001).
- [9] W. Xu, M.H. Jericho, I.A. Meinertzhangen and H.J. Kreuzer, "Digital in-line holography for biological applications", *Proc. Natl. Acad. Sci. USA* **98**, 11301-05 (2001).
- [10] I. Yamaguchi, J. Kato, S. Ohta and J. Mizuno, "Image formation in phase-shifting digital holography and applications to microscopy", *Appl. Optics* **40**, 6177-6186 (2001).
- [11] U. Schnars, "Direct phase determination in hologram interferometry with use of digitally recorded holograms", *J. Opt. Soc. Am. A* **11**, 2011-5 (1994).
- [12] A. Barty, K.A. Nugent, D. Paganin and A. Roberts, "Quantitative optical phase microscopy", *Opt. Lett.* **23**, 817-9 (1998).
- [13] E. Cuche, F. Bevilacqua and C. Depeursinge, "Digital holography for quantitative phase-contrast imaging", *Opt. Lett.* **24**, 291 (1999).
- [14] E. Cuche, P. Marquet and C. Depeursinge, "Simultaneous amplitude-contrast and quantitative phase-contrast microscopy by numerical reconstruction of Fresnel off-axis holograms", *Appl. Optics* **38**, 6994-7001 (1999).
- [15] P. Ferraro, S. De Nicola, A. Finizio, G. Coppola, S. Grilli, C. Magro and G. Pierattini, "Compensation of the inherent wave front curvature in digital holographic coherent microscopy for quantitative phase-contrast imaging", *Appl. Opt.* **42**, 1938-46 (2003).
- [16] L.F. Yu and L.L. Cai, "Iterative algorithm with a constraint condition for numerical reconstruction of a three-dimensional object from its hologram", *Journal of the Optical Society of America a-Optics Image Science and Vision* **18**, 1033-1045 (2001).
- [17] K. Matsushima, H. Schimmel and F. Wyrowski, "Fast calculation method for optical diffraction on tilted planes by use of the angular spectrum of plane waves", *Journal of the Optical Society of America a-Optics Image Science and Vision* **20**, 1755-1762 (2003).
- [18] W. Osten, T. Baumbach and W. Jueptner, "Comparative digital holography", *Opt. Lett.* **27**, 1764 (2002).
- [19] T.-C. Poon, "Three-dimensional image processing and optical scanning holography", *Adv. Imaging & Electron Phys.* **126**, 329-350 (2003).
- [20] I. Yamaguchi and T. Zhang, "Phase-shifting digital holography", *Opt. Lett.* **22**, 1268 (1997).
- [21] T. Zhang and I. Yamaguchi, "Three-dimensional microscopy with phase-shifting digital holography", *Opt. Lett.* **23**, 1221 (1998).

- [22] M. Jacquot, P. Sandoz and G. Tribillon, "High resolution digital holography", Opt. Commun. **190**, 87-94 (2001).
- [23] J.W. Goodman, *Introduction to Fourier Optics, 2nd ed* (New York, McGraw-Hill, 1996).
- [24] C.F. Lo, X. Peng and L.L. Cai, "Surface normal guided method for two-dimensional phase unwrapping", Optik **113**, 439-447 (2002).
- [25] M. Servin, J.L. Marroquin, D. Malacara and F.J. Cuevas, "Phase unwrapping with a regularized phase-tracking system", Appl. Opt. **37**, 1917-23 (1998).
- [26] M.A. Schofield and Y. Zhu, "Fast phase unwrapping algorithm for interferometric applications", Opt. Lett. **28**, 1194 (2003).
- [27] J. Gass, A. Dakoff and M.K. Kim, "Phase imaging without 2-pi ambiguity by multiwavelength digital holography", Opt. Lett. **28**, 1141-3 (2003).
- [28] Y.Y. Cheng and J.C. Wyant, "Two-wavelength phase shifting interferometry", Appl. Opt. **23**, 4539-43 (1984).
- [29] K. Creath, "Step height measurement using two-wavelength phase-shifting interferometry", Appl. Opt. **26**, 2810-6 (1987).
- [30] P. de Groot, "Extending the unambiguous range of two-color interferometers", Appl. Opt. **33**, 5948-53 (1994).
- [31] M.K. Kim, "Wavelength scanning digital interference holography for optical section imaging", Opt. Lett. **24**, 1693 (1999).
- [32] M.K. Kim, "Tomographic three-dimensional imaging of a biological specimen using wavelength-scanning digital interference holography", Opt. Express **7**, 305-10 (2000).
- [33] A. Dakoff, J. Gass and M.K. Kim, "Microscopic three-dimensional imaging by digital interference holography", J. Electr. Imag. **12**, 643-647 (2003).
- [34] D. Huang, E.A. Swanson, C.P. Lin, J.S. Schuman, W.G. Stinson, W. Chang, M.R. Hee, T. Flotte, K. Gregory, C.A. Puliafito and J.G. Fujimoto, "Optical coherence tomography", Science **254**, 1178-81 (1991).
- [35] A. Dubois, L. Vabre, A.C. Boccara and E. Beaurepaire, "High-resolution full-field optical coherence tomography with Linnik microscope", Appl. Opt. **41**, 805-12 (2002).
- [36] L. Yu and M.K. Kim, "Full-color three-dimensional microscopy by wide-field optical coherence tomography", Opt. Express **12**, 6632-6641 (2004).

Chapter 3

DIGITAL HOLOGRAPHY FOR BULK IMAGE ACOUSTOOPTICAL RECONSTRUCTION

Vladimir V.Petrov

Saratov State University, Saratov, Russia

Abstract The problem of moving image reconstruction from digital mathematical hologram created by the computer or written from real object by means of bulk acoustical field is considered. It is assumed that a complicated sound field, generated by a line of electro-acoustical transducers, affects the crystal in such a way that the resulting spatial distribution of the refractive index exactly repeats the function of the transparency space distribution in the optical hologram written from the object. The case of color image reconstruction when digital holograms are written for three separate channels is considered. The example of the discussed approach for bulk image recreation used to mobile phone color display application is considered.

Keywords: holography, acousto-optics, 3D moving imaging, thick hologram, color image, bulk image recreation

1. INTRODUCTION

A number of papers are devoted for the development of the methods of 3D moving image forming [1-25]. There exist many approaches in this way:

- Multi-view stereoscopic 3-D imaging;
- Computer-aided 3-D imaging;
- Computer generated holograms for 3-D displays;
- Optical scanning holography;
- Acousto-optically based 3-D holographic imaging;

- LCD based 3-D imaging;
- 3-D Microscopy;
- Rotating screen based 3-D scenes reconstruction;
- Multi-plane volumetric 3-D displays, and others.

But the problem of choosing the right approach still remains. There are commercial samples of 3-D information displays on the market, which are predominately based on the multi-view principle. The main shortcoming of that principle is the overlapping of two images in boundary zone, which makes the observation of the reconstructed bulk image problematic. The holographic principle is considered to be one of the most promising candidates for recreation of a bulk scene image. But the necessity to have a very high resolution of the material that carries the holographic information makes it difficult, from a technological point of view, to apply this principle for electronic bulk moving images recreation. Nevertheless, some scientific groups all over the world are working on the development of different variations of holographic method application. In this paper we continue the consideration of one of holography-based methods for electronically controlled moving scene image recreation.

2. MAIN ASSUMPTIONS

The general idea of the method in question consists in the formation of an optical hologram by means of a complicated acoustical field in a way by which the generated acoustical field possesses space distribution of refractive index, which exactly corresponds to the transmission function of the original optical hologram. The bulk holographic image is reconstructed due to acoustooptical (AO) effect when of a sound field is illuminated by a reference optical beam under the Bragg angle.

This principle was first proposed [17] for a surface sound wave field application in 1994. This method was applied in 1996 [21] for Bragg optical dynamic hologram generated by bulk acoustical waves for moving image reconstruction and further developed in [22 - 25].

There is a set of problems one can meet on the way of such system construction.

1) For acoustooptical interaction it is necessary to find a medium with low sound attenuation and a high figure of merit. This is important because of necessity to excite the sound field at very high frequencies so that the sound wave lengths become so short to be comparable with optical wavelength. In this case the spatial resolution can be comparable with one in a usual optically created hologram. Besides, the bandwidth of acoustooptical interaction must be very large to allow the excitation of a large range of spatial frequencies of the hologram to translate maximum data and to avoid loosing any information about the reconstructed image. In present time experimental works in realization of a very wide band (up to 3 GHz) and very high frequencies (up to 10 GHz) acoustooptical interaction [26] are

already known. At such high frequencies the sound wave lengths are already comparable with optical wavelengths. Of course up to now the sound attenuation is still very high in known materials.

2) To achieve the exact repetition (or approximation) of transparency spatial distribution of optical hologram in acoustically created optical hologram it is necessary to find the mechanism of complicated sound field distribution realization. One of possible ways is to use a digitized approach for a sound field exciting by an array or a line of electro-acoustical transducers. An electrical signal to each transducer is led from the driver where the time and spatial distribution of electrical field is numerically created. This distribution is previously calculated with the computer (or is taken from the real time recording system). This approach requires the creation of very small point sound radiators placed to an array or line with a period comparable with a sound wave length.

3) To calculate in real time such a complicated electrical signal distribution a very high-speed processor is required.

Consideration of the main principles of acoustooptical holographic moving color image reconstruction and the discussion of the main problems is the very aim of the present work.

3. SYSTEM ARCHITECTURE

One of possible variants of the scheme architecture for holographic moving image AO reconstruction in Fig.3-1 is presented.

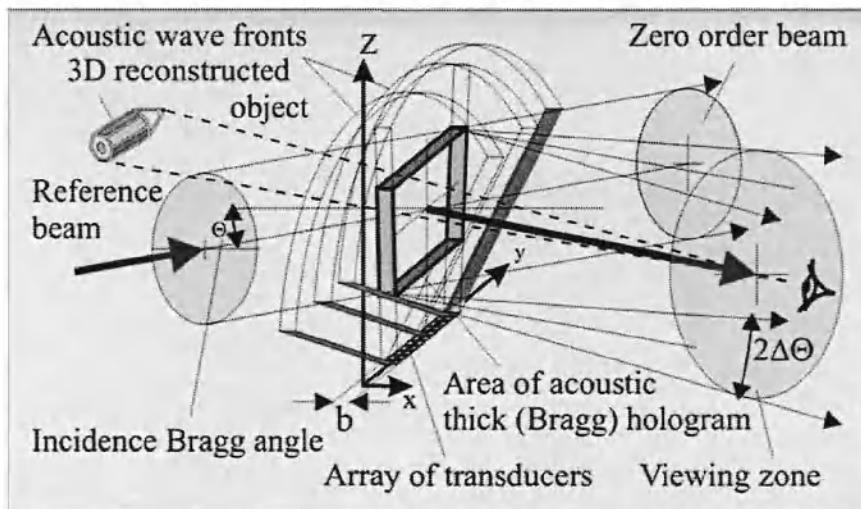


Fig. 3-1. Architecture of an acoustooptical set-up for 3D holographic image reconstruction

In Figure 3-2 the scheme of optical hologram that is formed by a complicated sound field is shown. The display itself represents the plate of AO interaction media with sizes $H \times D$ (See Fig. 3-2). On one of butt ends of this plate the multielement system of sound radiators (Array of transducers, Fig. 3-1) is disposed. The electrical impulse signals are directed to each element from the controlling processor that calculates and generates the necessary amplitude and phase of each signal in accordance with a special algorithm. Such an algorithm provides the formation of a certain sound field distribution (dynamic acoustically created optical hologram) in the far field area with sizes $A \times B$ at the maximum distance H from the transducer's plane (Fig.3-2). This sound field distribution can be calculated by the computer, taking into account the following factors: sound attenuation, pixel's shape distortion and other possible effects that deteriorate the recreating image. The set of electrical signals, which feed the sound radiators, can be also retranslated from the image receiving system (which is a subject of a separate consideration) so that the imaging scene is reconstructed in real time.

The spatial resolution of AO imaging reconstruction system mainly depends on the sound wave length λ , the period of an array l , the duration of electrical impulse τ_s , which feeds each sound radiator, and the duration of optical reference impulse τ_L which illuminates the hologram when recreating the image. Thus, disregarding the distortion effects, the spatial resolution in the plane parallel to the transducer's array (along y axis) is $\Delta y = l$, and along the normal to transducer's plane (x axis) is $\Delta x = V\tau_L$, where V is the sound velocity. The unique pixel duration τ_s must not exceed and must be synchronized with the duration of reference light impulse τ_L . The sizes of each piezo element are: a is the width (along y -axis) and b is the length. The spatial period l of transducer's array has to be comparable with the acoustical wavelength λ to provide the sufficient divergence γ of sound beam in the plane of hologram (see Fig 3-2.). The sequence of electronically controlled and acoustically created optical holograms moves in the space continuously from the transducer's plane along the x -axis. It is assumed that the whole hologram of each particular image is formed in the far field ($A \times B$) area. Once the particular hologram is formed, the optical light impulse is on and the particular image is recreated. The consequence of reference light flashes has the spatial period T , which is in fact defined by the time T of sound transmission from the transducer's plane to the end of the display ($T=H/V$). The reference optical beam from the laser is formed in time by an impulse modulator and in space by a lens. The moving 3D holographic image that is recreated due to acousto-optical effect can be seen by the observer in the direction of the first order of Bragg diffraction. An optical polarizer can be used for cutting out of unused zero order of diffracted light.

The Bragg regime is provided when the Klein-Cook parameter $Q \geq 2\pi$ ($Q = 2\pi L\lambda/A^2$), here λ - is the optical wavelength in a vacuum.

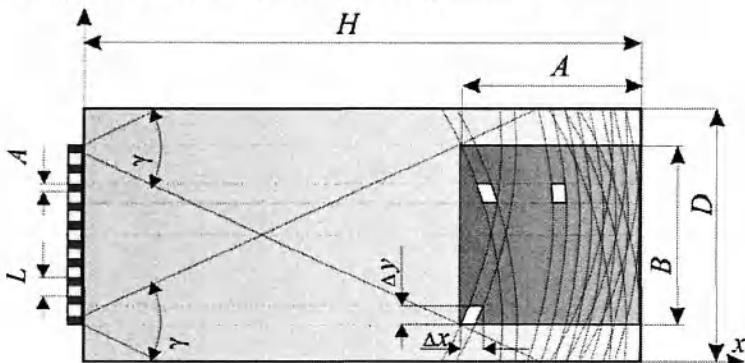


Fig. 3-2. The scheme of sound wave dynamic optical hologram forming

The calculations shows that a small laboratory model of holographic display with working area $2.56 \times 2.56 \text{ cm}^2$ and space resolution $175 \times 100 \mu\text{m}^2$ can be designed with the TeO_2 as photoelastic medium. The diffraction efficiency 0.16% per 1 mW of electrical power can be obtained.

4. COMPUTER SIMULATIONS

4.1 Acoustooptical cell is considered as a hologram

Presumably the AO cell was first considered as a kind of hologram by A. Korpel in [27-29] while applying acoustooptics for image Bragg visualization. It was shown that the AO cell possesses all the features of a hologram.

Indeed, every diffraction structure in general can be considered as a hologram. In the case with an AO cell let us consider a simple example.

Let us imagine that the electrical signal, which is directed to the electro-acoustical transducer of an AO cell is linearly frequency modulated. This electrical signal excites the correspondent acoustical wave, which is also linearly frequency modulated. The diffraction array, which is created by such an acoustical wave by means of the refractive index variation, can be approximated as a part of one-dimensional Fresnel zone plate with removed center. On the other hand the Fresnel zone plate represents the Fresnel hologram of a point source. Thus the linearly frequency modulated acoustical wave in fact represents part of a point source Fresnel hologram.

4.2 Notes concerning the Bragg acoustical hologram creation

Keeping in mind the above-mentioned assumptions, it is possible to define the Bragg acoustical hologram. The Bragg acoustical hologram represents a space distribution of the refractive index of photo-elastic medium in the plane (x,y) , which is created by an acoustical field, excited by a line of bulk sound wave radiators. This radiators line is disposed along the y coordinate at one of butt-ends of crystal plate. Each radiator has small size (a) along y – coordinate, that allows to consider it as a “point source”, but it remains continued length (the size b) along z – coordinate, which defines the length of acoustooptical interaction, and, consequently, defines the diffraction efficiency. The length of unique radiator (b) defines also the frequency bandwidth Δf of acoustooptical interaction in which the Bragg condition is fulfilled: $\theta = \arcsin(f_0\lambda_0/2Vn)$, where θ - the light incidence angle; f_0 – central frequency of excited sound waves; λ_0 - light wave length in vacuum; V - and n - the velocity of sound wave and refractive index, correspondingly.

There are several reasons to choose the Bragg regime of light diffraction on bulk acoustic waves:

1. To get the appropriate (high) efficiency of AO interaction;
2. To provide the image reconstruction exclusively in the first order of diffraction.
3. To reach very high sound frequencies (very small diffraction array's periods) to enhance the spatial resolution.
4. To obtain very big frequency bandwidth to enlarge the spectrum of spatial frequencies of hologram and to decrease the image distortion.
5. In the case of Bragg light diffraction on the bulk acoustic waves the acoustically created optical dynamic hologram in its features is equivalent to the thick 3D hologram (which is also called the hologram of Yu. Denisyuk) [30]. That means that such a hologram possess the feature of selectivity to the optical wavelength λ .
6. When using an anisotropic Bragg diffraction it is possible to cut out the useless zero order of diffraction by a polarizer as the polarization of image light is turned to $\pi/2$ in this case as compared with the zero (undiffracted light) order.

There is a shortcoming of Bragg acoustical hologram application that consists in the restriction by bandwidth Δf of frequencies spectrum, that “takes part” in the formation of a sound dynamic hologram. This restriction leads to a partial loss of information reflected by the hologram. But taking into account the possibility of a high frequency acoustical field application, it is possible to hope for realization of wide frequency band for which this restriction can be neglected.

4.3 Computer modeling of acoustically created optical hologram

In [17, 31] on the basis of [32, 33] the algorithm of numerical evaluation of acoustically created optical hologram was build. In [25] this theory was developed for the AO medium with sound attenuation; the thick acoustically created optical hologram was considered. In the same work a method of static modeling of dynamic electronically controlled hologram was developed.

It can be shown (see, for example [31]), that the hologram of multi-plane object is the superposition with the phase shift of unique plane hologram's set. That means that for computer modeling we can use a two-dimensional approach to analyze a step-by step transformation of an optical field reflected by the object to the electrical signal's set. The same approach can be applied to the reverse process – the transformation of an electrical signal set to the reconstructed bulk image.

If the object occupies the unique plane at the distance z from the hologram with transmission function $a(x,y)$, then the scalar field in the hologram plane can be represented as Fresnel-Kirchhoff integral [31-35]

$$\begin{aligned} \Psi_z(x, y, z) &= \\ &= \frac{\exp(jkz)}{jkz} \cdot \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [a(\xi, \eta)] \cdot \left(\exp \left\{ j \cdot \frac{\pi}{\lambda \cdot z} [(x - \xi)^2 + (y - \eta)^2] \right\} \right) d\xi d\eta, \quad (3-1) \end{aligned}$$

where $k = 2\pi/\lambda$, kz – phase of illuminating wave.

Denoting the convolution kernel of this integral as

$$\bar{h}_z(x, y) = \frac{\exp(jkz)}{jkz} \cdot \exp \left[j \cdot \frac{\pi}{\lambda \cdot z} \cdot (x^2 + y^2) \right], \quad (3-2)$$

the field at the object surface can be represented as 2-dimensional convolution of space impulse reply $\bar{h}(x,y)$ with 2-dimensional input signal $\exp(-jkz)[a(x,y)]$.

$$\Psi_z(x,y) = \exp(-jkz) \cdot [a(x,y)] * * \bar{h}_z(x,y) \quad (3-3)$$

As noted in [17], it is more convenient to use another space impulse reply $h_z(x,y)$, which is defined relatively $\bar{h}_z(x,y)$ by the phase shift [31]

In this case the equation (3-2) can be rewritten

$$h_z(x,y) = \exp(-jkz) \cdot \bar{h}_z(x,y) = \frac{1}{jkz} \cdot \exp\left[j \frac{\pi}{\lambda \cdot z} \cdot (x^2 + y^2)\right]. \quad (3-4)$$

In this case the field on the object surface is

$$\Psi_z(x,y) = a(x,y) * * h_z(x,y) \quad (3-5)$$

In the Fresnel approach the intensity distribution at the distance z from the object plane (hologram field) can be represented as [17]:

$$\begin{aligned} I_z(x,y,z) = & \\ &= B^2 + |a(x,y) * * h_z(x,y)|^2 + B \cdot \exp[-jkx \cdot \sin(\theta)] \cdot [a(x,y) * * h_z(x,y)] + \\ &+ B \cdot \exp[jkx \cdot \sin(\theta)] \cdot [a^*(x,y) * * h_z^*(x,y)], \end{aligned} \quad (3-6)$$

where B – amplitude of plane (reference) light wave, θ - light incidence angle; $**$ - 2-dimensional convolution operator. In fact the formula (3-6) represents the equation, which is known as a main holographic equation [36].

As the field I_z is represented by 2-D convolution, then it can be considered as the output of some 2-D linear system with impulse function $h_z(x,y)$ [36].

When calculating the discrete mathematical holograms let us follow the approach applied in [31,32]. The discrete impulse function h_{z_D} can be defined as

$$h_{z_D}(m,n) = \exp\left[j \frac{\pi}{\lambda \cdot z} \cdot X^2(m^2 + n^2)\right] = \exp\left[j \alpha^2 \frac{\pi}{M} \cdot (m^2 + n^2)\right], \quad (3-7)$$

where m,n – quantization indexes, $m = 0 .. M-1$; $n = 0 .. N-1$; X and Y – quantization steps along x and y coordinates, correspondingly. For calculation simplification $X=Y$ and $M=N$ is chosen; the α parameter is defined as

$$\alpha = x(M/\lambda z)^{1/2} \quad (3-8)$$

The transmission function $a(x,y)$ in this case must be quantized $a_{z_D}(m,n) = a(xM, yN)$. $m=xM$, $n=yN$.

For the reference beam it is useful to define the phase parameter β

$$\beta = 2\pi\alpha \cos(\theta) (\lambda z/M)^{1/2} \quad (3-9)$$

Then the discrete hologram can be rewritten as [17]

$$I_{z_D}(m, n) = \left| B \cdot \exp(j \cdot \beta \cdot M) + a_{z_D}(m, n) * * h_{z_D}(m, n) \right|^2 \quad (3-10)$$

The image reconstruction occurs when the hologram is illuminated by the same coherent plane wave (reference field) as the wave for the recording process. In this case the image is reconstructed at the same distance from the hologram as the registered one. So the formula (3-10) can be applied for the image reconstruction. For that the transmission function of an object $a(m, n)$ must be replaced by the transmission function of a hologram.

When calculating a discrete hologram, the 2-dimensional convolution operation $\text{conv}(a, h_{z_D})$ for two finite functions ($a_{z_D}(m, n)$) and ($h_{z_D}(m, n)$) in the equation (3-10), taking into account the convolution theorem [33,35], can be realized by fast Fourier transform [36].

$$\begin{aligned} \text{conv}(a, h_{z_0}) &= \\ &= a_{z_D}(m, n) * * h_{z_D}(m, n) = \\ &= \frac{1}{\sqrt{M \cdot N}} \cdot F^{-1}\{\sqrt{M \cdot N} \cdot A(k, l) \cdot H(k, l)\} \\ &= \frac{1}{\sqrt{M \cdot N}} \cdot F^{-1}\{\sqrt{M \cdot N} \cdot F[a(m, n)] \cdot F[h_{z_D}(m, n)]\}, \end{aligned} \quad (3-11)$$

where $A(k, l)$ and $H(k, l)$ - Fourier images of functions $a_{z_D}(m, n)$ and $h_{z_D}(m, n)$, correspondingly; and F and F' - the operators of right and back Fourier transform; $k=0, 1, \dots, M-1$; $l=0, 1, \dots, N-1$.

Then the equation (3-10) taking into account (3-11) can be rewritten as

$$I_{z_D}(m, n) = \left| B \cdot \exp(j \cdot \beta \cdot M) + \text{conv}(a_{z_D}, h_{z_D}) \right|^2, \quad (3-12)$$

4.4 Relationship between space distribution of hologram transmission function and time-space distribution of electrical signal's field

The important question concerns the problem how to find the appropriate set of electrical signals to excite the acoustical field, which exactly corresponds to the necessary space distribution of optical hologram transmission function. A good suggestion to solve this problem was done in [17], where it was obviously assumed that the equality of the fields with complicated space spectrums must lead to the equality of each spectrum component of these fields. Using this approach, let us find the function that connects the electrical signal's field to the acoustical field, which forms the hologram, taking into account the sound attenuation.

Suppose that, as it is shown in figure 2, the display's plane lies in the rectangular system of coordinates x, y in such a way that acoustical waves are excited in the direction x , and the line of bulk sound waves transducers is situated in the direction y . A reference light wave incidents to the display under some angle θ relatively to z -axis (so that the Bragg regime of acoustooptical interaction is realized). When reconstructing the image from dynamic acoustically created hologram, this angle must be equal to the Bragg angle (the angle between the wave vector of incident reference light beam and the wave front (along z axis) of sound waves).

Suppose $\psi(x, y)$ is a sound field of hologram and $f(t, y)$ is an electrical signals distribution field.

We will assume that the length of interaction is infinite (along z axis). Besides the plane sinusoidal acoustic wave $\exp[j(\Omega t - Kr)]$ travels in the direction of the sound wave vector K with components K_x and K_y ; r — is the position vector with coordinates x, y ; Ω - circular frequency of sound wave. Phase velocity V_p of sound wave can be expressed as

$$V_p = \frac{\Omega}{\sqrt{K_x^2 + K_y^2}}. \quad (3-13)$$

If signal

$$f(t, y) = \exp[j(\Omega t + uy)] \quad (3-14)$$

is applied to the transducer's line (u – the space frequency, that is changed in y direction), then time-changeable acoustic field in crystal taking into account sound attenuation can be written as

$$\psi_i(x,y) = \exp[j(\Omega t - K_x x + uy) - \alpha_a x], \quad (3-15)$$

where α_a – sound attenuation index in x direction. Sound wave number in y direction is $K_y = -u$. Taking into consideration (3-13), we have

$$K_x = \sqrt{\frac{\Omega^2}{V_p^2} - u^2} \quad (3-16)$$

At the moment of time t_i the acoustical field

$$\psi_{t_i}(x,y) = \exp[j(\Omega t_i - K_x x + uy) - \alpha_a x] \quad (3-17)$$

will be excited by electrical signal $f(t,y)$. The relationship between these two fields can be found by applying two-dimensional Fourier transforms. The arbitrary bit of sound field $\psi_{t_i}(x,y)$ can be extended to plane components as

$$\psi_{t_i}(x,y) = \frac{1}{4\pi^2} \cdot \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Psi_{t_i}(k_x, k_y) \cdot \exp[j(k_x x + k_y y)] dk_x dk_y, \quad (3-18)$$

where $\Psi_{t_i}(k_x, k_y)$ - the complex amplitude of individual plane wave component; it can be found as back Fourier transform

$$\Psi_{t_i}(k_x, k_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi_{t_i}(x,y) \cdot \exp[-j(k_x x + k_y y)] dx dy, \quad (3-19)$$

here $k_x = -K_x$; $k_y = -K_y = u$. As it was already noted, for excitation of sound plane wave $\Psi_{t_i}(x,y) = \exp[j(\Omega t_i + k_x x + k_y y) - \alpha_a x]$, it is necessary to direct to the sound radiators line the signal $f(t,y) = \exp[j(\Omega t + k_y y)]$. Therefore, to get the sound field $\exp[j(k_x x + k_y y)]$, it is necessary to apply the electrical signal $\exp[j(\Omega t - \Omega t_i + \kappa_y y) + \alpha_a V_p t]$. So it is inferred that the electrical signal

$$f_0(t,y) = \Psi_{t_i}(k_x, k_y) \exp\{j[\Omega(t - t_i) + k_y y] + \alpha_a V_p t\} \quad (3-20)$$

will excite the plane sound wave component in which the sound attenuation will be compensated along the coordinate x :

$$\Psi_{t_i 0}(x, y) = \Psi_{t_i}(k_x, k_y) \exp[j(k_x x + k_y y)], \quad (3-21)$$

where $\Psi_{t_i}(k_x, k_y) = \text{const}$ for fixed k_x and k_y .

The arbitrary bit of sound field $\psi_{ti}(x, y)$ will be generated when to the transducer's line the electrical signal is applied which is found as a Fourier transform (3-18) from signal's set $f_0(t, y)$ (3-20), that excite the plane components of acoustical field $\Psi_{t_i 0}(x, y)$ (3-21).

$$f(t, y) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Psi_{t_i}(k_x, k_y) \exp\{j[\Omega(t - t_i) + k_y y] + \alpha_a V_p t\} dk_x dk_y, \quad (3-22)$$

It follows from the equation (3-13) that

$$\Omega = V_p (k_x^2 + k_y^2)^{1/2}. \quad (3-23)$$

Substituting (3-23) into (3-22) we get

$$\begin{aligned} f(t, y) &= \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Psi_{t_i}(k_x, k_y) \exp\{j[V_p(k_x^2 + k_y^2)^{1/2}(t - t_i) + k_y y] + \alpha_a V_p t\} dk_x dk_y \\ &= \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Psi_{t_i}\left[\left(\frac{\Omega^2}{V_p^2} - k_y^2\right)^{1/2}, k_y\right] \frac{\Omega \exp\{j[\Omega(t - t_i) + k_y y] + \alpha_a V_p t\}}{V_p^2 [(\Omega^2/V_p^2) - k_y^2]^{1/2}} dk_y d\Omega. \end{aligned} \quad (3-24)$$

On the other hand, signal $f(t, y)$ can be also represented by sinusoidal components in frequency domain (u, Ω) by using the Fourier transformation

$$f(t, y) = \frac{1}{4\pi^2} \cdot \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(\Omega, u) \cdot \exp[j(\Omega t + uy) + \alpha_a V_p t] du d\Omega. \quad (3-25)$$

Comparing (3-24) and (3-25), it is possible to get

$$\begin{aligned} F(\Omega, u) \Big|_{u=k_y} &= \\ &= \Psi_{t_i} \left[\left(\frac{\Omega^2}{V_p^2} - k_y^2 \right)^{1/2}, k_y \right] \cdot \frac{\Omega}{V_p^2 [(\Omega^2/V_p^2) - k_y^2]^{1/2}} \cdot \exp(-j\Omega t_i). \end{aligned} \quad (3-26)$$

This equation, obtained through the method proposed in [17], connects the angular spectrum $F(\Omega, u)$ of electrical signal $f(t, y)$, applied to the sound radiators line, and angular spectrum $\Psi(K_x, K_y)$ of excited acoustical field. Correspondingly, fields $f(t, y)$ and $\psi(x, y)$ are related as back Fourier transforms from functions $F(\Omega, u)$ and $\Psi(K_x, K_y)$. As it is seen from the equations (3-20) and (3-22), the compensation of sound attenuation in x -direction leads to the necessity to put analogous attenuation in time for signals that is applied to the transducers line.

Relation (3-26) allows also taking into account the coordinates relationship in frequency domain of angular spectrums transformation of electrical field and excited acoustical one.

4.5 Steps for digital modeling of bulk image capturing and reconstruction

In brief we can describe the algorithm of recording and reconstruction of electronic digital hologram's process as follows:

Recording process of electronic hologram

Step 1. Admission that the 3-dimensional (3D) recording object hologram can be represented as a superposition (with phase shifting) of single-plane hologram [31].

Step 2. Expression of the scalar field in the hologram plane through the Fresnel-Kirchhoff integral and representation it as 2D convolution of pulse spatial reply (reference field) with 2D input (object) signal.

Step 3. Representation at the Fresnel approach of the field intensity distribution at the given distance from the object plane (mathematical Fresnel hologram).

Note.

This mathematical Fresnel hologram represents the field intensity distribution at the hologram plane. Such field intensity distribution can be used then for the creation of physical hologram. For example, illuminating with this field of the photo plate can do it. It is alluring also to admit that with this field is possibly to create by some way the momentary distribution of acoustical waves in photoelastic medium. Avoiding now the discussion of mechanism of optical induction of sound wave hologram, we just assume that such necessary sound waves distribution is done*.

* The question of sound wave hologram recording is the topic of separate interesting problem, which is not the aim of present chapter and it will be discussed in future publications.

Step 4. It is assumed that the spatial distribution of elasto-optical medium refractive index variation, which is created by sound waves excited by the array of transducers at the fixed moment of time, exactly repeats the space distribution of optical hologram transmission function variation. This assumption means the equivalence of each spectrum components of these optical and acoustical fields. So the Fourier transform of obtained field distribution in hologram plane is written. This equation evaluated in such a way that it takes into account the compensation of sound attenuation influence of each component of acoustical field.

Step 5. The equation, which defines the correspondence between angular spectrum of acoustical field that related to recorded hologram and real acoustical field (that possesses with features of movement along one coordinate, divergence and attenuation) that excites the electrical signals in the transducer's array, is specially evaluated.

Step 6. By back Fourier transform of last equation the expression for time-space electrical signals field distribution can be written.

Step 7. Simulated set of electrical signals can be then used for 3D holographic image reconstruction. It is assumed that in real device these signals are amplified and transmitted to the reconstruction system.

Reconstruction process represents the inverted steps of recording process:

Step1. Two dimensional space-time depended field of electrical signals is applied to the electro-acoustical transducer's array;

Step 2. By Fourier transform of the expression for these signals time-space distribution the spectrum components of electrical signal's can be found.

Step 3. Using the equation for coordinates transformation which defines the correspondence between angular spectrums of real (attenuated, moved and divergent) acoustical field excited by transducers and the field that corresponds to object hologram, the spectrum components of last one are found;

Step 4. By back Fourier transform the sound field two-dimensional distribution can be written;

Step 5. The reconstructed optical 3D image can be formed by using the equation for mathematical hologram when the hologram is illuminated by the same reference optical field as it is done during the registration process.

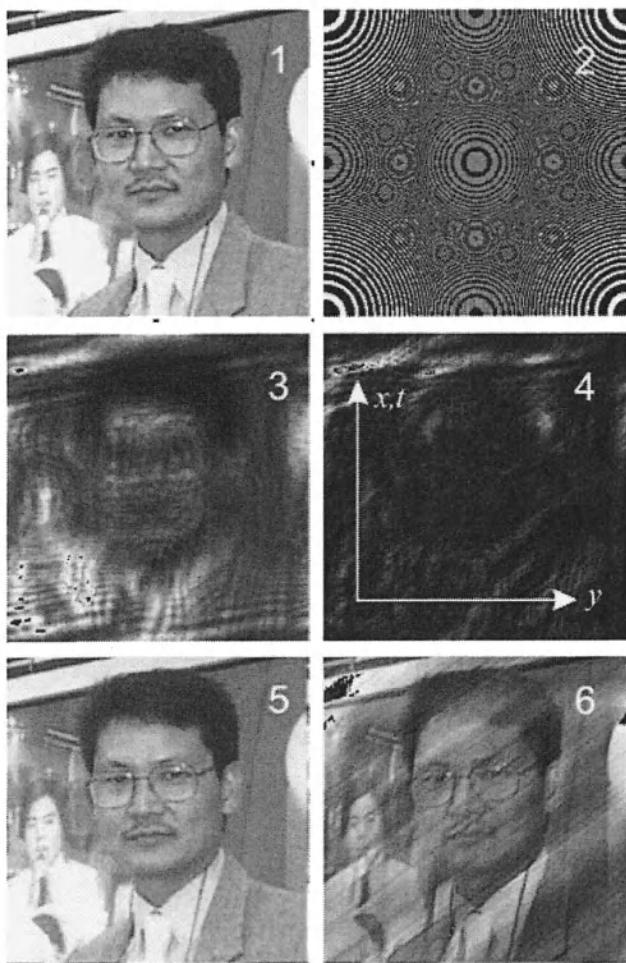


Fig. 3-3. Illustrations of several steps of mathematical hologram recording and bulk image reconstruction

1 – one plane of 3D object image; 2 – convolution kernel of Fresnel-Kirchhoff integral; 3- mathematical Fresnel hologram of an object; 4 - time (t) - space (y) distribution of electrical signals applied to the transducer's line, taking into account sound attenuation along x axis; 5 – object image reconstructed from optical hologram 3; 6 – the image reconstructed from acoustically created hologram, formed from the set of electrical signals shown on picture 4.

5. MODELING RESULTS

According to the algorithm described in chapter 3.4 the mathematical modeling of hologram creation process and following image reconstruction was fulfilled.

In figure 3-3 the results of several steps of mathematical hologram recording and bulk image reconstruction are shown. Here (1) is one plane of 3D object image. For numerical modeling we used just one plane image so that to avoid too complicated calculations and to save the process time. It is possible to do as it is known that hologram of 3D object can be mathematically built from holograms of unique object planes given with certain phase shift. (2) is the convolution kernel of Fresnel-Kirchhoff integral space distribution. (3) represents the mathematical Fresnel hologram of an object which, in fact can be created in any known way – by recording on the photo material or by formation with complicated sound field. Picture (4) shows the space (x,y) or space–time (y,t) distribution of electrical signals applied to the transducer's line, taking into account sound attenuation along x axis (see also Fig.3-2); (5) illustrates the 3D object image reconstructed from optical hologram (3). On picture (6) the image reconstructed from acoustically created hologram, formed from the set of electrical signals shown on picture (4) and taking into account the sound attenuation in crystal is presented.

The appearance of additional noise when recreating the image from acoustically build hologram might be probably stipulated by regular mistake accumulation when numerical modeling in the restricted special frequency range and due to applying the transformation algorithm describing the movement of divergent sound field with certain velocity. The effect of downing of the reconstructed image intensity along the sound wave propagation coordinate is taken into consideration and avoided by applying of special algorithm in the process of numerical simulation.

6. COLOR IMAGE PROCESSING

Let us consider the color image processing when recording the hologram and reconstructing the image on the example of the display for mobile phone application.

The crystal that represents the medium for acoustooptical interaction is placed inside the cover of the cell phone. The light sources for reference beams with red (R), green (G), and blue (B) colors are placed at the down pad part of a phone cell, as it is shown in figures 3-4 and 3-5.

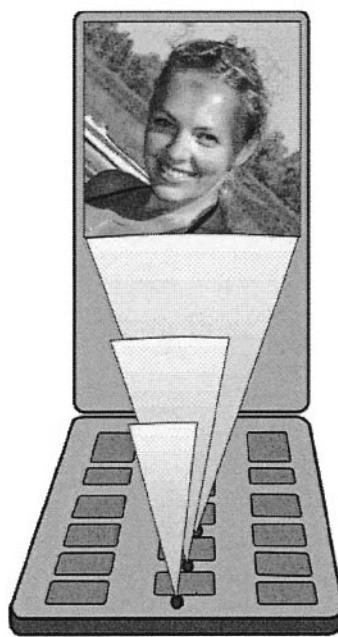


Fig. 3-4. Acoustooptical holographic display mobile phone application

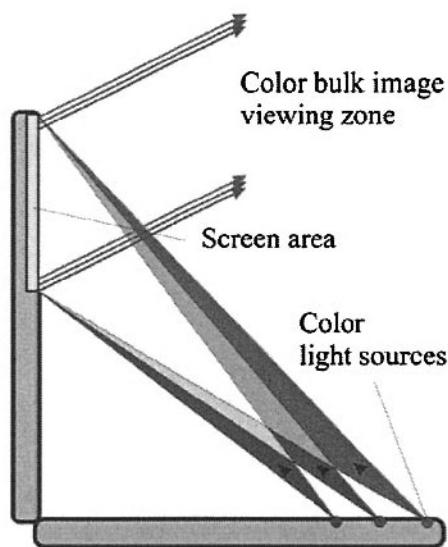


Fig. 3-5. Three color reference beam forming

In such suggested approach the reference beam of each color illuminates the whole hologram that specially formed for that color instead of illuminating the screen with pixel-by-pixel or line-by-line methods as it been used in other known systems.

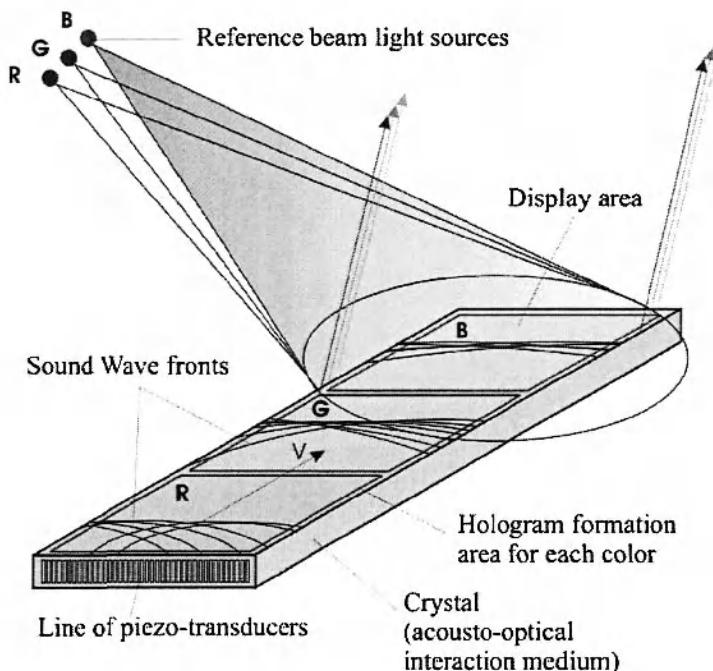


Fig. 3-6. Three-color dynamic hologram forming with bulk acoustical field excited by a line of piezo-elements

The principle of three-color holograms creation is illustrated in the Fig. 3-6. It is seen from this picture that when applying the electrical signals to the transducer's line by short impulses τ_s , the sound wave fronts generated in the crystal form the line of pixels in far field zone. By this wave the whole hologram for each color is formed by line-by-line. The acoustical wave front moves along the x - axis with the velocity of sound V . Once the moving hologram that corresponds to certain color reached of the display area (see Fig. 3-2), then the reference beam of corresponding color is switched on with duration equal to τ_L .

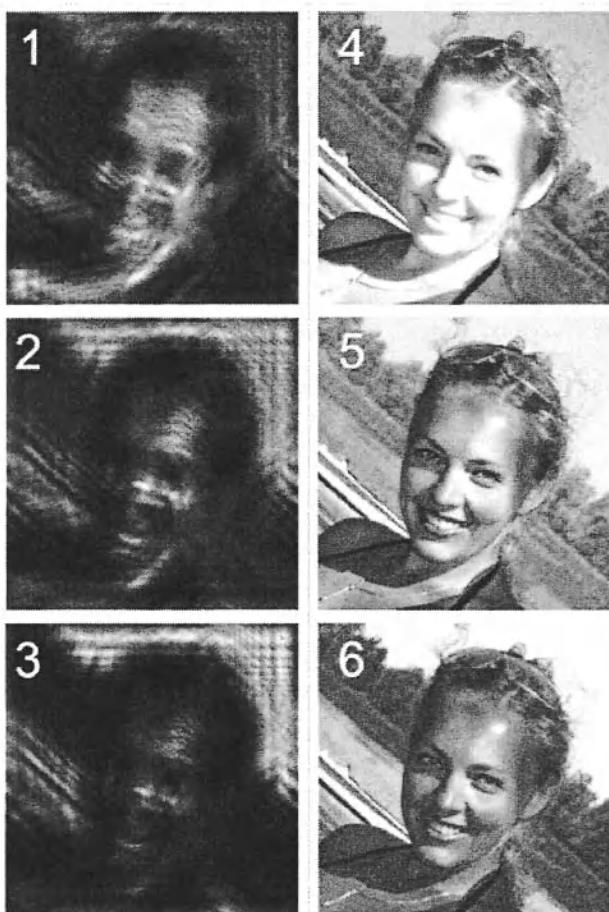


Fig. 3-7. Acoustically created digitally modeled holograms for Red (1), Green (2) and Blue (3) channels and corresponding Red (4), Green (5), and Blue (6) reconstructed images

7. CONCLUSION

At present work the holographic approach for the creation of bulk color moving image is considered on the basis of acoustooptical interaction of reference beams with acoustically build hologram. Such a hologram represents the diffraction array of refractive index variations that is formed by complicated sound field, which excited by the line of piezoelectric transducers. The possible variant of architecture of such kind of system is discussed and calculation results of main parameters of laboratory model of such a display are considered. There are also discussed the main problems on

the way of suggested display realization. The main inference that can be made is that offered approach can be applied with today known materials and technologies for the creation of small size bulk image displays like, for example, mobile phone displays.

REFERENCES

1. E.Leith et al., J.SMPTE **74**, *Requirements for a wavefront reconstruction television facsimile system*, 893-896, 1965.
2. G.Goetz et al., IEEE Trans. Electron. Devices **20** (11), *Recent progress in a real time three-dimensional display*, 1973.
3. R.V.Johnson, SPIE Proc. **222**, Laser Scanning and Recording for Advanced Image and Data Handling (SPIE, Bellingham, Wash.,), *Electronic focusing in the Scophony scanner*, 15-18, 1980.
4. D.Psaltis et. al., Opt. Eng. **23**, *Optical image correlation with a binary spatial light modulator*, 698-701, 1984.
5. G.Y.Sirat and D.Psaltis, Opt. Lett. **10** (1), *Conoscopic Holography*, 4 –6, 1985.
6. F.Mok et al. Opt. Lett. **11**, *Real-time computer generated holography means of liquid-crystal television spatial light modulator*, 748-750, 1986.
7. V.G.Komar, O.B.Serov, *Art holography and holographic cinematography*, Moscow "Isskustvo" - Publishing house, 1987.
8. P.St. Hilaire et al., J. Opt.Soc.Am. **A9** (11), *Synthetic aperture holography: A novel approach to three dimensional displays*, 1969-1977, 1992.
9. F.Wyrowski and O.Bryngdahl, J. Opt. Soc. Am. **A6** (8), *Specl-free reconstruction in digital holography*, 1171-1174, 1989.
10. I.N.Kompanets, S.A.Gonchukov, *Volumetric 3-D Liquid-Crystal Displays*, Information Display, **20**, 5 p.24-26, 2004.
11. S.A.Benton, SPIE Proc.**IS-8**. (SPIE Institute of Holography, Bellingham, Wash.Experiments in holographic video, 247 - 267 (1991); P.St.Hilaire et. al., SPIE Proc. **1667-33**, *Color images with the MIT holographic video display*, 1992.
12. S.Bains, Laser focus world..**29**. No 1, *Radial scanning produces 3-D image on flat screen*, 41-42, 1993.
13. P.St. Hilaire, Opt. Eng. **34** (10), *Scalable optical architecture for electronic holography*, 2900-2911, 1995.
14. P.Hilaire, Optics & Photonics News. **8**. No 8, *Holographic video: The ultimate visual interface?*, 35-39, 1997.
15. M.W.Halle, SPIE Proc. **2176**. Practical holography VIII, *Holographic stereograms as discrete imaging systems*, 73-84, 1994.
16. C. Chinnock, Laser focus world. **29**. No 9, *Volumetric imaging provides a walk-around view*, 20-22, 1994.

17. L.Onural, G.Bozdagı, A.Atalar, Optical Eng. **33**. No 3, *New high-resolution display device for holographic three-dimensional video: principles and simulations*, 835 – 844, 1994.
18. T.-C.Poon et.al., Opt. Eng. **34**, *Three-dimensional microscopy by optical scanning holography*, 1338-1344, 1995.
19. J.Kulick et.al., J.Opt.Soc. Am. A **12**(1), *Partial pixels: A three-dimensional diffractive display architecture*, 73-83, 1995.
20. K.Maeno et al., SPIE Proc. **2652**, Practical holography X, SA. Benton, ed. (SPIE, Bellingham, Wash.), *Electro-holography display using 15 megapixel LCD*, 15 – 23, 1996.
21. V.V.Petrov, *Set-up for electronic forming of three dimensional holographic image*, Patent of Russian Federation № 2117975, Priority 13.11.1996.
22. V.V.Petrov, Proc. SPIE, **5821**, Current Research on Image Processing for 3D Information Displays, V.Petrov, ed., *3D image reconstruction based on acoustically created hologram*, 60-75, 2005.
23. V.Petrov, Acoustooptical Club Proc. St.Petersburg, *Acoustooptical display for the creation of 3-D holographical image*, 95, 1997.
24. V.V.Petrov, D.A.Egorushkin, *Modeling of holographic television and computer displays*. Scientific paper's collection. Saratov, 85-87, 1997.
25. V.V.Petrov. *Holographic Video. Acoustooptical approach*, Saratov, "GosUNZ" - Publishing-house, 77 pages, 2002.
26. V.V.Petrov, M.A.Grigor'ev, and A.V.Tolstikov, Optics and Spectroscopy, **89** No 3, *Experimental Study of Acoustooptical Interaction in Lithium Niobate in the Frequency Range from 7.5 to 10.5 GHz*, 463-468, 2000.
27. A.Korpel, Appl. Phys.Lett. **9**. 425, 1966.
28. A. Korpel, Int.J.Nondestr. test., **1**. 337, 1970.
29. A.Korpel, L.W.Kessler, M.Ahmed, J. Ac.Soc.Am., **51**, 1972.
30. Yu.N.Denisuk, DAN SSSR, **144**, *About the reflection of optical features of object in the wave radiation field that is scattered by this object*, 1275, 1962; Optics and spectroscopy , **15**., p. 522, 1963; **18**, p.152, 1965.
31. L.Onural and P.D.Scott, Opt.Eng. **26** (11), *Digital decoding of in-line holograms*, 1124 – 1132, 1987.
32. L.Yaroslavsky, N.Merzlyakov, *Methods of digital holography*, M.: Nauka, 192 p, 1977
33. L Yaroslavsky, Digital signal processing in optics and holography, M.: Radio i svyaz, 296 p, 1987.
34. H.J.Caulfield, *Handbook of Optical Holography*, **1**, Academic Press, N.Y, 376 p., 1979.
35. M.B.Vinogradova, O.V.Rudenko,A.P.Suhorukov, *Theory of Waves*, "Nauka", Moscow, 384 p., 1979.
36. J.W. Goodman. *Introduction to Fourier optics*. McGraw-Hill, New York, 1968.

Chapter 4

DIGITAL HOLOGRAPHY APPLIED TO MICROSCOPY

Christian Depeursinge

*Institute of Imaging and Applied Optics, Ecole Polytechnique Fédérale de Lausanne,
CH-1015 Lausanne, Switzerland.*

Abstract: Digital Holographic Microscopy (DHM) is an imaging technique offering both sub-wavelength resolution and real time observation capabilities. The reconstruction of the wavefront from the hologram provides the amplitude and the absolute phase of the wave diffracted by the microscopic objects. Absolute phase contrast yields longitudinal accuracies as low as one nanometer in air or even less in dielectric media. The lateral accuracy and the corresponding resolution can be kept at a sub-micron level by the use of a high Numerical Aperture (N.A.) microscope Objective (M.O.). In the present state of the art, it can be kept commonly below 600nm. The principles of hologram formation, acquisition and wavefront reconstruction from digital holograms, acquired in a non-scanned modality, are described in details. The role of the M.O. in the capture of high spatial frequencies components of diffracted light and their restitution for high fidelity imaging is developed. A variety of applications of this new type of optical microscopy are described: material research, surface and interface sciences, microtechnologies, micro-optics and MOEMS. Applications to cell dynamics studies: nano-movements and cyto-architectures deformations, are demonstrated. Tomographic images of tissues can be also obtained by DHM.

Key words: Coherent Imaging, Digital Holography, wavefront reconstruction, microscopy,

1. INTRODUCTION

Optical instruments have been invented to see at small scale. Commonly called “microscopes”, they have been used since many centuries. In occidental countries, such instruments are known from the end of the 16th century to improve human vision. Microscopy was soon identified (18th century) as an essential tool to develop investigations in biology and medicine. Cells and micro-organisms were identified for the first time in the 19th century. These observations can be considered as the beginning of the development of modern biology and medicine. Today microscopy, electron or optical, with wavelength extending from infrared to EUV: extreme UV and X-rays, plays an essential role in the development of micro- and nanotechnologies: photolithography, MEMS and MOEMS (Micro- Optical, Elecroanical and Mechanical Systems). Its use in material and surface sciences is also in strong progression. Nowadays, the perspective to reach the nanometer scale with optical instruments, easier to handle than electronical microscope, appears more and more clearly as an incitation to invent and develop new technological approaches. Digital Holographic Microscopy (designated by DHM in the following) is one of them.

Most of the early optical instruments ignored the undulatory character of light and were based on the laws of geometrical optics and on the Snell-Descartes law of refraction. The development of optical instruments was intimately tied up to the principle of real and virtual images formations for direct vision with the human eye or image pickup with a camera. Even in the present days, two centuries after the development of the principle of geometrical optics, most of the efforts are still put on optical design and realization of components such as lens systems and mirrors, optically corrected to improve the quality of real images taken by cameras. The quest for better design and systems achieving improved stigmatic requirements still appears as the “holy Grail” of optical imaging.

The discovery of the undulatory nature of light and the development of the concept of coherence were at the base of a better understanding of image formation in optics. Holography brought for the first time the evidence that recording an image was not indissociably tied up with the formation of a real image on a photographic plate or an electronic camera. Gabor had a first intuition, long before the availability of long coherence sources in optics (Gabor, 1948). This intuition proved its real value with the development of laser technology, which brought long coherence sources in the laboratories. Holography became wide spread as a practical tool in metrology, where high spatial resolution was made possible by exploiting phase data. Long coherence sources also enabled the realization of large dimension holograms. It was discovered that holography made possible the restoration

of the third dimension of objects, delivering a full three-dimensional image. In microscopy, holography was deemed, a long time ago, to bring the capability of “seeing in 3D” microscopical objects (Pluta, 1988). The capture of holograms was however so delicate that no real exploitation of the idea was pursued. In the field of biology and medicine, holography, applied to cell and tissue investigations in 3D, appeared as an attractive tool, but was rarely used in its early, non digital phase (De Haller et al., 1995).

The idea of the reconstruction of a wavefront from the hologram, yielding the exact replica of the object, was a great assess of the development of holography. Its practical implementation by the recourse to digital processing methods appeared as an appealing and fundamentally new perspective in holography. The introduction of informatics was favoured by the rapid development of digital processing techniques which brought computing facilities on a wider basis by the large spread of PC's or digital signal processors. In Digital Holography (DH) the complex wavefront is reconstructed digitally by using a computer. The data are provided by a hologram captured and digitalized with an electronic camera. The most recent informatics tools have permitted the rapid calculation and simulation of wave propagation. This circumstance appeared as the second technological breakthrough, after laser invention, which has promoted digital processing of holographic data. These developments, designated as “digital holography”, have presented the great advantage of providing quantitative data as well as simplified optical manipulations and improved flexibility (see (Goodman et al., 1967), (Kronrod et al., 1972), (Yaroslavskii et al., 1980), (Schnars et al., 1994), (Schnars, 1994), (Coquoz et al., 1995), (Coquoz, 1994)). The consequences of this evolution or revolution in optics was to push holography out of optics lab's, making it more practical in the industrial environment and research institutes.

For the first time also, several constraints usually considered as mandatory in classical image forming optics, could be relaxed. It was henceforth possible to correct lens defects, aberrations in particular, by pure digital means. The perspective to postpone to a data processing step, most of the difficulties met in the practical implementation of sophisticated stigmatic optical systems appears, from now on, as appealing and prone to open new avenues in optical design.

The fact that the image acquisition in optics has become progressively dissociated from the step of forming a real image on a camera or the human eye, is a major outcome of modern coherence imaging and will constitute a fundamental breakthrough in optical imaging technology. We present here the evidence of this evolution in the field of optical microscopy, but it is likely that similar evolution will reach a growing number of vision systems

at the macro scale as well. Vision systems, material inspection, biology, micro-manipulation and micro-robotics or even “micro- or nano-workshops” are concerned by the development of so-called “coherence imaging” and could have a major technological impact.

One attracting aspect of coherence imaging based on digital processing of holograms is its magnificent completion in its role of vision system, giving access to full 3D vision capabilities, and bringing ipso-facto, its utmost flexibility in data manipulation and faculty of adaptation to vision problems, more specifically in microscopy, but also in metrology, visual inspection and so on. In biology and medicine, these breakthroughs paved the way to the use of (DHM) for the observation and analysis of cells and tissues.

1.1 Digital Holographic Microscopy

Application of digital holography to imaging microscopic objects has been proposed by several authors in the past.

Cuche et al. (Cuche et al., 1997) has proposed, in 1997 already, an approach based on the use of separate object and reference wave. Improved signal to noise ratio could be obtained in principle for absorbing specimens by having recourse to an external reference wave, the intensity and polarization of which can be controlled and adapted to the object wave characteristics. This separation has the consequence that the source must have some degree of coherence in space and time which limits the choice of the source to gas or solid state lasers, semiconductor lasers in particular. The interest of the proposed optical setup has been that high lateral resolutions could be achieved by intercalating high N.A. objectives between the specimen and the CCD camera. Lateral resolution, which could be made less than a few hundreds of nanometers. Combined with an axial resolution in the range of only a few nanometers, DHM certainly has appeared, in many applications, as attractive tools for the observation of surfaces, materials, MEMS or MOEMS (Micro- Optical, Electronic and Mechanical Systems) at the nano- scale. The holographic approach seems more robust and simple than scanning white light interferometers which require a vertical scanning mechanism and special objectives like Mirau interferometric objectives or a pair of objectives in a Linnick configuration. In biology, DHM has appeared as a unique tool to evaluate shape and deformations at the nano-scale and very tiny variations of refractive index. This achievement was made effective by developing a modified microscope design including an external reference beam and yielding slightly off-axis digital holograms of microscopic objects. Essential was also to develop the theory and the

corresponding computer program to correct for the phase modifications introduced by the use of a microscope objective intercalated between the sample and the CCD chip (Cuche, Marquet et al., 1999; Dubois, Joannes and Legros, 1999; Indebetouw et al., 1999). An interactive computer based environment allows to reconstruct easily wavefronts from intricate digital holograms resulting from wavefront deformation by optical components and the combination with a reference wave, itself resulting from complex wavefront deformation along the optical path of the optical setup. This combination of optical design and software for digital hologram processing is certainly a major contribution of our group to DHM. Careful consideration of the phase correction introduced by the M.O. intercalation has been a major preoccupation of our research group. To reach that goal, effective Optical Transfer Functions (OTF) of high N.A. objectives have been evaluated experimentally and the obtained data are introduced in the reconstruction algorithm (Marian, 2005t). Most of the perturbing aberrations are minimized with the above-described procedure and high quality reconstruction of wavefronts is henceforth a reality.

An elegant approach, also using a microscope objective, has been proposed by Indebetouw (Indebetouw et al., 1999), (Indebetouw et al., 2001). The formation of a hologram is achieved by taking the interference of the image of the specimen with a reference wavefront obtained by spatially filtering the object beam. A major advantage of this approach is that the source is incoherent in space and time and therefore can be very simple and cheap. Though appealing in its principle, spatially filtering the object wave causes a very important loss of light intensity of the reference wave, which is largely unfavourable to the signal to noise ratio and ultimately to image quality.

Phase interpretation: Generally the phase of the wavefront emanating from a macroscopic object cannot be interpreted directly because of the important phase diversity caused by the irregularity of the object surface and the presence of an important speckle contribution originating from the presence of multiple scattering centres in the object. Some data on the phase variations of the wavefront could however be retrieved from multiple hologram taken in the time course or at several distances. These phase variations, restored from the speckle pattern, could only be interpreted in term of a deformation or illumination wave modifications. This approach is commonly developed in Electronic Speckle Interferometry (ESPI).

In microscopy, the perspectives appear much more favorable and in a sense unique: the important speckle noise due to the plurality of scattering centres prevailing at the Macro- scale becomes much less important. In Digital Holographic Microscopy (DHM) (Cuche, Bevilacqua et al., 1999;

Cuche, Marquet and Depeursinge, 1999), the method developed rests upon the fact that the relatively low number of scattering centers exists in a volume of size comparable to a few wavelengths (tens of microns). Multiple scattering appears therefore as a negligible perturbation and do not generate significant speckle noise from the volume of the observed sample. The first Born approximation can be considered as a valid approximation and this trade-off appears as a favourable circumstance for microscopic observations. The phase of the complex reconstructed wavefront could be interpreted for the first time in term of the photon time of flight or optical pathlength (Cuche, Marquet and Depeursinge, 1999). Phase provides indirect data on the topology or morphology of objects down to the subwavelength range and can be used for 3D presentation of the microscopic object.

The originality of our approach is based on two ingredients giving access to microscopic world: both a modified microscope design yielding digital holograms of microscopic objects has been proposed and an interactive computer program has been developed to reconstruct easily wavefronts from the acquired digital holograms. Simultaneous amplitude and quantitative phase contrast images are obtained from the reconstructed complex wavefront. The reconstruction method computes the propagation of the complex optical wave front diffracted by the object and is used to determine the refractive index and/or shape of the object with accuracy in the nanometer range along the optical axis.

2. DIFFRACTION THEORY

2.1 Free space propagation

Huyghens law states that each point of a wave front can be considered as the source of a spherical wave outgoing to infinity. The superposition of all these waves generates at a further time t a propagated wavefront.

In a mathematical language, we say that the field $u(\mathbf{r}, t)$ at a position \mathbf{r} and at a time t is related to the field $u(\mathbf{r}', t')$ by the integral equation:

$$u(\mathbf{r}, t) = \iiint_{-\infty}^{\infty} \int u(\mathbf{r}', t') G(\mathbf{r}, t, \mathbf{r}', t') d\mathbf{r}' dt' \quad (4.1)$$

where $G(\mathbf{r}, t, \mathbf{r}', t')$ is the Green function, solution of the wave equation:

$$\nabla^2 u - \frac{1}{c^2} \frac{\partial^2 u}{\partial t^2} = -4\pi\delta(\mathbf{r}, t) \quad (4.2)$$

Considering the harmonic decomposition of the time variations of the wavefield:

$$u(\mathbf{r}, t) = \int_{-\infty}^{\infty} U(\mathbf{r}, \omega) e^{-i\omega t} d\omega \quad (4.3)$$

The propagation equation (4.2) will transform in the following Helmholtz equation:

$$(\Delta^2 + k^2) G_k(\mathbf{r}, \mathbf{r}') = -4\pi\delta(\mathbf{r} - \mathbf{r}') \quad (4.4)$$

where we have put:

$$k = \omega/c \quad (4.5)$$

Passing in spherical coordinates, (4.4) can be solved exactly for the free space propagation of the wave in 3D:

$$G_k^{(\pm)}(\mathbf{r} - \mathbf{r}') = \frac{e^{\pm ik|\mathbf{r} - \mathbf{r}'|}}{|\mathbf{r} - \mathbf{r}'|} \quad (4.6)$$

These are the Green functions for the 3D Helmholtz equation. They describe the harmonic response of an “ingoing” (G_k^-) and “outgoing” (G_k^+) spherical wave. A general solution to the Helmholtz equation is given by a linear superposition of ingoing and outgoing waves. The spatial dependence of the harmonic decomposition of the wave $U(\mathbf{r})$ obeys the integral equation:

$$U(\mathbf{r}) = \iiint_{-\infty}^{\infty} U(\mathbf{r}') G(\mathbf{r}, \mathbf{r}') d^3 \mathbf{r}' \quad (4.7)$$

2.2 Diffraction theory¹

In general, (4.7) does not provide a practical method to compute the propagated field, since it assumes that $U(\mathbf{r}')$ is already known at every space point \mathbf{r}' . It rather appears as an implicit equation giving $U(\mathbf{r})$. A much

¹ For fundamental aspects, see Born, M. and Wolf, E., Principle of optics (Cambridge Univ. Press, 1999)., Goodman, J., Introduction to Fourier Optics (McGraw-Hill, 2005).

more useful equation would be the one that permits to compute the 3D field distribution from the field determined on a sub-region of space. This goal can be reached by applying Gauss theorem applied to vector fields:

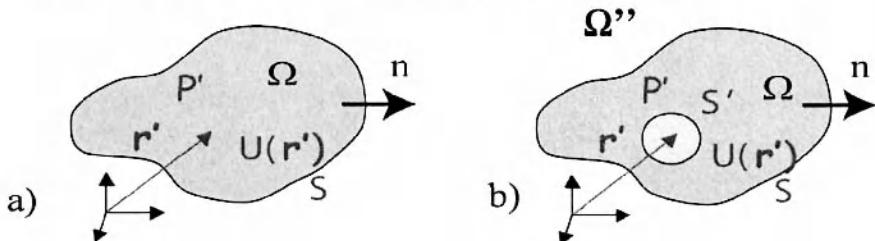


Figure 4-1. a) Ω is a volume surrounded by the surface S and over which the vector fields U and V are continuous. Ω'' is the volume complementary to Ω . b) Same volume Ω where a spherical void with surface S' and surrounding point r' is considered.

Gauss theorem can be applied to the vector fields U and V supposed to be continuous over Ω :

$$\iiint_{\Omega} \left(U \Delta^2 V - V \Delta^2 U \right) d^3 r = - \iint_S \left(U \frac{\partial}{\partial n} V - V \frac{\partial}{\partial n} U \right) dS \quad (4.8)$$

If U and V satisfy the Helmholtz equation the left member of equ.(4.8) cancels, so we have:

$$\iint_S \left(U \frac{\partial}{\partial n} V - V \frac{\partial}{\partial n} U \right) dS = 0 \quad (4.9)$$

Further, as a particular solution of Helmholtz equation, V will be simply chosen equal to the outgoing Green function $G_k^+(\mathbf{r}-\mathbf{r}')$. Let us also consider a new surface S (Figure. 4.1 b)) surrounding Ω and composed of S of Figure 4.1 a) and the additional surface S' of a small sphere surrounding P' . The computation of the surface integral of $\frac{\partial}{\partial n} G_k^+(\mathbf{r}-\mathbf{r}')$ over the small sphere S' will give 4π . At the limit of S' approaching P' , (4.9) will become:

$$\begin{aligned} \iint_{S \cup S'} \left(U \frac{\partial}{\partial n} G_k^+ - G_k^+ \frac{\partial}{\partial n} U \right) dS = \\ \iint_S \left(U \frac{\partial}{\partial n} G_k^+ - G_k^+ \frac{\partial}{\partial n} U \right) dS - U(\mathbf{r}') \cdot 4\pi = 0 \end{aligned} \quad (4.10)$$

and therefore:

$$U(\mathbf{r}') = \frac{1}{4\pi} \iint_S \left(U \frac{\partial}{\partial n} G_k^+ - G_k^+ \frac{\partial}{\partial n} U \right) \quad (4.11)$$

By considering an integration volume Ω'' complimentary to Ω , it may be easily shown that

$$\frac{1}{4\pi} \iint_{S''} \left(U \frac{\partial}{\partial n} G_k^+ + G_k^+ \frac{\partial}{\partial n} U \right) = 0 \quad (4.12)$$

Where S'' is the limiting surface of Ω'' . It tends to transform in an infinite plane when Ω grows to infinity. S'' becomes as close as possible to the infinite plane S : $S'' \sim S$. Finally, it follows from (4.12) that the two parts of the integral (4.11) over S are in fact equal and that the final expressions of the field $U(\mathbf{r}')$ is:

$$U(\mathbf{r}') = \frac{1}{2\pi} \iint_S \left(U \frac{\partial}{\partial n} G_k^+ \right) dS = \frac{-1}{2\pi} \iint_S \left(G_k^+ \frac{\partial}{\partial n} U \right) dS \quad (4.13)$$

These are the Rayleigh-Sommerfeld diffraction formula of the First and Second kind respectively.

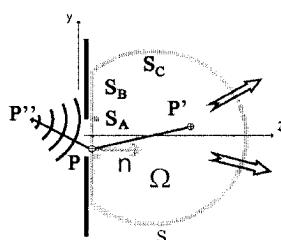


Figure 4-2. Surfaces S and volume Ω considered for the calculation of the Field U at point P' diffracted by a semi-transparent window S_A cut in an opaque screen.

We apply Rayleigh-Sommerfeld formula of the first kind to the calculation of $U(\mathbf{r}')$, by computing the surface integral on S , surrounding the volume Ω (see *Figure 4-2*). S is composed of three sub-surfaces: S_A , S_B and S_C which are respectively the window in the screen letting the field propagate in the volume Ω , the screen stopping the irradiation and the sphere delimiting the volume Ω . Letting the volume Ω grow to infinity, it can be inferred from the fact that the irradiation energy is finite in space and time, that the contribution of surface S_C goes to zero. Similarly the contribution of S_B is assumed to be zero. The only contribution to the integrand originates from S_A where \mathbf{n} is along z:

$$\begin{aligned}\frac{\partial}{\partial n} G_k^+(\mathbf{r}-\mathbf{r}') &= \frac{\partial}{\partial z} \frac{\exp(ik \cdot |\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|} \\ &= \left(ik - \frac{1}{|\mathbf{r}-\mathbf{r}'|} \right) \cdot \frac{\exp(ik \cdot |\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|} \cdot \frac{\partial}{\partial z} |\mathbf{r}-\mathbf{r}'|\end{aligned}\quad (4.14)$$

where $|\mathbf{r}-\mathbf{r}'| = \sqrt{(x-x')^2 + (y-y')^2 + (z-z')^2}$

and $\frac{\partial}{\partial z} |\mathbf{r}-\mathbf{r}'| = (z-z')/|\mathbf{r}-\mathbf{r}'| = \cos(\mathbf{n}, (\mathbf{r}-\mathbf{r}'))$

As $k = 2\pi/\lambda \ll 1/|\mathbf{r}-\mathbf{r}'|$ in the far field region and $\cos(\mathbf{n}, (\mathbf{r}-\mathbf{r}')) \approx -1$ close to the optical axis and for small apertures (this assumption is not valid for high N.A. objectives), we have the following approximation:

$$\begin{aligned}\frac{\partial}{\partial n} G_k^+(\mathbf{r}-\mathbf{r}') &= \left(ik - \frac{1}{|\mathbf{r}-\mathbf{r}'|} \right) \cdot G_k^+(\mathbf{r}-\mathbf{r}') \cdot \cos(\mathbf{n}, (\mathbf{r}-\mathbf{r}')) \approx \\ &\approx (-ik) \cdot G_k^+(\mathbf{r}-\mathbf{r}')\end{aligned}\quad (4.15)$$

and finally:

$$\begin{aligned}
 U(\mathbf{r}') &= \frac{1}{2\pi} \iint_{S_A} \left(U \frac{\partial}{\partial n} G_k^+ \right) dS = \frac{-ik}{2\pi} \iint_{S_A} \left(U(\mathbf{r}) \cdot G_k^+(\mathbf{r} - \mathbf{r}') \right) dS = \\
 &= -\frac{i}{\lambda} \iint_{S_A} \left(U(\mathbf{r}) \cdot G_k^+(\mathbf{r} - \mathbf{r}') \right) dS
 \end{aligned} \tag{4.16}$$

which is known as the Huyghens-Fresnel integral.

Mathematically this means that the Field $U(\mathbf{r}')$ in the plane z' can be calculated from the field in plane z : $U(\mathbf{r})$, with the expression of the Huyghens Fresnel integral.

Suppose, in the following, that we concentrate on the mathematical relationship between the fields in two planes distant of $z' - z = d$ and that will be, conventionally, parameterized in the plane perpendicular to the optical axis by ξ and η in the target plane T , and x and y in the origin plane (Source plane): S .

In the paraxial approximation we have:

$$\begin{aligned}
 |\mathbf{r} - \mathbf{r}'| &= \sqrt{d^2 + (x - \xi)^2 + (y - \eta)^2} \\
 &\approx d \left[1 + \frac{1}{2} \left(\frac{x - \xi}{d} \right)^2 + \frac{1}{2} \left(\frac{y - \eta}{d} \right)^2 \right]
 \end{aligned} \tag{4.17}$$

and

$$k = \frac{2\pi}{\lambda} = \frac{n\omega}{c} \quad (n: \text{index of refraction}, \omega: \text{pulsation}, c: \text{light velocity}) \tag{4.18}$$

The Huyghens-Fresnel integral can be developed in the following form:

$$\begin{aligned}
 U_T(\xi, \eta) &= -\frac{i}{\lambda} \iint_{S_A} (U_s(\mathbf{r}) \cdot G_k^+(\mathbf{r} - \mathbf{r}')) dS = \\
 &= -\frac{i \exp(ikd)}{\lambda d} \iint_{S_A} \left(U_s(x, y) \cdot \exp \left\{ \frac{ik}{2d} [(x - \xi)^2 + (y - \eta)^2] \right\} \right) dx dy = \\
 &= -\frac{i \exp(ikd)}{\lambda d} \exp \left[\frac{ik}{2d} (\xi^2 + \eta^2) \right] \cdot \\
 &\quad \cdot \iint_{S_A} \left(U_s(x, y) \cdot \exp \left[\frac{ik}{2d} (x^2 + y^2) \right] \exp \left[-i \frac{k}{d} (x\xi + x\eta) \right] \right) dx dy \\
 &= -\frac{i \exp(ikd)}{\lambda d} \exp \left[\frac{ik}{2d} (\xi^2 + \eta^2) \right] \cdot \mathbf{F}_F \mathbf{q}_{Fourier} \left\{ U_s(x, y) \cdot \exp \left[\frac{ik}{2d} (x^2 + y^2) \right] \right\} = \\
 &= -i \cdot \exp(ikd) \cdot \mathbf{F}_F^\sigma [U_s(x, y)]
 \end{aligned} \tag{4.19}$$

Where the 2D Fresnel transform is defined as:

$$\begin{aligned}
 \mathbf{F}_F^\sigma [U(x, y)] &= \\
 &= \frac{1}{\sigma^2} \exp \left[\frac{i\pi}{\sigma^2} (\xi^2 + \eta^2) \right] \cdot \mathbf{F}_F \mathbf{q}_{Fourier} \left\{ U(x, y) \cdot \exp \left[\frac{i\pi}{\sigma^2} (x^2 + y^2) \right] \right\}
 \end{aligned} \tag{4.20}$$

where we have posed

$$\sigma = \sqrt{\lambda d} = \sqrt{2\pi \frac{d}{k}} \tag{4.21}$$

The Fresnel transform is a function of the continuous parameter σ . If σ goes to infinity, Fresnel transform become identical to Fourier transform and, on the other side, if σ goes to zero, Fresnel transform becomes the “identity” transform multiplied by i . For more information concerning the mathematical aspects, see (Liebling et al., 2003), (Liebling, Blu et al., 2004), (Liebling, 2004).

The Fresnel transform establishes an invertible and bijective correspondence between the Propagated field $U_T(\xi, \eta)$ in the target plane $\xi, \eta: (T)$ and the source field $U_S(x, y)$ in the source plane $x, y: (S)$. This property establishes the well-posed nature of holographic imaging and thereby contrasts with the “ill-posed” character of intensity based imaging modalities.

Discretized, the Fresnel transform take the form:

$$U_T(m\Delta\xi, n\Delta\eta) = -i \exp(2i\pi d/\lambda) \cdot DF_{Fresnel} T[U_S(k\Delta x, l\Delta y)] \quad (4.22)$$

where $\Delta\xi, \Delta\eta$, are the sampling intervals in the target plane and $\Delta x, \Delta y$ the sampling intervals in the source plane. m, n, k, l are integers. The following expression of the Discrete Fresnel Transform can be given:

$$\begin{aligned} DF_{Fresnel} T[U(k\Delta x, l\Delta y)] &= \\ &= \frac{1}{\sigma^2} \cdot \exp\left\{\frac{i\pi}{\sigma^2} [m^2 \Delta\xi^2 + n^2 \Delta\eta^2]\right\} \cdot \\ &\cdot DFT\left[U(k\Delta x, l\Delta y) \cdot \exp\left\{\frac{i\pi}{\sigma^2} [k^2 \Delta x^2 + l^2 \Delta y^2]\right\}\right]_{m,n} \end{aligned} \quad (4.23)$$

where DFT is the Discrete Fourier Transform. The inverse discrete Fresnel transform, which can be computed in a discrete form, gives the explicit expression for the source wavefield of the object:

$$U_S(k\Delta x, l\Delta y) = i \exp(-2i\pi d/\lambda) \cdot DF_{Fresnel}^{-1} [U_T(m\Delta\xi, n\Delta\eta)] \quad (4.24)$$

$$\begin{aligned} DF_{Fresnel}^{-1} [U(m\Delta\xi, n\Delta\eta)] &= \\ &= \sigma^2 \exp\left\{-\frac{i\pi}{\sigma^2} [k^2 \Delta x^2 + l^2 \Delta y^2]\right\} \cdot \\ &\cdot DFT^{-1}\left[U(m\Delta\xi, n\Delta\eta) \exp\left\{-\frac{i\pi}{\sigma^2} [m^2 \Delta\xi^2 + n^2 \Delta\eta^2]\right\}\right]_{k,l} \end{aligned} \quad (4.25)$$

DFT or *DFT*[†] can be easily and efficiently computed with a fast decimation algorithm (FFT), making these propagation computations in a fraction of a second.

3. HOLOGRAM FORMATION AND WAVEFRONT RECONSTRUCTION

3.1 Hologram formation

The hologram of an object is obtained by recording the interference of the wave diffracted by the object and a reference wave, provided that a non-zero degree of mutual coherence can be generated between the illuminating beam and the reference beam. In Digital Holography, this interference pattern is captured by a camera, today mainly CCD or CMOS cameras, and directly transmitted to the computer after digitization. The signal digitized is generally a non-linear function of the light intensity impinging on the camera plane. Nevertheless CCD provides a response approximately linear with intensity. In all cases, the signal can be linearized after calibration.

The field resulting from the superposition of the reference wave R and object wave O is simply given by $(R+O)$ if both fields are linearly polarized and collinear. The hologram recorded intensity is given by the simple mathematical expression:

$$I_H(x, y) = (R + O)^* \cdot (R + O) = |R|^2 + |O|^2 + R^* O + R O^* \quad (4.26)$$

The two first terms of this development are called “zero order terms” and are slowly varying in time or space and do not depend on the phase of R and O . They can be easily eliminated by high pass, time or spatial filters. The last two terms, on the contrary are called “interference terms” or “crossed terms”. They are sensitive to the phase difference between R and O and will be processed to retrieve the phase information

3.2 Plane and spherical object and reference wave

In general R and O waves are spherical waves, plane waves or a combination of them. Plane waves are described mathematically by:

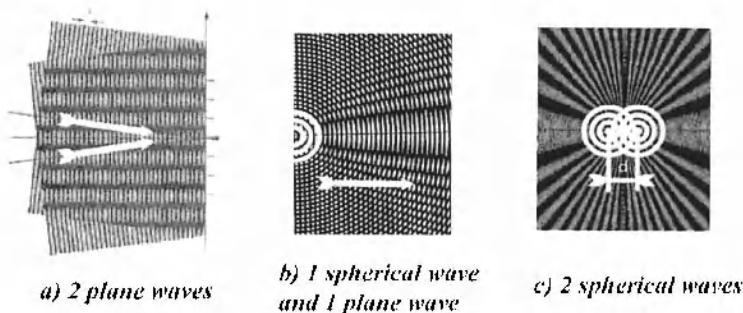
$$U(\mathbf{r}) = \exp(i \cdot \mathbf{k} \cdot \mathbf{r}) \quad (4.27)$$

and spherical waves (centred at \mathbf{r}') by:

$$U(\mathbf{r}) = \exp(i \cdot k \cdot |\mathbf{r} - \mathbf{r}'|) / |\mathbf{r} - \mathbf{r}'| \quad (4.28)$$

$U(\mathbf{r})$ may represent optionally the object wave $O(\mathbf{r})$ or the reference wave $R(\mathbf{r})$. The result of the superposition of both waves is the combination of plane and/or spherical waves and is illustrated on *Figure 4-3*. The combination of two plane waves gives planes parallel to the sum of \mathbf{k}_1 and \mathbf{k}_2 . They are standing waves if $|\mathbf{k}_1| = |\mathbf{k}_2|$. Their interception by the hologram plane gives a straight parallel fringe pattern.

The combination of spherical waves with plane waves (see *Figure 4-3 b*) gives a series of paraboloids with axis in the direction of the wavevector of the plane wave. Their intersection with the hologram plane gives a Fresnel pattern in a plane normal to the axis and parabola in a plane parallel to the axis.



- Figure 4-3.* Standing wave patterns created by the interference of
- Two plane waves with different k -vectors,
 - A spherical and a plane wave,
 - Two spherical waves with different centers (separated by distance d)

Finally, the combination of two spherical waves; object and reference plane, yield a set of hyperboloids with the axis joining the two sources emitting the spherical wave and distant of d . Their intersections with a plane normal to the axis give a typical Fresnel pattern. If the hologram plane is parallel to the axis, the fringes take the form of a set of hyperbola. If the

distance of the axis to the plane divided by d is large enough, the fringes are close to a set of straight parallel fringes, the spacing of which are inversely proportional to d .

These various configurations can be exploited to generate various type of holograms that can be digitized and processed to reconstruct the wavefronts generated by the object.

3.3 Wavefront reconstruction

Reconstructing the object wavefield can be achieved by forming the product:

$$R \cdot I_H = R \cdot (R + O)^* \cdot (R + O) = R \cdot \left(|R|^2 + |O|^2 \right) + |R|^2 O + R^2 O^* \quad (4.29)$$

The first term on the right part of (4.29) corresponds to the reconstruction of the zero order term and should be eliminated as completely as possible. The second term corresponds to the virtual image and restores exactly in principle the object wave after computation of the object wavefield propagated over the exact distance d separating the object from the hologram. The third term corresponds to a wave outgoing from the hologram and converging to the real image of the object, which appears as a mirror image of the virtual image (with respect to the hologram). The real image can be in principle retrieved by multiplying I_H by R^* and propagating it in the opposite direction by a distance $-d$, providing the wavefield O^* in the real image plane.

Propagating RI_H over a distance d is illustrated on *Figure 4-4*. In general, the reconstructed wavefront O in the object plane appears as mixed-up with the zero-order term and the blurred real image. This situation creates important perturbations and non-recoverable errors in the restoration of the object wavefield. It is therefore essential to filter out the zero-order term and the real image.

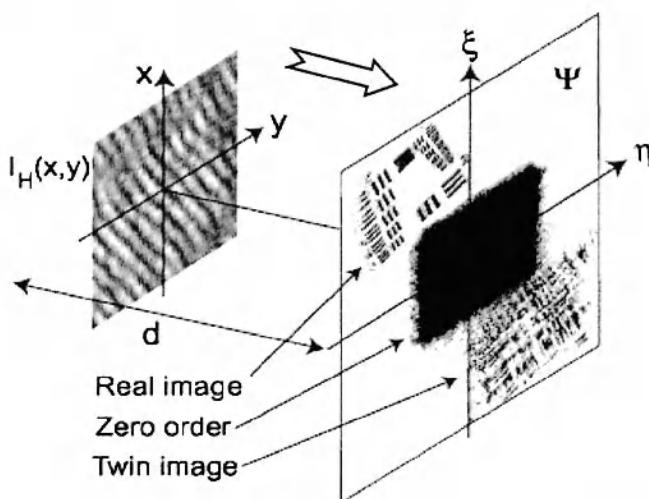


Figure 4.4. Digitized hologram with its reconstruction at distance d of the object (USAF test target, intensity image). Courtesy of E. Cuche, (Cuche, 2000).

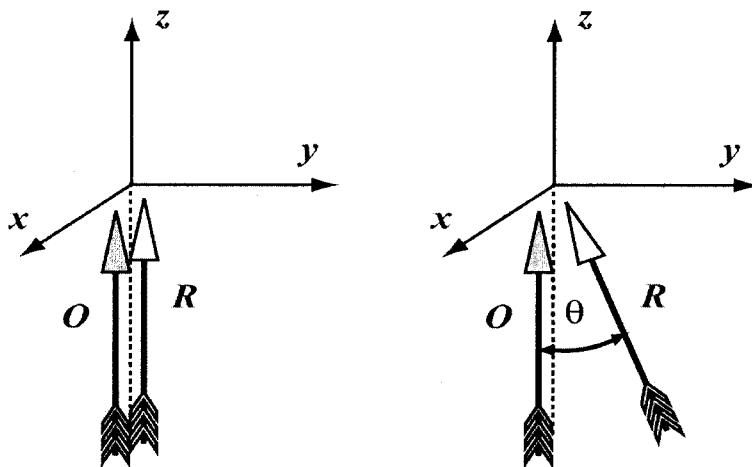
Provided that the fourth term R^*O can be isolated by a suitable filter, the object wave $O(x,y)$ can be easily retrieved as a complex field in the plane of the hologram by multiplying the hologram intensity $I_H(x,y)$ by the reference wave R :

$$O(x,y) = I_{H \text{ filtered}}(x,y) \cdot R(x,y) / |R \cdot R^*| \quad (4.30)$$

O exactly matches the wave originally emitted by the object, which is normally located behind the hologram with respect to the observer and therefore is considered as a virtual image. In digital holography, this operation can be achieved by processing numerically the data from one or several holograms acquired successively in space or time. This filtering operation can be performed in the time and / or in the space domain.

3.3.1 In-line and off-axis holography

Holograms can be taken with reference waves propagating in parallel with object waves, this is called “in-line” holography. They can be taken also with a reference wave tilted by an angle θ : this arrangement is called “off-axis” holography. These variants are illustrated on Figure 4-5.



*Figure 4-5. Left: In-line holography: Object and Reference wave propagate in parallel
Right: Off line holography: Object wave and Reference wave propagate with an angle θ .*

In-line holography is often used to image and locate particles in depth. This approach is based on the fact that the particles constitute a sparse medium and that the illuminating wave is only weakly attenuated by the medium. The illuminating wave can therefore play the role of reference wave. This situation is generally found with high energy radiations like Extreme UV and X-rays. If the object is strongly attenuating or strongly diffracting, it is necessary to add a reference wave which can be superposed to the object wave with a beam splitter (see for example lensless holographic microscopes described in section 5.1). Elimination of the twin image is a difficulty for in-line holography. It may be solved either by filtering it in the time domain by heterodyning the reference wave, or by using strongly diverging beams such as those used in lensless holographic microscopy. In this last case the magnification for the virtual and real images are very different and there interferences are less pronounced.

Off-axis holography is generally used in situations where the object is cumbersome and cannot be traversed by the reference beam. In the case of digital holography, the off-axis geometry presents the advantage of providing a simple method to filter out the terms in the hologram corresponding to the real image (see section 3.3.3).

3.3.2 Filtering in the time domain

In the time domain, filtering requires the acquisition of several holograms while changing the phase difference between R and O : this

approach is called *phase shifting digital holography*. Three or four phase shifts (multiple of $\pi/2$) are generated by a phase shifter. A simple mean to realize such a phase shifter is using a piezo- activated mirror driven by up-stepping or down-stepping voltages. This technique has been used in digital holographic microscopy to form images of microscopic objects with microscope objectives (Dubois, Joannes and Legros, 1999; Joannes et al., 2003; Mills et al., 2005; Yamaguchi et al., 1997; Yamaguchi et al., 2002). Other means include the heterodyning of the reference beam with a tandem of frequency shifting acousto-optic modulators (Gross et al., 2005; Le Clerc et al., 2000). Stepping up or down the phase shift eliminates, after accumulation of holograms, the R^*O term and RO^* term respectively. A minimum of three (most often four) holograms is needed to obtain rejection of unwanted terms.

3.3.3 Filtering in the spatial domain

Simpler instrumentation can be achieved by processing data in the spatial domain: modulation techniques, which are required in the time domain, are not needed in the spatial domain. Introducing a small angle between the propagation direction of the object and reference wave makes correlations, normally measurable in the time domain, appear in the spatial domain: heterodyning in particular is achieved by tilting the reference wave (Cuche, Marquet and Depeursinge, 1999),(Cuche, 2000). This tilt in off-axis holography, generates a wave component in the x - y plane, which introduces a high frequency spatial modulation of the hologram intensity.

Considering a plane wave, inclined on the x -axis at an incidence angle of θ , the complex expression of the reference wavefield in the x - y plane is:

$$R(x, y) = |R| \cdot \exp(i \cdot k \cdot \sin \theta \cdot x) \quad (4.31)$$

And the hologram intensity:

$$I_H = (|R|^2 + |O|^2) + |R| \exp(-ik \sin \theta \cdot x) O + |R| \exp(ik \sin \theta \cdot x) O^* \quad (4.32)$$

Fourier analysis of the spatially developed hologram provides a simple way to filter out unwanted contributions to the hologram and makes straightforward the object wavefront reconstruction. Analysing the spatial spectrum of I_H and assuming that the object spectrum O is limited to less than half of the linear extension of the Fourier domain, we observe that the

second and third terms of (4.32) correspond to the object spectrum and to its conjugate in the Fourier domain. They are translated to the left for the virtual image O and to the right for the real image O^* , by a value equal to $k \sin\theta$. This shift in the Fourier domain can therefore be adjusted by changing k or θ . From the point of view of the extension of the Fourier representation of the object wavefield, i.e. ultimately, the resolution of the holographic image, the position that minimizes aliasing in the estimation of the object spectrum, is at the mid-distance to the extension of the Fourier domain. Nevertheless, this optimal choice is a theoretical limit and cannot be reached in practice, because of the presence of the Fourier contribution of the object spectrum at the zero order ($|O|^2$): this contribution is equal to the autocorrelation of the object spectrum and therefore occupies up to twice the spectrum of the object left in the Fourier domain (Goodman, 2005),(Marquet, 2003). Indeed, only one third of the 1D Fourier domain should be accessible to the determination of the Fourier spectrum of the object. Nevertheless, this estimation is pessimistic and provides a low limit to the spectrum extension of the object. The truth is somewhere between these two limits.

In order to illustrate the reconstruction of the wavefront in digital holography, the result of the digital acquisition of the hologram of a USAF test target is given on *Figure 4-6* (a). Part (b) represents the Fourier analysis of this hologram in an off-axis configuration. This analysis makes the different diffraction orders appear in the Fourier plane: the zero order of diffraction can be observed at the middle of the Fourier domain, near zero spatial frequencies. The +1 and -1 orders of diffraction are visible in the form of diffusely scattered spots, symmetrically on one side and the other of the zero order. They correspond to the R^*O and RO^* terms respectively, which contain respectively the spectral data necessary to reconstruct the wavefields in the virtual and real image planes. Part (c) finally shows the propagated wavefield in the plane of the virtual image: a focused image of the USAF test target can be seen on the amplitude image reconstructed from the R^*O term of the Fourier decomposition of the hologram. The other parts of the image show the propagated zero order terms, which appear as a bright spot at the center of the image and finally the unfocused image of the real image of the target. Some artefacts originating from parasitic scattering centers and multiple scattering may give rise to undesired spot and flakes over the image. It is therefore of uppermost importance to “clean” the reconstructed image by selecting carefully the contributing terms in the hologram spectrum.

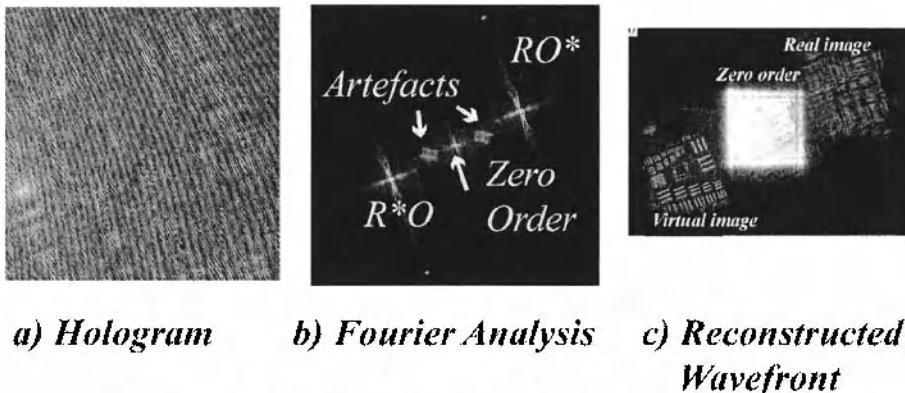


Figure 4-6. Illustration of the reconstruction process (Cuche, 2000).

Figure 4-7 illustrates the filtering of the hologram in the Fourier domain: Figure 4-7 (left) shows the delineated zone on the Fourier transform of the hologram. The criteria for the zone delimitation are, to some extent, arbitrary. They result from an interactive procedure by which the operator make a choice of what appears as being part of the spectrum of the object. After delimitation of the spectral components of the object, the wavefield is reconstructed and the image of the USAF target appears fully restored on the image.

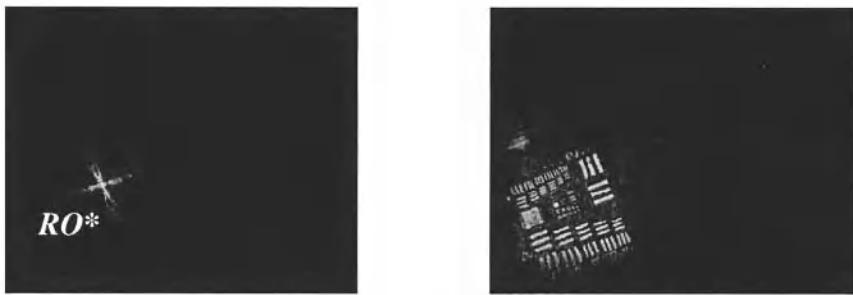


Figure 4-7. Role of filtering out RO^* : delineating the zone around the spot attributed to RO^* permits reconstructing only the virtual image (Cuche, 2000).

3.3.4 Hologram apodization

In coherent optics, as well as in signal processing, it is well known that the windowing or truncating of the signal in the space or time domain, gives rise to undesired fringes or oscillations, which obscure the image. A simple method has been proposed (Cuche et al., 2000) for designing apodized apertures with a transmission profile which follows a curve defined using a

cubic spline interpolation. The method has been applied in digital holographic microscopy to perform digitally the apodization of the aperture of holograms recorded by a CCD camera. The transmission of the apodized aperture is entirely defined by four parameters, which are adjusted iteratively to minimize intensity and phase fluctuations appearing in the images obtained by numerical reconstruction of the holograms. Apodization reduces the standard deviation of the reconstructed phase distribution from 1.6 nm to 0.15 nm, practically canceling the perturbations introduced by the windowing of the camera.

3.4 The problem of hologram sampling

The problem of hologram sampling may have appeared from the beginning of digital holography developments as an insurmountable difficulty: with a few microns between pixels, the spatial sampling provided by CCD or MOS camera are ten to fifty less dense than the sampling by other recording media used in holography: polymers, photorefractive materials or photographic plates. Therefore electronic cameras have soon appeared as media not suited to record off-axis holograms.

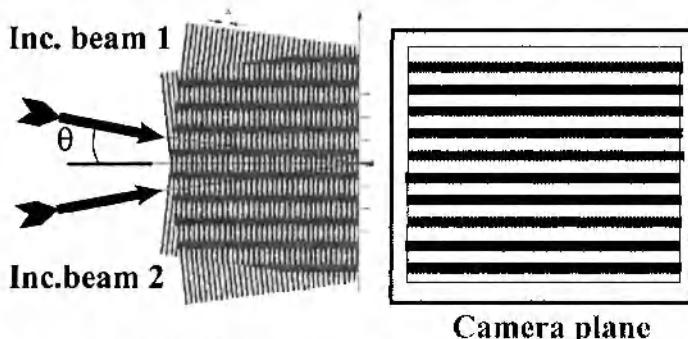


Figure 4-8. The pixel period of the camera determines the density of fringes in the hologram that can be correctly sampled by application of Shannon theorem.

Figure 4-8. illustrates the limitation introduced by electronic cameras in hologram sampling: according to Shannon theorem the shortest spatial period of fringes that can be correctly sampled is just twice the distance between two pixels, which in the present state of the art is around 3 microns for the highest density cameras. A periodicity of 6 microns appears as the shortest period sampled. The resulting angle between two interfering plane waves is therefore around 6°, which means that optical systems with only small aperture can be tolerated.

3.5 The role of Optical components

Some optical components may play an important role in the operation of a digital holographic microscope. A particular importance must be attached to the adaptation of the sampling capacity of the camera to the characteristics of the hologram. It was shown in the previous section that the pixel density of MOS could only sample properly fringes spaced out of more than approximately 6 microns, which is around ten times the wavelength λ . This means that the lateral component of the wavevector \mathbf{k} of the beam incident on the hologram plane should not be greater than grossly a tenth of its axial value. To image microscopic objects with the maximum resolution, the waves radiated by the object must be collected in the largest achievable solid angle. These high angle waves evidently do not satisfy the conditions required for proper sampling by the camera. A lens inserted between the object and the camera can play this role.

High Numerical Aperture (N.A.) microscope Objectives (M.O.) provide a simple mean to adapt the sampling capacity of the camera to the information content of the hologram. As illustrated on *Figure 4-9*, a lens or M.O. achieves a reduction of the k_x , k_y components of the \mathbf{k} vector components in the x,y plane perpendicular to the optical axis. The reduction factor is given by the magnification M of the M.O. The new components k'_x , k'_y of the \mathbf{k}' wavevector of the beam after the M.O., can be made as small as required by the Shannon theorem applied to the sampling capacity of the camera. Using a high magnification objective, the match can be optimized. In the same time, by maximizing the N.A, transverse resolution can be pushed to the limit of diffraction and sub-micron resolution can be easily achieved (ordinarily better than 600nm).

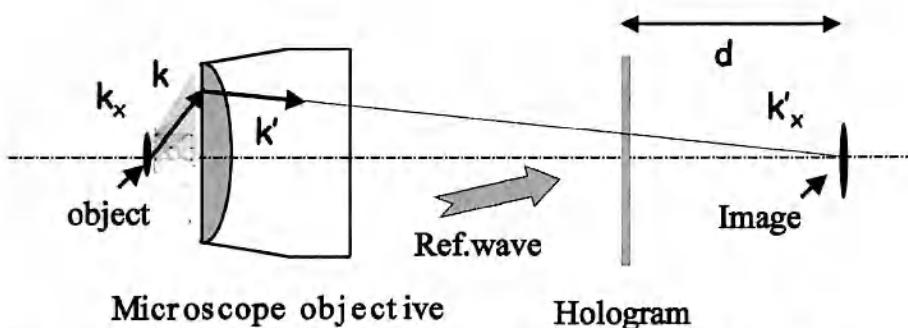


Figure 4-9. The insertion of a lens or microscope objective into the object beam permits to devide the k_x and k_y components of the diffracted beam by the magnification factor M of the M.O. The sampling of the beam by a digital camera can therefore be adapted to the angular spectrum of the beam emitted by the microscopic object.

After the introduction of the M.O., the reconstruction of the object wavefield yields the image of the microscopic object given by the M.O.. This image will be described by $O_i(\xi_i, \eta_i)$.

$O_i(\xi_i, \eta_i)$ can be computed by the convolution of the real object wavefield $O_o(\xi_o, \eta_o)$ by the transfer function $h(\xi_i, \eta_i, \xi_o, \eta_o)$:

$$O_i(\xi_i, \eta_i) = \iint h(\xi_i, \eta_i, \xi_o, \eta_o) \cdot O_o(\xi_o, \eta_o) d\xi_o d\eta_o \quad (4.33)$$

In a rough approximation, i.e. if we neglect the lateral extension of the point spread function and the influence of aberrations, The point spread function may be simplified in the following manner:

$$h(\xi_i, \eta_i, \xi_o, \eta_o) = C \cdot \exp \left[\frac{i\pi}{\lambda D} (\xi_i^2 + \eta_i^2) \right] \cdot \delta(\xi_i - M \cdot \xi_o) \cdot \delta(\eta_i - M \cdot \eta_o) \quad (4.34)$$

$$\text{with } D = f \frac{d_0}{d_i} \quad (4.35)$$

(f : focal length, d_0 object to lens distance, d_i lens to image distance.)

The conclusion is that, in a crude approximation the wavefront obtained after image reconstruction in the ξ_i, η_i plane can simply corrected by a multiplication by the inverse of the coefficient in (4.34):

$$\Phi(\xi_i, \eta_i) = C^{-1} \cdot \exp \left[\frac{-i\pi}{\lambda D} (\xi_i^2 + \eta_i^2) \right] \quad (4.36)$$

4. RECONSTRUCTION ALGORITHMS: NUMERICAL ASPECTS

4.1 Wavefront reconstruction

In a discrete form, the digital hologram is formed by sampling, over the plan $x-y$, the intensity of the wavefield resulting from the superposition of

the object wave $O(k\Delta x, l\Delta y)$ and a reference wave $R(k\Delta x, l\Delta y)$: $I_H(k\Delta x, l\Delta y)$ is the hologram intensity, sampled on the matrix of points $(k\Delta x, l\Delta y)$ corresponding to a subset of the matrix of pixels of the CCD or CMOS camera.

According to (4.30), $I_H(k\Delta x, l\Delta y)$ must be filtered in order to isolate the contribution of the OR^* terms to the hologram. As explained in section 3.3.3, the spectrum of $I_H(k\Delta x, l\Delta y)$ is first computed and a mask is interactively defined and applied subsequently to all reconstructions performed in real time.

Assuming that the reference wave is precisely known by its complex data matrix $R(k\Delta x, l\Delta y)$, the object wavefield O_H in the plane of the hologram could be computed as:

$$O_H(k\Delta x, l\Delta y) = I_{H \text{ filtered}}(k\Delta x, l\Delta y) R(k\Delta x, l\Delta y) / |R|^2 \quad (4.37)$$

In practical situation, however, the real reference wave $R(k\Delta x, l\Delta y)$ cannot be exactly known and the concept of a tentative reference wave, called “digital reference wave” $R_D(k\Delta x, l\Delta y)$ has been introduced (Cuche, Marquet and Depeursinge, 1999) to enable the concept of “parameter adjustment”. The proposed concept of digital reference wave is original and brings the flexibility needed to allow for a precise adjustment of the parameters for the digital reconstruction. It is a pure mathematical concept, having a representation in the computer only and mimicking the optical reference beam. This approach has allowed the correct restitution of the phase information in DHM. The digital reference wave is computed on the basis of a mathematical expression describing the plane wave and containing two parameters which are the components of the wavevector \mathbf{k} : k_x and k_y :

$$R_D(k\Delta x, l\Delta y) = |R_D| \cdot \exp[i(k_x \cdot k\Delta x + k_y \cdot l\Delta y)] \quad (4.38)$$

After having replaced the true reference wave $R(k\Delta x, l\Delta y)$ by the digital reference wave $R_D(k\Delta x, l\Delta y)$, the object wavefield O_T can be reconstructed by propagating the wavefield from the hologram plane x-y to the target plane $\xi-\eta$: (T)

$$\begin{aligned} O_T(m\Delta\xi, n\Delta\eta) = & -i \exp(2i\pi d/\lambda) \cdot DF_{Fresnel} T [O_H(k\Delta x, l\Delta y)] = \\ & -i \exp(2i\pi d/\lambda) \cdot DF_{Fresnel} T \left[I_{H_{\text{filtered}}}(k\Delta x, l\Delta y) R_D(k\Delta x, l\Delta y) / |R_D|^2 \right] \end{aligned} \quad (4.39)$$

4.2 Adjustment of reconstruction parameters

Many difficulties in the reconstruction procedure arise from the experimental errors made on the determination of the parameters of the reference wave (Cuche, Marquet and Depeursinge, 1999). The best approach is to resort to an adjustment procedure where the parameters are adjusted iteratively till phase will be constant on a calibration zone of the object itself or of a special calibration target.

A computer program may provide an interactive environment and automatic procedures to adjust easily the reconstruction parameters, in order to get true phase data on the microscopic field: a special calibration technique can be implemented for use on some selected part of the image, where constant phase data can be provided. For the purposes of phase-contrast imaging, a numerical procedure has been developed to compensate the deformation of the object wavefront produced by the microscope objective or other optical component of the microscope. Optical aberrations introduced by the optical components can also be corrected digitally in this way. The method requires the adjustment of several reconstruction parameters, which may be performed automatically (Colomb et al., 2006).

In more details, the method requires the acquisition of a single hologram of the object or test object, from which the phase distribution can be obtained simultaneously with intensity distribution at the surface of the object (Cuche, Marquet and Depeursinge, 1999). Then, by successive adjustments, $R_D(k\Delta x, l\Delta y)$ can be rendered as close as possible to $R(k\Delta x, l\Delta y)$. To achieve this goal, a calibration zone must be defined: the microscopic object to be reconstructed is very often surrounded by a zone, which is flat and serves as a reference plane. If this is not possible, the object can be removed and the calibration operation can be achieved in the absence of the object (transmission DHM) or on a flat mirror (reflection DHM). Once this zone is delimitated in an interactive way, the parameters of the digital

reference wave are determined so that the calibration zone appears as flat: this operation is sometimes called "flattening". An advantage of the developed approach based on the acquisition of a single hologram, is to offer both robustness and low noise, so that very high accuracies and resolutions can be achieved on the determination of the reconstruction parameters. More details can be found in the texts of Cuche et al.: (Cuche, 2000; Cuche, Bevilacqua and Depeursinge, 1999; Cuche, Marquet and Depeursinge, 1999).

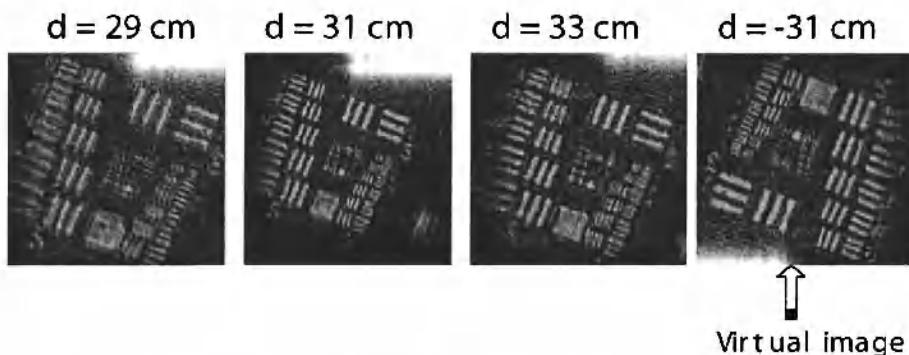


Figure 4-10. Electronic focusing obtained by adjusting the reconstruction distance d from 29 cm to 33cm: optimal focusing is obtained for $d=31\text{cm}$. Reconstructing the image for $d=-31\text{cm}$ yields the focused real image, rotated by π (Cuche, 2000).

After completion of this calibration operation, there is still one parameter to be adjusted: the reconstruction distance d . For this step, the criterion is the sharpness of the focused image. This adjustment can be done manually by changing the distance d in the reconstruction process or automatically by applying known techniques. An example of electronic focusing is shown on *Figure 4-10*. An efficient technique is based on wavelets or Fresnelets (Liebling and Unser, 2004)

4.3 Introduction of a microscope objective and/or lenses in the reconstruction algorithm

As explained in section 3.5, collecting the highest spatial frequencies of the waves diffracted by the microscopic objects is necessary to restore the object wavefront with high fidelity and obtaining highly resolved images. High N.A. objectives, even preferably oil or solid immersion lenses are needed to reach this goal. Solid Immersion Lens (SIL) may have N.A. much larger than 1, (reaching 2. or even 3, for some SIL based on the use of high refractive materials: GaP). This lens should be placed in close proximity to

the specimen, intercalated between the sample and the recombining beam splitter and the CCD chip. To complete the reconstruction process, the task consists in compensating the phase spread introduced by the M.O. over the plan perpendicular to the optical axis.

The phase spread function, is basically described by a quadratic dependence on the distance to the optical axis. This is a so-called paraboloidal wavefront centered on the axis. It is a good approximation of the phase spread of spherical beams, such as the beam focused by a lens. The role of a lens or microscope objective in the modification of the complex shape of the beam or so called phase spread function, can be described by a ray transfer matrix, or more simply in the case of a thin, almost perfect lens, by a parameter D related to $\Phi(m\Delta\xi, n\Delta\eta)$, called the “digital phase mask”: a complex function correcting for the phase spread function or complex shape change introduced by the microscope objective. From the development of section 3.5, we have established that

$$\Phi(m\Delta\xi, n\Delta\eta) = C^{-1} \exp \left\{ -\frac{i\pi}{\lambda D} \left((m\Delta\xi)^2 + (n\Delta\eta)^2 \right) \right\} \quad (4.40)$$

can be taken as a phase correcting factor or phase mask for the M.O. induced phase modification. More complex expressions of the digital phase mask $\Phi(m\Delta\xi, n\Delta\eta)$ can be considered to correct complex lens aberrations.

The expression of the reconstructed waveform becomes finally:

$$\begin{aligned} O_o(m\Delta\xi, n\Delta\eta) &= \\ &= -i \exp(2i\pi d/\lambda) \cdot \Phi(m\Delta\xi, n\Delta\eta) \cdot \\ &\quad \cdot DF_{Fresnel} T \left[I_{H \text{ filtered}}(k\Delta x, l\Delta y) R_D(k\Delta x, l\Delta y) / |R_D|^2 \right] = \\ &= -i \exp(2i\pi d/\lambda) \cdot \Gamma(m\Delta\xi, n\Delta\eta) \cdot R_D(k\Delta x, l\Delta y) / |R_D| \\ &\quad \cdot DF_{Fresnel} T \left[I_{H \text{ filtered}}(k\Delta x, l\Delta y) / |R_D| \right] \end{aligned} \quad (4.41)$$

where it is possible to show (Colomb, Cuche, Charrière, Kühn, Aspert, Montfort, Marquet and Depeursinge, 2006) by going in the \mathbf{k} -space that the digital reference wave can be formally pulled out of the argument of the Fresnel transform, giving rise to a quadratic term which can be combined with $\Phi(m\Delta\xi, n\Delta\eta)$ and form another quadratic term $\Gamma(m\Delta\xi, n\Delta\eta)$.

Indeed, it can be shown (Colomb, Cuche, Charrière, Kühn, Aspert, Montfort, Marquet and Depeursinge, 2006) that the approach based on the use of a phase mask $\Gamma(m\Delta\xi, n\Delta\eta)$ can be generalized: $\Gamma(m\Delta\xi, n\Delta\eta)$ can be computed as polynomial expression of m and n . This approach lends itself to the development of an automatic procedure for aberrations compensation in DHM. Reconstruction yields correct and accurate phase distribution, even in the presence of important and high order aberrations.

The reconstructed wavefront is a complex field that can be evaluated in amplitude and phase, therefore delivering two images: one amplitude image and one phase image. The phase image $\varphi(x,y)$, computed according to (4.42), presents an uppermost interest because it really provides true nanometer scale images of microscopic objects. The phase image of a USAF test target is shown as an example on Figure 4.11:

$$\varphi(x, y) = \arctan \left[\frac{\text{Im}(O_o(x, y))}{\text{Re}(O_o(x, y))} \right] \quad (4.42)$$

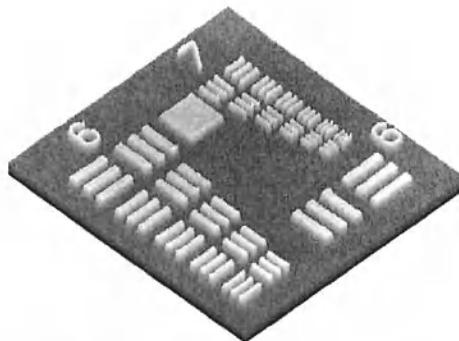


Figure 4.11. Phase image $\varphi(x, y)$ of an USAF test target showing the metallic chromium layers deposited on the glass substrate. The thickness, around 50 nm, can be accurately determined.

5. INSTRUMENTS

A large variety of instruments have been proposed to take holograms of microscopic objects and reconstruct digitally the corresponding wavefronts. The optical systems proposed to image such small fields have been often lensless systems. The reason of that choice is most certainly that no lens or microscope objective is needed to form images of microscopic objects. The

inconvenience of spectrally limited optical materials and the defects and various aberrations of M.O. are therefore avoided in theory. In practice however, other inconveniences are generated by the lensless combination of object and reference beams

5.1 Lensless configurations

Strongly diverging wavefields such as spherical waves emitted by punctual object and reference sources have been proposed to obtain high magnifications. Holograms of microscopic objects are directly formed and collected on a CCD camera. Separate object and reference waves were used by Haddad (Haddad et al., 1992), Boyer et al. (Boyer et al., 1996) and also by Takaki et al. (Takaki et al., 1999): a point source situated in the plane of the microscopic preparation and in its immediate neighborhood provides a spherical wave which interferes with the light diffracted by the object. The hologram therefore results from the interference of “*grosso modo*” two spherical waves, one emanating from a pinhole and used as a reference wave and the other directly diffracted by the object. In order to keep the sampling capacity of the CCD camera compatible with the spatial frequencies content of the hologram, the pinhole of the reference wave must be very close to the specimen and/or the camera quite distant from the specimen. This causes practical difficulties: the preparation must include the reference point source. Takaki et al. (Takaki and Ohzu, 1999) have proposed to realize a two beam recording system by inserting a beam recombination cube between the preparation and CCD chip: one beam points slightly outside the preparation and the other illuminates the preparation at short distance in the same plane. The optical design is presented on *Figure 4-12*.

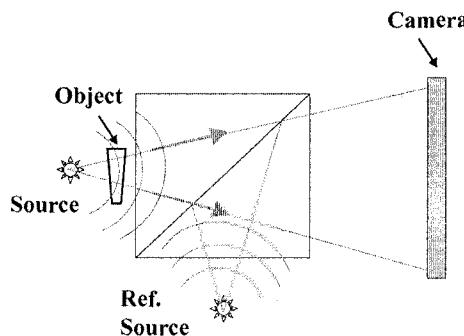


Figure 4-12. Lensless holographic microscope

A similar configuration for digital holographic microscopy have been proposed (Kebbel et al., 2001). The CCD collects “Fourier” holograms,

which ultimately yield an object image by Fourier inverse transformation. Corrections to the data collected on the CCD are often mandatory because of the strong divergence of the object and reference beams.

Another approach (Xu et al., 2001) has been to combine the illuminating and the reference beam in a single beam. This so-called “in-line” geometry eliminates the recourse to an additional reference wave. This arrangement, in many aspects, matches the primary idea of Gabor (Gabor, 1948). The reconstruction process is based on the inversion of Kirchhoff-Helmoltz transform in a on-axis geometry. Though very simple in its principle and practical realisation, the experimental device gives rise, in our opinion, to severe difficulties, both on the theoretical plan and practical implementation. Firstly, the application of the method assumes that the illumination beam is only slightly attenuated or perturbed by the object or preparation so that the interference of the reference beam and the beam scattered by the object remains large enough and unaberrated. This is not the case of relatively thick preparations such as dense biological samples. The basic hypothesis validating the use of in-line holography to observe sparse objects, bumps into the difficulties caused by multiple scattering and heterogeneity of the diffractive index of the specimen and the recording of a hologram becomes problematic. More precisely, the validity of the first Born approximation may be very rapidly invalidated in practice for biological objects of more than several tens of micron in thickness (Marquet, 2003). In particular, the hypothesis that the reconstruction of the wavefront can be performed from a single hologram at various depths (variable reconstruction distance d), providing the 3D arrangement of the object, is valid only at the limit of objects composed of a small number of isolated scattering centres: sparse preparation (as it was shown in more recent papers of Xu and Al. (Xu et al., 2002; Xu et al., 2003)). This hypothesis is invalidated in the majority of biological applications involving tissues thicker than a few tens of microns. In practice, the amplitude of the reference wave and the object wave significantly deviate from the expectations from the first Born approximation. This may lead to a wrong evaluation of the cross term linking the object and reference wave, used for wavefront reconstruction.

Another major drawback met for all the “lensless” configurations proposed for digital holographic microscopy, is that the effective N.A. is limited in practice far below 1, which keeps, to our knowledge, the values of the resolution much larger than one micron, a poor figure in contemporary microscopy. This is inevitable for the systems using a beam recombination cube. For the other optical set-ups, the main reason causing these restrictions is certainly that the angle of acceptance of the CCD camera is kept to relatively low values, because of the presence of an air gap between the CCD chip and the sample. When the sample is imbedded in a liquid or solid

dielectric medium as in the case of biological samples, the existence of the critical angle of refraction for the outgoing wave diffracted by the sample constitutes a natural mechanism limiting the acceptance angle of the camera chip. Moreover the sensitivity of the camera for grazing incidence is falling fast.

5.2 Configuration with lenses or Microscope Objectives

As shown in section 3.5, the recourse to a lens or a M.O. facilitates much the sampling of the hologram. For that reason, it was proposed by Cuche, (Cuche, Marquet and Depeursinge, 1999) to incorporate a M.O. in the optical setup in order to collect all the wavevector components of the beam diffracted by the object. It can be noted that M.O. have also been incorporated in setups for “in-line” holography in order to generate a punctual light source, with or without additional pinhole, in the neighbourhood of the observed specimen (Xu, Jericho, Meinertzhangen and Kreuzer, 2001).

De Nicola, P.Ferraro et al. have further developed similar experimental arrangements: (De Nicola et al., 2002), (De Nicola et al., 2003), (Ferraro et al., 2003), (Coppola et al., 2004), (Ferraro, Coppola et al., 2004), (Ferraro, De Nicola, Finizio et al., 2004), (Ferraro, De Nicola, Coppola et al., 2004), (Ferraro et al., 2005). Application to imaging of transparent microscopic objects is described also by (Palacios et al., 2005).

In the field of biology, Kemper et al. have obtained images of hepatocytes with a transmission DHM adapted from a commercial microscope (Carl et al., 2004).

Similarly MK Kim and coworkers, in recent years, have developed optical setups for microscopy imaging by digital holography, which incorporates M.O. (Dakoff et al., 2003), (Yu et al., 2005), (Yu et al., 2005). By scanning the wavelength according to a technique previously developed (Kim, 1999), (Kim, 2000), these authors have been able to achieve sectioning at a microscopic scale. Application of that sectioning technique with resolution less than one micron (750nm) has been given recently (Montfort, 2005).

Frank Dubois has proposed a Mach-Zehnder configuration incorporating two matched M.O. (Dubois, Joannes and Legros, 1999; Dubois et al., 2004), This configuration allows for matching the object and reference wavefronts and relaxes more severe requirements on spatial coherence of the source. Applications have been developed in observation of microcrystal growth or cells in microgravity conditions(Dubois, Joannes, Dupont et al., 1999).

In parallel works, Ting-Chung Poon, Guy Indebetouw, and others have developed for many years interesting approaches based either on scanning

and non scanning techniques. Scanning techniques have been proposed first (Poon et al., 1995), (Schilling et al., 1995), (Poon et al., 1996), (Schilling et al., 1997), (Indebetouw et al., 1998), (Kim et al., 2000), (Poon, 2003; Poon et al., 2003), (Indebetouw et al., 2005), (Poon et al., 2005). A non-scanning technique has also been proposed, which is based on temporal heterodyning and has been called “spatiotemporal” microholography. A similar approach has been recently proposed (Popescu et al., 2004), but without the formation of a hologram. It was called Fourier phase microscopy. Similar approach, without phase shifting, but phase extraction by Hilbert transform was called Hilbert phase microscopy (Ikeda et al., 2005).

Two types of microscopes have been developed by our group for the observation of transparent objects on one side, or opaque, nevertheless light reflecting or diffracting microscopic objects on the other side.

5.2.1 Transmission Microscope

Transmission setup is used to observe transparent or semi-transparent microscopic objects. Typically micro-lenses, MOEMS devices or biological objects like cells or tissue specimens can be imaged with such an instrument. A typical arrangement is shown on *Figure 4-13*. It consists in a punctual source delivering a light beam with appropriate coherence properties to a beamsplitter, which separates the beam into an illuminating beam and a reference beam. A lens focuses the illuminating beam onto the specimen to be observed. Diffracted beam is collected by the M.O. and normally imaged in a focal plane defined by the position of the M.O. with respect to the specimen, its magnification factor and by the optional use of a field lens. This part of the optical arrangement involving the illumination beam and the specimen is not very different from a conventional transmission microscope. A reference beam has been added to the configuration, which is optically completely separated from the imaging beam. This option has the great advantage of providing completely controllable optical levels and polarisation states, even if the specimen is strongly absorbing or scattering. The “cross terms” in the hologram, which carry the useful wavefront data, can therefore be sampled in optimal conditions by the camera and further digitalized for numerical processing. The camera can be optionally a CCD or CMOS camera with enough pixels to provide a 512 x 512 or more often today, a 1024 x 1024 matrix of data representing the sampled hologram. The camera is placed at some distance d of the focussing plane, so that the Fresnel hologram can be sampled in optimal conditions and reconstructed according to the algorithm given in section 4.

The hologram acquisition is performed at video frequency (25 Hz or even more, depending on the performances of the camera) and the reconstruction rate is about 15 images per second. The method requires the adjustment of several reconstruction parameters, which may be performed automatically.

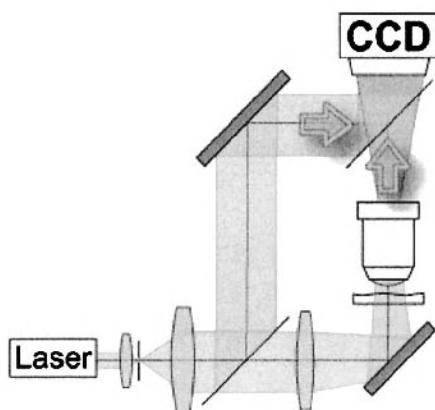


Figure 4-13. DHM in transmission: Optical setup for the capture of an hologram formed with the beam transmitted trough a semi-transparent object.

5.2.2 Reflection Microscope

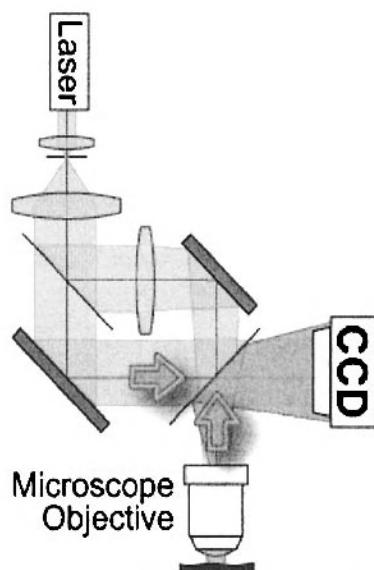


Figure 4-14. DHM in reflection: Optical setup for the capture of an hologram formed with the back-reflected or backscattered light from the specimen.

In an alternative approach, a reflection setup can also be used to observe opaque specimens (see *Figure 4-14*). In that case, the beam reflected on the facets of the object or, more generally, backscattered by the scattering centers in the object gives the optical signal. Similarly, a punctual source delivers a light beam to a beamsplitter, which separates the beam into an illuminating beam and a reference beam. The illuminating beam impinges on the specimen through the M.O., which also collects the reflected or backscattered light from the specimen. An image of the specimen is formed at a distance d in front of, or behind the focused image.

5.2.3 Short coherence holographic microscopy

Short coherence properties of the source can be exploited to “slice” the specimen. The so-called “coherence gating” is a detection modality consisting in filtering out, in the time or spatial domain, the crossed terms $O^*R + OR^*$.

These terms have non-zero values only in the hologram region where the propagation time of the wave from the object to the camera matches the propagation time of the reference wave. In the specimen, this condition delimitates a section, usually a plane, wherefrom the wavefront is emitted. This “gating” therefore can be used to give a full tomographic image of the specimen. Full 3D imaging can be achieved, provided that a sufficient backscattered signal can be obtained from the whole specimen. For turbid media such as biological tissues, such a requirement is met over a few hundreds of micrometers.

The use of short coherence has been proposed several years ago by (Cuche, Poccia and Depeursinge, 1997) and (Indebetouw et al., 2000) and later by (Pedrini et al., 2001), (Pedrini et al., 2002). Application to the visualization of tissue structures, eye cornea in particular, was demonstrated by (Massatsch et al., 2005), (Massatsch, 2003).

An interesting feature of digital holography applied to microscopy is the outstanding flexibility brought by digital data processing. An example of this property is given by the possibility to synthesize by digital data processing ultra-short coherence gate obtained by the superposition of wavefronts reconstructed from holograms taken at different wavelengths. The gate can be made theoretically as short as desired, provided that a spectral range wide enough can be scanned. In practice, slices as thin as 750nm could be obtained by scanning the visible range: 480-700 nm (Montfort, 2005). An advantage over the “white light” approach consisting in illuminating the specimen with very broad spectral range sources, is that exact corrections can be brought to the phase of the reconstructed wavefronts, on the basis of a

reference measurement. In particular, the dispersion laws governing the dependence of refractive indexes on wavelength render these corrections mandatory. Ultra-high precision is at the expense of these precautions.

These developments have been proposed by (Kim, 1999), (Kim, 2000), (Dakoff, Gass and Kim, 2003), (Yu and Kim, 2005), (Yu and Kim, 2005) and recently by (Montfort, 2005), with an outstanding 750nm resolution in depth.

6. APPLICATIONS

6.1 Material Sciences

6.1.1 Fresnel reflection

DHM provides extremely precise data concerning both the phase and amplitude of the reflected or refracted light beam at the surface or interface of dielectric, semiconductor or metallic materials. These properties can be used to identify the dielectric properties of material involved in the realisation of homogeneous or heterogeneous structures. Metallic or semiconductor reflection data (see *Figure 4-15*) are characterized by a phase shift that can be precisely measured and can be very helpful in the determination of electronic properties of these metallic or semiconductor compounds. The behaviour of photons in photonic crystals, polaritons in semiconductors and electrons in metals (plasmons) and probable future meta-materials are hot topics for DHM investigations. The propagation of coherent radiation in transmission and reflection through multi-dielectric layers or on stratified media is also a highly motivating topic for research laboratories and industry.

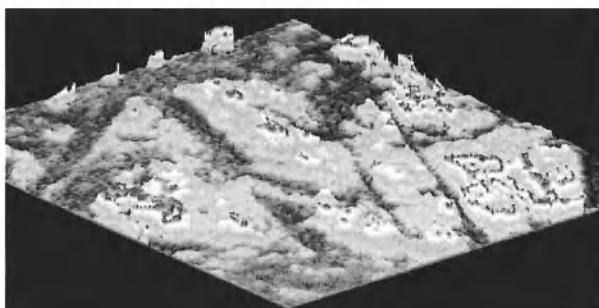


Figure 4-15. Phase image of the surface of glass material covered with heterogeneous ZnO deposit.

6.1.2 Birefringence properties

Finally, birefringence of material and large molecules can be visualized from the capture of a single hologram by using simultaneously two reference beams (Colomb et al., 2002), (Colomb et al., 2004), (Colomb, 2005t). This invaluable property of holography makes possible the detailed study of stresses and strains inside dielectric materials like fiber optics guides (Colomb et al., 2005) and laser processed Si, SiO₂, glass and other substrates

6.2 Surface sciences

Roughness measurements, quality control of polished surfaces down to the nanometer scale, detection of surface defects such as lines, marks and scratches, are examples of applications of DHM to surface investigations. Precise profilometry of surfaces plays an important role in many industrial applications. The advantages of DHM technology are:

- Non-touch
- Fast data acquisition, non-scanning technique
- Robust

These advantages make DHM attractive for the optimization of industrial processes and “on-line” quality control. Examples of surface quality evaluation are presented on *Figure 4-16* (mirror polish) and *Figure 4-17* (machined surface) with linear profiles and computed RMS deviations and peak values.

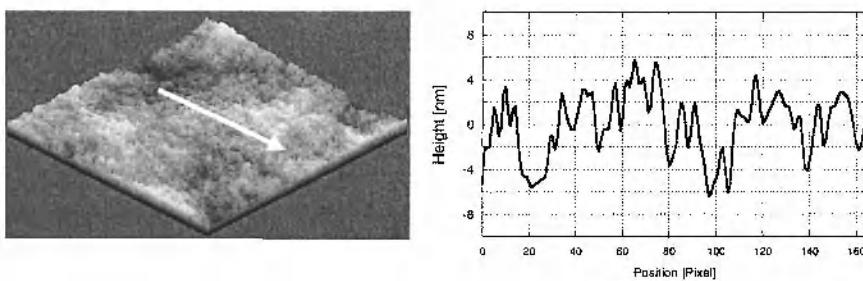


Figure 4-16. Example of polish quality control: a linear profile (white arrow) is presented in the right part of the figure. Roughness (RMS): 2.465 nm, peak value: 12.180 nm.
(courtesy of Lycée Tec)

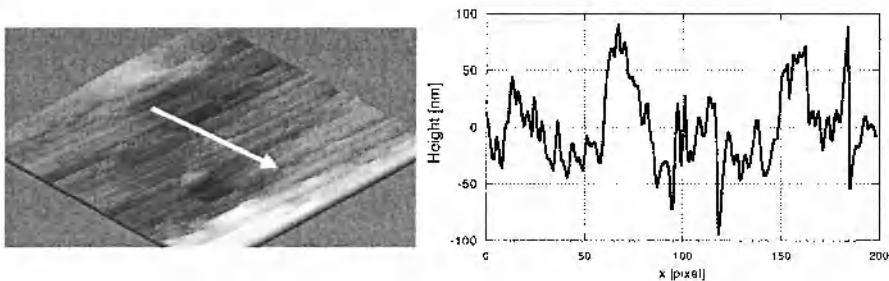


Figure 4-17. Example of surface profilometry showing a texture due to machining.
Roughness (RMS): 26.5 nm, peak value: 189.4 nm. (courtesy of Lycée Tec)

6.3 Micro-optics, MEMS and MOEMS

Mechanical, Optical, and Electronical Microsystems (MOEMS), and, in a similar domain, micro-optics appear as technologies in constant and rapid evolution. They require metrological tools adapted to research and development, as well for quality control. DHM brings some determinant advantages in response to that particular demand: it offers accuracies comparable to interferometers, but with a more practical access to the data. Its robustness makes its implementation in a laboratory environment very easy, practically as simple as any type of optical microscope. The digital reconstruction process involved in DHM makes it a versatile tool to obtain rapidly from a single hologram a wide range of information on microlenses such as surface topography, diffracted wave front, phase function, aberrations, radius of curvature, lens height and surface roughness.

Compared to classical phase shifting interferometry, DHM offers similar performances in terms of resolution, precision, repeatability and field of view, but can be considered as an attractive solution as a result of five main features:

- Higher acquisition rate resulting from the fact that a complete description of the complex wavefront is obtained from a single hologram.
- Reduced sensitivity to external perturbations (vibration and ambient light), since the capture time can be reduced to a few tens of microseconds.
- The accuracy is not intrinsically limited by the precision of the control of moving parts, such as piezo-electric transducers.
- A DHM instrument can be used without adaptations to investigate a wide variety of micro optical component shapes, including cylindrical, square and strong aspheric lenses.

- Ease of use and flexibility for implementations in automated processes for quality control, resulting from the robustness of the technique regarding positioning tolerances.

Examples of phase image of a microlenses array and micro-optical realization are given in *Figure 4-18* (Charrière et al., 2006).

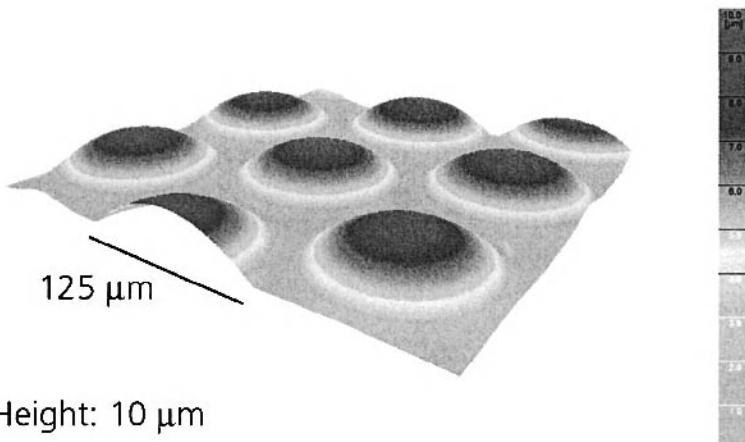


Figure 4-18. 3D presentation of micro-lenses array after phase unwarping.

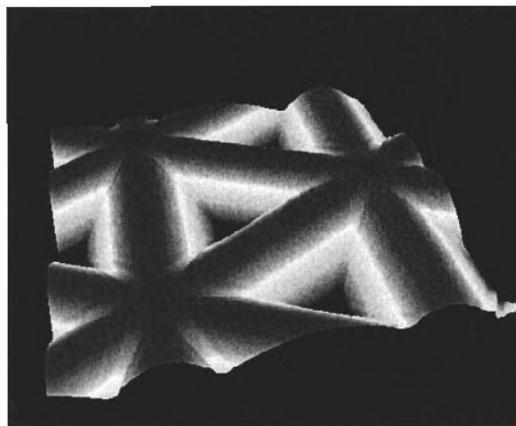


Figure 4-19. Other example of micro-optical realization; "cat's eye", unwarped phase. (Courtesy of Lycée Tec and Dr. Klaus Reimer Fraunhofer-Institut für Siliziumtechnologie, Itzehoe)

6.4 Biology: Cells and tissues

In biology absolute phase-contrast imaging of living cells in culture, for which the reconstructed phase distribution provides a quantitative and precise measurement of the 2D distribution of optical pathlength at the surface of the specimen. Quantitative informations concerning cell morphology and volume, as well as the protein content potentially can be derived from DHM images, provided that adequate processing of data will be chosen. Using high N.A., a sub-micron transverse resolution has been achieved. An accuracy of approximately half a degree has been estimated for phase measurements. In reflection geometry, this corresponds to a vertical resolution less than approximately 1 nanometer at a wavelength of 632 nanometers. In the transmission geometry, the resolution for thickness measurements depends on the refractive index of the specimen and a resolution in the range of a few tens of nanometers has been estimated for living cells in cultures.

Figure 4-20: shows an example of a 3D image of a preparation of neuron cells in culture, obtained by DHM: thickness of sample has been estimated to 7 microns, with an optical pathlength resolution of around 40 nm. The field of view is around $30 \times 30 \mu\text{m}$. For more details see (Marquet et al., 2005)

Living neurons in culture have been imaged with unrivalled resolutions by one of us (Pierre Marquet). Minute deformations and movements of

living cells could be detected and monitored (Rappaz et al., 2005). Both morphology and functional studies can be carried out on living preparations. The method is completely non invasive, because the illumination levels are very low, much less than in fluorescence microscopy. Cultures can be kept alive for several hours and even days. Therefore, long-term evolution of cells can be studied. On the other hand, fast acquisition (at video

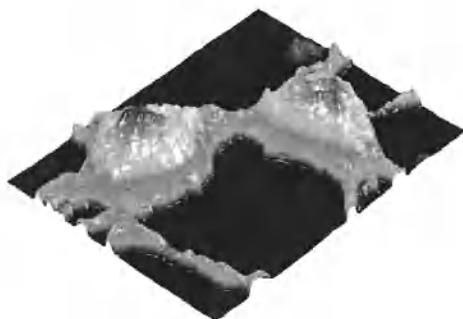


Figure 4-20. DHM phase image of two neurons in culture. The phase shift or optical pathlength is coded in gray scale. Upper scale limit corresponds to an OPL of 10 microns, which corresponds approximately to 7 micrometers effective height. Courtesy of Pierre Marquet (Marquet, 2003).

rate or even faster) is feasible, because the method is non-scanning and the acquisition is made in parallel with a large bandwidth, showing tiny movements of the membranes or cytoplasm.

Image of two neurons in culture, obtained by Digital Holographic Microscopy (DHM) is presented on *Figure 4-20* (Marquet, 2003). The third dimension is presented in coded grey levels and represents the phase shift induced by the variations of the Optical Path Length (OPL) of the beam crossing the cell. Details of the topography and movements of the cell can be observed with DHM in the range of a few tens of nanometers.

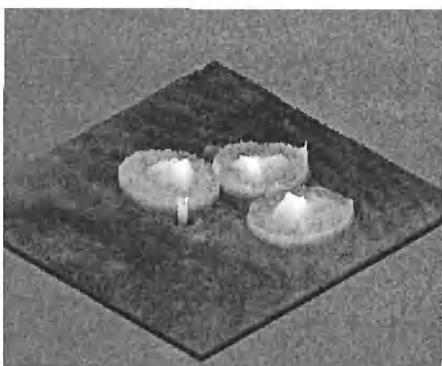


Figure 4-21. DHM phase image of three yew pollen grains. This image clearly shows the cytoplasm and nuclear content of the grain. The abrupt increase of OPL at the frontier of the grain originates from the elevated refracted index in the shell of the grain (Charrière et al., 2005).

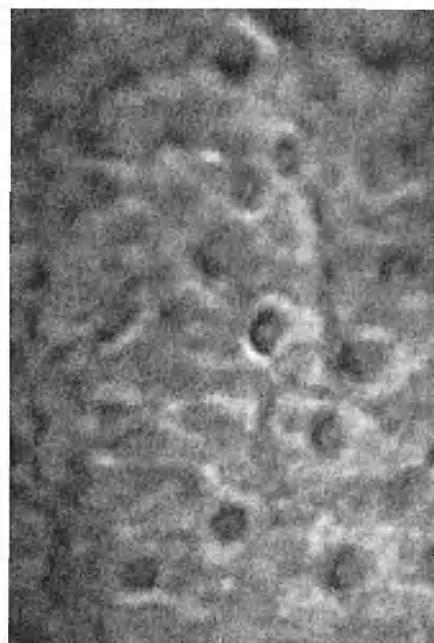


Figure 4-22: Eye epithelium: the DHM amplitude image shows from the beam reflected by the cellular body surrounding the nuclei appearing in dark (Massatsch, 2003)

Absolute phase image of a variety of biological objects can be easily obtained by DHM. *Figure 4-21* shows phase images of pollen grains: many useful details can be observed, which are specific to each variety of plant or tree. Similar, highly contrasted phase images have been obtained for bacteria and other single-celled or multi-cellular organisms.

DHM also provides images of tissues. An example of tissue imaging is given at *Figure 4-22*: a section of porcine eye epithelium is shown. It was obtained in short coherence illumination. The thickness of the imaged tissues depends on the strength of the beam coherently backscattered by the tissue structure, compared to the intensity the incoherently backscattered signal. The very

weak coherently backscattered signal compared to incoherently backscattered signal makes imaging through "turbid media" impossible because of the very disadvantageous ratio of both intensities. The coherently backscattered signal can be as low as 10^{-3} to 10^{-6} of the multiply scattered signal. The contrast is basically so much degraded that the recourse to coherent detection appears as mandatory to eliminate incoherently scattered light and restore an acceptable contrast for imaging purposes. Holography provides a simple access to mutual coherence between the scattered field and the reference beam. Reconstruction of the wavefront from the hologram plays the role of a filter suppressing the background constituted of the multiply scattered photons. Spatial heterodyning of scattered light by the recourse to an off-axis reference beam provides the opportunity to improve tremendously the signal to noise ratio. Digital focussing capabilities provided by numerical reconstruction at various distances offers the possibility to scan the object in depth.

7. CONCLUSIONS AND PERSPECTIVES

The interest of Digital Holography applied to Microscopy can now be considered as established in its theoretical fundaments and in a variety of practical applications: material and surface sciences, micro- and nanotechnologies, Microsystems: MEMS and MOEMS Polarisation microscopy has also been made effective by use of DHM and birefringence measurements from a single hologram.

For industrial needs: metrological and quality control, DHM is performing almost ideally in many conditions: It can be rapidly adapted to changing requirements because of its flexibility provided by digital processing: digital wavefront processing, digital focussing. It provides absolute and calibrated data, rendering possible the comparison with a gauge, archived or accessible through the net. DHM lends itself to quality control and development tasks.

Several biological applications have been already proved: cell and tissue observations, visualisation of protein deposit, surface structuring and patterning in view of biocompatibility and biomaterials developments. The quantitative study of cytoarchitecture of cells and more complex arrangement of cells is essential in the understanding of biophysical mechanisms. In particular, the genesis of pathology: cancer, neurodegeneration, indirect effects of inflammation, oedema, viral infections...can gain much of cell morphology studies. Cell dynamics is a preferential domain where DHM technology will bring new input: fast imaging capabilities, non-scanning and non-touch features of DHM appear as

determinant advantages in the context of biology. Demonstration that minute deformations and movements of living cells can be detected and monitored has been already brought. Both morphology and functional studies can be carried out on living preparations. The method is completely non invasive, because the illumination levels are much lower than fluorescence based imaging techniques, including multi-photon microscopy. Cultures can be kept alive for several hours and days.

We think that DHM has a large potential in future developments: in material research, nanophotonics, new material properties: in photonic crystals, plasmonics, meta-materials developments and so on. In biology it will help determinately in discovering and precising biophysical aspects of cell metabolism, growth and death, apoptosis, plasticity, dendrite spines burgeoning, signal trafficking, vesicles migration.

DHM technology is now ready for use at a large scale, in research and industrial laboratories. It does not need particular operating conditions: clean or antivibratory environment. Very fast acquisition rate is possible, making feasible the observation of fast movements, vibrations, oscillation of membranes in micro-systems. The hologram acquisition is normally performed at the video frequency (25 Hz) or but can reach much higher rates (up to 10'000 holograms per second). For repetitive events, time resolution can reach the sub-picosecond level by sampling pulsed laser illuminated holograms. The reconstruction rate is also very fast: more than 15 images per second. DHM offers diverse and important applications in microscopy: surface studies, and biological applications for the determination of cell volumes combined with refractive index.

The increasing number of research groups involved in DHM development is a indicator the success and future potential of this maging technique.

ACKNOWLEDGMENTS

My gratitude is expressed to all the coworkers involved in DHM development: Pia Massatch*, Anca M. Marian*, Frédéric Montfort*, Tristan Colomb*, Florian Charrière*, Jonas Kühn*, Nicolas Aspert**, Mikhail Botkine** Sébastien Bourquin**, François Marquet**, Etienne Cuche**, Yves Emery**, Pierre Marquet***, and Pierre J. Magistretti**** (* Imaging and Applied Optics Institute, Ecole Polytechnique Fédérale de Lausanne, CH-1005 Lausanne, Switzerland, ** Lyncée Tec SA, PSE A, CH-1015 Lausanne, Switzerland, *** Centre de Neurosciences Psychiatriques, Département de psychiatrie DP-CHUV, Site de Cery, 1008 Prilly, Switzerland, and Brain

Mind Institute, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland.

We acknowledge the Swiss National Science Foundation: SNSF grant 21-67068.01, and the Topnano 21 and nano-micro program of the Swiss Initiative for Innovation: Project 5198.2 and 6101.3, in collaboration with Lyncée Tec SA (www.lynceetec.com).

REFERENCES

- Born, M. and Wolf, E., Principle of optics (Cambridge univ. press, 1999).
- Boyer, K., Solem, J. C., Longworth, J. W., Borisov, A. B., and Rhodes, C. K., 1996, Biomedical three-dimensional holographic microimaging at visible, ultraviolet and X-ray wavelengths, *Nature Medicine* **2**(8), p. 939-941.
- Carl, D., Kemper, B., Wernicke, G., and von Bally, G., 2004, Parameter-optimized digital holographic microscope for high-resolution living-cell analysis, *Applied Optics* **43**(36), p. 6536-6544.
- Charrière, F., Cuche, E., Marquet, P., and Depeursinge, C., 2005, Cell refractive index tomography by digital holographic microscopy, *Optics Letters (in press)*,
- Charrière, F., Colomb, T., Montfort, F., Cuche, E., Emery, Y., Marquet, P., and Depeursinge, C., 2006, Characterization of microlenses by digital holographic microscopy, *accepted for publication in Applied Optics*,
- Colomb, T., 2005, Numerical aberrations compensation and polarization imaging in Digital Holographic Microscopy, Thesis No EPFL, Lausanne.
- Colomb, T., Cuche, E., Montfort, F., Marquet, P., and Depeursinge, C., 2004, Jones vector imaging by use of digital holography: simulation and experimentation, *Optics Communications* **231**(1-6), p. 137-147.
- Colomb, T., Dahlgren, P., Beguin, D., Cuche, E., Marquet, P., and Depeursinge, C., 2002, Polarization imaging by use of digital holography, *Applied Optics* **41**(1), p. 27-37.
- Colomb, T., Dürr, F., Cuche, E., Marquet, P., Limberger, H. G., Salathe, R. P., and Depeursinge, C., 2005, Polarization microscopy by use of digital holography: application to optical-fiber birefringence measurements, *Applied Optics* **44**(21),
- Colomb, T., Cuche, E., Charrière, F., Kühn, J., Aspert, N., Montfort, F., Marquet, P., and Depeursinge, C., 2006, Automatic procedure for aberrations compensation in digital holographic microscopy and applications to specimen shape compensation, *accepted in Applied Optics*,
- Coppola, G., Ferraro, P., Iodice, M., De Nicola, S., Finizio, A., and Grilli, S., 2004, A digital holographic microscope for complete characterization of microelectromechanical systems, *Measurement Science & Technology* **15**(3), p. 529-539.
- Coquoz, O., 1994, Endoscopic Holography with a Multicore Optical Fiber Applied to Biomedical Imaging, Thesis No 1277 EPFL, Lausanne.

- Coquoz, O., Conde, R., Taleblou, F., and Depeursinge, C., 1995, Performances of Endoscopic Holography with a Multicore Optical- Fiber, *Applied Optics* **34**(31), p. 7186-7193.
- Cuche, E., 2000, Numerical Reconstruction of Digital Holograms: Application to Phase contrast Imaging and Microscopy, Thesis No no 2182 EPFL, Lausanne.
- Cuche, E., Poscio, P., and Depeursinge, C., 1997, Optical tomography by means of a numerical low-coherence holographic technique, *Journal of Optics-Nouvelle Revue D Optique* **28**(6), p. 260-264.
- Cuche, E., Bevilacqua, F., and Depeursinge, C., 1999, Digital holography for quantitative phase-contrast imaging, *Optics Letters* **24**(5), p. 291-293.
- Cuche, E., Marquet, P., and Depeursinge, C., 1999, Simultaneous amplitude-contrast and quantitative phase-contrast microscopy by numerical reconstruction of Fresnel off-axis holograms, *Applied Optics* **38**(34), p. 6994-7001.
- Cuche, E., Marquet, P., and Depeursinge, C., 2000, Aperture apodization using cubic spline interpolation: application in digital holographic microscopy, *Optics Communications* **182**(1-3), p. 59-69.
- Dakoff, A., Gass, J., and Kim, M. K., 2003, Microscopic three-dimensional imaging by digital interference holography, *Journal of Electronic Imaging* **12**(4), p. 643-647.
- De Haller, E. B., von Bally, G., and Depeursinge, C., 1995, High-resolution holography and biopsy: preliminary results, *Bioimaging* **3**, p. 76-87.
- De Nicola, S., Ferraro, P., Finizio, A., and Pierattini, G., 2002, Wave front reconstruction of Fresnel off-axis holograms with compensation of aberrations by means of phase-shifting digital holography, *Optics and Lasers in Engineering* **37**(4), p. 331-340.
- De Nicola, S., Ferraro, P., Finizio, A., Grilli, S., and Pierattini, G., 2003, Experimental demonstration of the longitudinal image shift in digital holography, *Optical Engineering* **42**(6), p. 1625-1630.
- Dubois, F., Joannes, L., and Legros, J. C., 1999, Improved three-dimensional imaging with a digital holography microscope with a source of partial spatial coherence, *Applied Optics* **38**(34), p. 7085-7094.
- Dubois, F., Joannes, L., Dupont, O., Dewandel, J. L., and Legros, J. C., 1999, An integrated optical set-up for fluid-physics experiments under microgravity conditions, *Measurement Science & Technology* **10**(10), p. 934-945.
- Dubois, F., Requena, M. L. N., Minetti, C., Monnom, O., and Istasse, E., 2004, Partial spatial coherence effects in digital holographic microscopy with a laser source, *Applied Optics* **43**(5), p. 1131-1139.
- Ferraro, P., Grilli, S., and Alfieri, D., 2005, Extended focused image in microscopy by digital holography, *Optics Express* **13**(18), p. 6738-6749.
- Ferraro, P., Coppola, G., De Nicola, S., Finizio, A., and Pierattini, G., 2003, Digital holographic microscope with automatic focus tracking by detection sample displacement in real time, *Optics Letters* **28**(14), p. 1257-1259.

- Ferraro, P., De Nicola, S., Finizio, A., Pierattini, G., and Coppola, G., 2004, Recovering image resolution in reconstructing digital off-axis holograms by Fresnel-transform method, *Applied Physics Letters* **85**(14), p. 2709-2711.
- Ferraro, P., Coppola, G., Alfieri, D., De Nicola, S., Finizio, A., and Pierattini, G., 2004, Controlling images parameters in the reconstruction process of digital holograms, *Ieee Journal of Selected Topics in Quantum Electronics* **10**(4), p. 829-839.
- Ferraro, P., De Nicola, S., Coppola, G., Finizio, A., Alfieri, D., and Pierattini, G., 2004, Controlling image size as a function of distance and wavelength in Fresnel-transform reconstruction of digital holograms, *Optics Letters* **29**(8), p. 854-856.
- Gabor, D., 1948, A new microscopic principle, *Nature* **161**, p. 777-778.
- Goodman, J., Introduction to Fourier Optics (McGraw-Hill, 2005).
- Goodman, J. and Lawrence, R. W., 1967, Digital image formation from electronically detected holograms, *Appl.Phys.Lett.* **11**, p. 77-79.
- Gross, M., Goy, R., Forget, B. C., Atlan, M., Ramaz, E., Boccara, A. C., and Dunn, A. K., 2005, Heterodyne detection of multiply scattered monochromatic light with a multipixel detector, *Optics Letters* **30**(11), p. 1357-1359.
- Haddad, W. S., Cullen, D., Solem, J. C., Longworth, J. W., McPherson, A., Boyer, K., and Rhodes, C. K., 1992, Fourier-Transform Holographic Microscope, *Applied Optics* **31**(24), p. 4973-4978.
- Ikeda, T., Popescu, G., Dasari, R. R., and Feld, M. S., 2005, Hilbert phase microscopy for investigating fast dynamics in transparent systems, *Optics Letters* **30**(10), p. 1165-1167.
- Indebetouw, G. and Klysubun, P., 1999, Space-time holography: A three-dimensional microscopic imaging scheme with an arbitrary degree of spatial coherence, *Appl.Phys.Lett.* **75**, p. 2017-2019.
- Indebetouw, G. and Klysubun, P., 1999, Space-time digital holography: A three-dimensional microscopic imaging scheme with an arbitrary degree of spatial coherence, *Applied Physics Letters* **75**(14), p. 2017-2019.
- Indebetouw, G. and Klysubun, P., 2000, Imaging through scattering media with depth resolution by use of low-coherence gating in spatiotemporal digital holography, *Optics Letters* **25**(4), p. 212-214.
- Indebetouw, G. and Klysubun, P., 2001, Spatiotemporal digital microholography, *Journal of the Optical Society of America a-Optics Image Science and Vision* **18**(2), p. 319-325.
- Indebetouw, G., El Maghnouji, A., and Foster, R., 2005, Scanning holographic microscopy with transverse resolution exceeding the Rayleigh limit and extended depth of focus, *Journal of the Optical Society of America a-Optics Image Science and Vision* **22**(5), p. 892-898.
- Indebetouw, G., Kim, T., Poon, T. -C., and Schilling, B. W., 1998, Three-dimensional location of fluorescent inhomogeneities in turbid media by scanning heterodyne holography, *Optics Letters* **23**(2), p. 135-137.
- Joannes, L., Dubois, F., and Legros, J. C., 2003, Phase-shifting: schlieren high-resolution quantitative schlieren that uses the phase-shifting technique principle, *Applied Optics* **42**(25), p. 5046-5053.

- Kebbel, V., Hartmann, H. J., and Juptner, W., 2001, Application of digital holographic microscopy for inspection of micro-optical components, *Proc.SPIE* **4398**, p. 189-198.
- Kim, M. K., 1999, Wavelength-scanning digital interference holography for optical section imaging, *Optics Letters* **24**(23), p. 1693-1695.
- Kim, M. K., 2000, Tomographic three-dimensional imaging of a biological specimen using wavelength-scanning digital interference holography, *Optics Express* **7**(9), p. 305-310.
- Kim, T. and Poon, T. -C., 2000, Three-dimensional matching by use of phase-only holographic information and the Wigner distribution, *Journal of the Optical Society of America a-Optics Image Science and Vision* **17**(12), p. 2520-2528.
- Kronrod, M. A., Merzlyakov, N. S., and Yaroslavskii, L. P., 1972, Reconstruction of an hologram with a computer, *Sov.Phys.Tech.Phys.* **17**, p. 333-334.
- Le Clerc, F., Collot, L., and Gross, M., 2000, Numerical heterodyne holography with two-dimensional photodetector arrays, *Optics Letters* **25**(10), p. 716-718.
- Liebling, M., 2004, On Fresnelets, Interference, fringes and digital holography, Thesis No 2977 EPFL, Lausanne.
- Liebling, M. and Unser, M., 2004, Autofocus for digital Fresnel holograms by use of a Fresnelet-sparsity criterion, *Journal of the Optical Society of America a-Optics Image Science and Vision* **21**(12), p. 2424-2430.
- Liebling, M., Blu, T., and Unser, M., 2003, Fresnelets: New multiresolution wavelet bases for digital holography, *Ieee Transactions on Image Processing* **12**(1), p. 29-43.
- Liebling, M., Blu, T., and Unser, M., 2004, Complex-wave retrieval from a single off-axis hologram, *Journal of the Optical Society of America a-Optics Image Science and Vision* **21**(3), p. 367-377.
- Marian, A., 2005, Measurement and Interpretation of the 3D Amplitude Point Spread Function of Lenses and Microscope objective, Thesis No 3288 EPFL, Lausanne.
- Marquet, P., 2003, Développement d'une nouvelle technique de microscopie holographique digitale. Perspectives pour l'étude de la plasticité neuronale., Thesis UNIL Lausanne,
- Marquet, P., Rappaz, B., Magistretti, P. J., Cuche, E., Emery, Y., Colomb, T., and Depeursinge, C., 2005, Digital holographic microscopy: a noninvasive contrast imaging technique allowing quantitative visualization of living cells with subwavelength axial accuracy, *Optics Letters* **30**(5), p. 468-470.
- Massatsch, P., 2003, Low coherence Digital Holographic Microscopy: Application to vision in turbid media and Biomedical Imaging, Thesis No 2797 EPFL, Lausanne.
- Massatsch, P., Charrière, F., Cuche, E., Marquet, P., and Depeursinge, C. D., 2005, Time-domain optical coherence tomography with digital holographic microscopy, *Applied Optics* **44**(10), p. 1806-1812.
- Mills, G. A. and Yamaguchi, I., 2005, Effects of quantization in phase-shifting digital holography, *Applied Optics* **44**(7), p. 1216-1225.
- Montfort, F., 2005, Tomography using Multiple Wavelengths in Digital Holographic Microscopy, Thesis No 3345 EPFL, Lausanne.

- Palacios, F., Ricardo, J., Palacios, D., Goncalves, E., Valin, J. L., and De Souza, R., 2005, 3D image reconstruction of transparent microscopic objects using digital holography, *Optics Communications* **248**(1-3), p. 41-50.
- Pedrini, G. and Schedin, S., 2001, Short coherence digital holography for 3D microscopy, *Optik* **112**(9), p. 427-432.
- Pedrini, G. and Tiziani, H. J., 2002, Short-coherence digital microscopy by use of a lensless holographic imaging system, *Applied Optics* **41**(22), p. 4489-4496.
- Pluta, M., Advanced Light Microscopy (Elsevier, New York, 1988), Vol. II.
- Poon, T. -C., Three-dimensional image processing and optical scanning holography, in *Advances in Imaging and Electron Physics*, Vol 126 (2003), pp. 329-350.
- Poon, T. -C. and Indebetouw, G., 2003, Three-dimensional point spread functions of an optical heterodyne scanning image processor, *Applied Optics* **42**(8), p. 1485-1492.
- Poon, T. -C., Wu, M. H., Shinoda, K., and Suzuki, Y., 1996, Optical scanning holography, *Proceedings of the Ieee* **84**(5), p. 753-764.
- Poon, T. -C., Akin, T., Indebetouw, G., and Kim, T., 2005, Horizontal-parallax-only electronic holography, *Optics Express* **13**(7), p. 2427-2432.
- Poon, T. -C., Doh, K. B., Schilling, B. W., Wu, M. H., Shinoda, K., and Suzuki, Y., 1995, 3-Dimensional Microscopy by Optical Scanning Holography, *Optical Engineering* **34**(5), p. 1338-1344.
- Popescu, G., Deflores, L. P., Vaughan, J. C., Badizadegan, K., Iwai, H., Dasari, R. R., and Feld, M. S., 2004, Fourier phase microscopy for investigation of biological structures and dynamics, *Optics Letters* **29**(21), p. 2503-2505.
- Rappaz, B., Marquet, P., Cuche, E., Emery, Y., Depeursinge, C., and Magistretti, P. J., 2005, Dynamic measurement of the intracellular refractive index and cell morphometry of living cells with digital holographic microscopy, *Optics Express (in press)*.
- Schilling, B. W. and Poon, T. -C., 1995, Real-Time Preprocessing of Holographic Information, *Optical Engineering* **34**(11), p. 3174-3180.
- Schilling, B. W., Poon, T. -C., Indebetouw, G., Storrie, B., Shinoda, K., Suzuki, Y., and Wu, M. H., 1997, Three-dimensional holographic fluorescence microscopy, *Optics Letters* **22**(19), p. 1506-1508.
- Schnars, U., 1994, Direct Phase Determination in Hologram Interferometry with Use of Digitally Recorded Holograms, *Journal of the Optical Society of America a-Optics Image Science and Vision* **11**(7), p. 2011-2015.
- Schnars, U. and Juptner, W., 1994, Direct Recording of Holograms by a Ccd Target and Numerical Reconstruction, *Applied Optics* **33**(2), p. 179-181.
- Takaki, Y. and Ohzu, H., 1999, Fast numerical reconstruction technique for high-resolution hybrid holographic microscopy, *Applied Optics* **38**(11), p. 2204-2211.
- Xu, W., Jericho, M. H., Meinertzhangen, I. A., and Kreuzer, H. J., 2002, Digital in-line holography of microspheres, *Applied Optics* **41**(25), p. 5367-5375.
- Xu, W., Jericho, M. H., Kreuzer, H. J., and Meinertzhangen, I. A., 2003, Tracking particles in four dimensions with in-line holographic microscopy, *Optics Letters* **28**(3), p. 164-166.

- Xu, W. B., Jericho, M. H., Meinertzhangen, I. A., and Kreuzer, H. J., 2001, Digital in-line holography for biological applications, *Proceedings of the National Academy of Sciences of the United States of America* **98**(20), p. 11301-11305.
- Yamaguchi, I. and Zhang, T., 1997, Phase-shifting digital holography, *Optics Letters* **22**(16), p. 1268-1270.
- Yamaguchi, I., Matsumura, T., and Kato, J., 2002, Phase-shifting color digital holography, *Optics Letters* **27**(13), p. 1108-1110.
- Yaroslavskii, L. P. and Merzlyakov, N. S., Methods of digital holography (Consultants Bureau, New York, 1980).
- Yu, L. F. and Kim, M. K., 2005, Wavelength-scanning digital interference holography for tomographic three-dimensional imaging by use of the angular spectrum method, *Optics Letters* **30**(16), p. 2092-2094.
- Yu, L. F. and Kim, M. K., 2005, Wavelength scanning digital interference holography for variable tomographic scanning, *Optics Express* **13**(15), p. 5621-5627.

Chapter 5

PHASE-SHIFTING DIGITAL HOLOGRAPHY

Principles and applications

Ichirou Yamaguchi

*Department of Electronic Engineering, Faculty of Engineering, Gunma University, Kiryu,
Gunma 376-8515, Japan*

Abstract: In digital holography holograms are recorded by a CCD and image reconstruction is performed by a computer. It is free from tedious photographic processing and delivers three-dimensional distributions of both amplitude and phase quantitatively. Its main limitation that is caused by much lower resolution of CCDs than photographic materials has been substantially overcome by phase-shifting digital holography that reduces the spatial frequency of hologram by employing the in-line setup and directly evaluates complex amplitude at the CCD plane to eliminate the conjugate and zero-th order images appearing in the off-axis setup which has been commonly used. In this chapter we first describe its basic principle of image formation. It is followed by applications to microscopy, color holography, and data compression for storage, transmission, and real-time display of holographic data. Then its applications to measurements of shape and deformation of diffusely reflecting surfaces are discussed in comparison with conventional holographic interferometry and electronic speckle pattern interferometry.

Key words: Interferometry, holography, image processing, microscopy, color image, data compression, speckle, shape and deformation measurement

1. INTRODUCTION

Recent remarkable and quick advances in digital cameras and computers have been substantially improving practical utilities of digital holography that uses digital recording of holograms and digital reconstruction of 3-dimensional images. Especially, drastic increases of pixel numbers of CCD

and computational power have enabled us to produce the reconstructed images with higher quality in only a few seconds. For guaranteeing the quality of the reconstructed image it is necessary to use as many pixels for the final image as possible. In the off-axis setup that was first used in digital holography¹⁻³ the reconstructed image is accompanied by the zero-order and the conjugate images and hence only one-ninth of the total pixels are utilized for the final image. It is due to direct use of the interference intensity leading to ambiguity of the related phase distribution. This ambiguity can be eliminated by directly deriving the complex amplitude of the object wave by means of phase-shifting interferometry⁴, which has been mainly applied to analysis of phase objects and optical surfaces. However, it can also be used for derivation of complex amplitude when both the real amplitude and phase show variations.

The phase-shifting digital holography was first successfully applied to diffusely reflecting objects for which both the amplitude and phase distributions exhibit random speckle-like variations⁵. The image reconstruction process also becomes simpler and straightforward because no spatial frequency filtering is necessary as in the off-axis method. In this paper we first discuss image formation by phase-shifting digital holography that is then applied to 3-d microscopy, color holography, and data compression. Then we describe measurement of surface shape and deformation of diffusely reflecting surfaces using phase-shifting digital holography in conjunction with the conventional methods such as holographic and speckle interferometry. The latter field can be regarded as an extension of conventional phase-shifting techniques that are applied for holographic and speckle interferometry. The present method requires simpler setups with an aid of a computer not only in phase analysis but also in imaging where phase information is essential. Thus phase-shifting digital holography realizes a new flexibility of quantitative acquisition, processing, transmission, and display of coherent optical information by means of modern digital processing and communication channels and optoelectronic devices.

2. IMAGE FORMATION

2.1 Hologram recording and analysis

The basic setup for phase-shifting digital holography is illustrated in Fig.5-1. A laser beam is divided into two paths. One beam is expanded to illuminate an object. The light scattered from the object is combined at the

CCD with the collimated reference beam after reflection at the PZT mirror controlled by a computer through a beam splitter to ensure the in-line configuration. At least three interference patterns are acquired after stepwise phase shifts of the reference beam. For analysis of image formation we adopt the coordinate system depicted in Fig.5-2. The object wave at the CCD plane is represented as a Fresnel transform of the complex amplitude at the object plane $U_O(x', y')$ by

$$U(x, y) = \iint U_O(x', y') \exp\left[i k z_O + ik \frac{(x - x')^2 + (y - y')^2}{2z_O}\right] dx' dy' \quad (5-1)$$

where integration is carried out over infinity and $z_0 (>0)$ is the distance from the object plane to the CCD. The interference intensity detected by the CCD is given by

$$I_H(x, y; \delta) = |U_R(x, y) \exp(i\delta) + U(x, y)|^2 = |U_R|^2 + |U|^2 + 2 \Re[U_R U^* \exp(i\delta)] \quad (5-2)$$

where $U_R(x, y)$ is the complex amplitude of the reference beam and δ is the mean phase difference between the object and the reference waves. By using the phase-shifting procedure in which the reference phase is shifted by a step of $\pi/2$ at least three times, we can derive the complex amplitude of the object wave such as

$$U(x, y) = \frac{1}{4U_R^*} \{ I_H(x, y; 0) - I_H(x, y; \pi) + i[I_H(x, y; \pi/2) - I_H(x, y; 3\pi/2)] \} \quad (5-3)$$

in the case of four-step algorithm, and

$$U(x, y) = \frac{1-i}{4U_R^*} \{ I_H(x, y; 0) - I_H(x, y; \pi/2) + i[I_H(x, y; \pi/2) - I_H(x, y; \pi)] \} \quad (5-4)$$

in the three-step algorithm. If we compare these expressions with conventional phase analysis using the phase-stepping method, we find it rather unique that we can treat the whole process with complex numbers also in numerical computations.

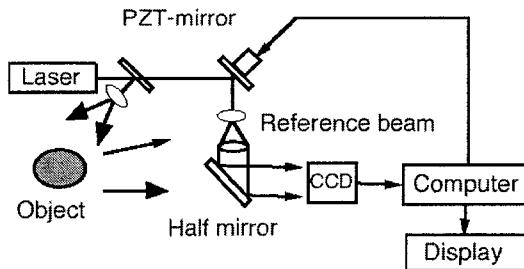


Figure 5-1. Setup of the phase-shifting digital holography

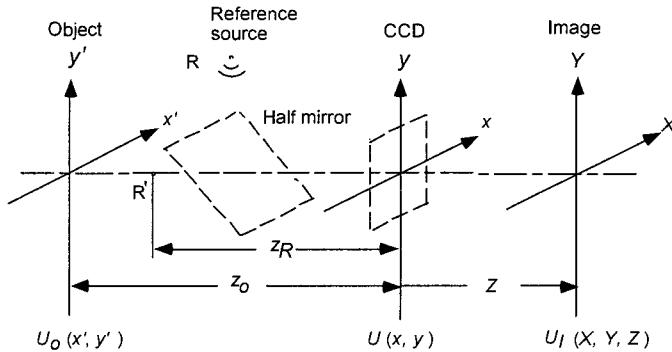


Figure 5-2. Coordinate system for phase-shifting digital holography

2.2 Image reconstruction

The reconstruction is performed by the Fresnel transformation of the derived complex amplitude in such a way that

$$U_I(X, Y, Z) = \iint U(x, y) \exp \left[ikZ + ik \frac{(X-x)^2 + (Y-y)^2}{2Z} \right] dx dy \quad (5-5)$$

where the integration is carried out over the area of CCD. We assume first the sufficient extension and ideal resolution of the device. Here the collimated reference beam is assumed with $z_R = \infty$. If we substitute Eq. (5-1) into Eq. (5-5), we find that the image plane is determined from the condition that the quadratic term of x and y in the exponent vanishes such as $Z=z_o$ where the complex amplitude becomes

$$U_I(X, Y, -z_O) = U_O(X, Y) \quad (5-6)$$

if we neglect limitation of the finite size of CCD array. Image reconstruction using Eq. (5-5) can be numerically accomplished by regarding it as either a Fourier transform or a convolution integral and replacing the integration by summation. In the former algorithm to be called now the single FFT method, the sample interval of the image is given by $\lambda Z/L$ where L is the size of CCD, while the latter, named here the double FFT method because it uses FFT twice, keeps the sampling interval equal to be the pixel pitch of CCD independent of the reconstruction distance. Recently relationships between these algorithms have been clarified together with proposal of a new algorithm using double Fresnel transforms⁶. It is especially useful in color digital holography where reconstructed image at three wavelengths are superposed^{7, 8}. In the case of diffusely reflecting objects only the phase distribution $\arg\{U(x, y)\}$ is enough to reconstruct images of almost the same quality as will be mentioned below in more detail^{9, 10}.

If we record an object by using a geometry shown in Fig.5-1, the resolution is given by

$$\Delta x = \lambda |z_O| / L \quad (5-7)$$

while the focal depth of the reconstructed image is represented by

$$\Delta z = 2\lambda z_o^2 / L^2 \quad (5-8)$$

The maximum object size to be recorded is equal to $N\Delta x$ with the pixel number N along the x-direction in the single FFT algorithm, while it becomes equal to that of CCD in the double FFT method. The effects of CCD parameters on image resolution were investigated numerically¹¹.

2.3 Three-dimensional microscopy

For attaining higher resolution we can insert a microscope objective between the object and the CCD and the defocused image of the object is superposed with the reference wave at the CCD. The reconstruction is carried out by the Fresnel transformation. In this case we can attain the same magnification and resolution as obtained from direct imaging with the objective¹². The second setup for microscopy employs a divergent reference as shown in Fig.5-3. The confocal system where the reference point source and the object are located at nearly the same distance from the CCD is employed to reduce the spatial frequency of the hologram¹¹. In this case we

have to multiply the integrand of Eq.(5-5) by the parabolic phase function expressed by

$$U_M(x, y) = \exp\left(ik \frac{x^2 + y^2}{2z_M}\right) \quad (5-9)$$

with the distance z_M of the point reference from the CCD plane. In conventional imaging by a lens, this parabolic phase function is provided by a lens. The coefficient of the quadratic term of x and y in the exponent of Eq.(5-5) vanishes at

$$Z = Z_I = -1/\left(\frac{1}{z_O} + \frac{1}{z_R}\right) \quad (5-10)$$

The condition that the linear term of Eq.(5-5) vanishes leads to the magnification factor given by

$$m = \left(1 + \frac{z_O}{z_M}\right)^{-1} \quad (5-11)$$

Although this property is well known in conventional holography, it can be used more flexibly in digital holography. Figure 5-4 shows examples of the amplitude and phase images of onion peels both focused and defocused¹¹. The wavelength is 514 nm of an Ar-laser. Instead of a CCD camera we employed a CCD mounted on a card (SONY CCB-M27B, 768x494 pixels, 6.45x4.84 μm^2) so that the cubic beam splitter (5x5x5 mm³) could approach the device as close as possible. Four phase-shifted images were first stored in a frame memory with 512x512x8 bits. The focused intensity image exhibits the same quality as that taken with a microscope. In the phase images the optical thickness distribution of the transparent parts of cells can be found. It has a saw-tooth cross-section to deliver the distribution after phase-unwrapping. The resolution observed from a test target proved to be 3 μm , while the theoretical value was 2.5 μm . Thus by phase-shifting digital holography we can realize a 3-dimensional microscopy applicable for various kinds of objects with a simple setup and without any mechanical movement except for phase-shifting as well as straightforward image reconstruction without frequency filtering needed for off-axis configuration to separate the desired image.

For digital holographic microscopy using complex amplitude information we also can mention scanning holographic microscopy that detects the complex amplitude by optical heterodyne technique where the reference

beam is frequency shifted and an object is scanned by an X-Y stage under focused illumination¹³.

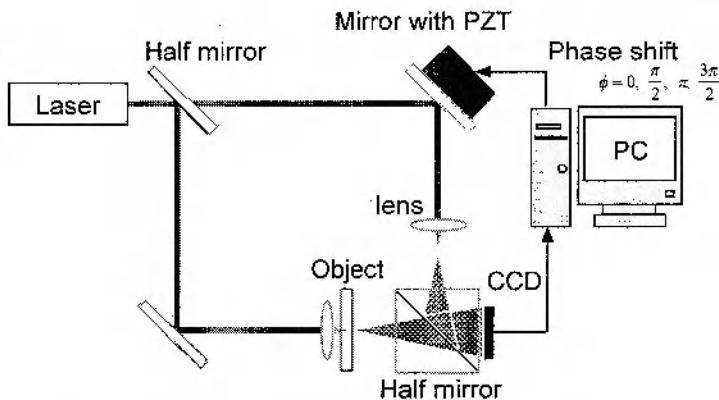


Figure 5-3. Setup for microscopy using phase-shifting digital holography without imaging lens

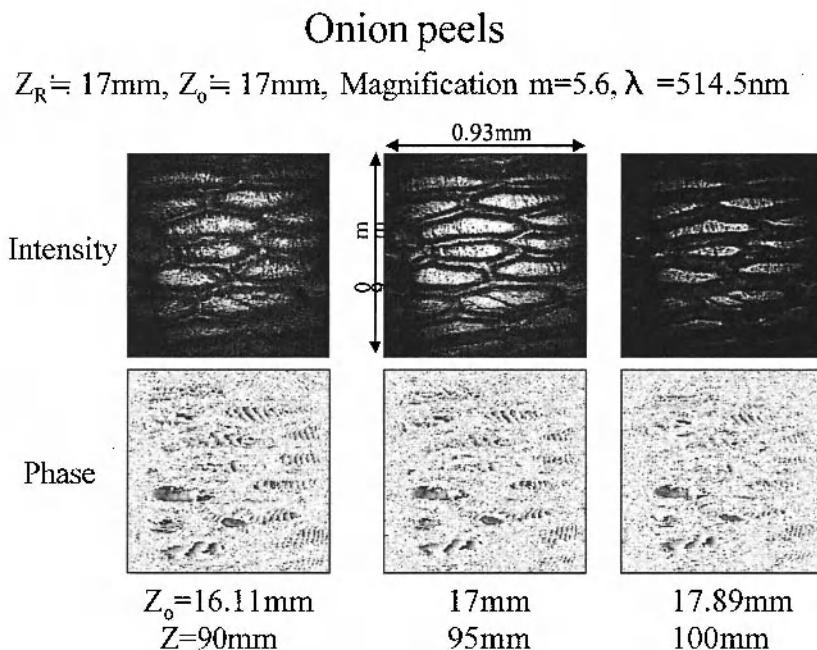


Figure 5-4. Reconstructed images of onion peels at various reconstruction distances

2.4 Color holography

Phase-shifting digital holography was also applied to color holography by using a multi-line laser and a color CCD shown in Fig.5-5⁷. With the phase-shifting of the reference beam in-line holograms for three wavelengths emitted from a He-Cd laser are recorded simultaneously to derive the complex amplitude at each wavelength and then the three monochromatic images corresponding to each wavelength are reconstructed and combined into full-color images in the computer. Laser power variation over the wavelengths can be compensated for in the reconstruction process. We compared the images reconstructed by two algorithms using a single FFT and the double FFT methods with each other by both experiments and numerical simulations. Phase shift was correctly provided for the middle wavelength and errors arising at the other two wavelengths proved not to cause serious deterioration in the reconstructed images. We also employed an achromatic phase shifter consisting of polarization components to eliminate the errors⁸. The optical system for this method, however, becomes more complicated.

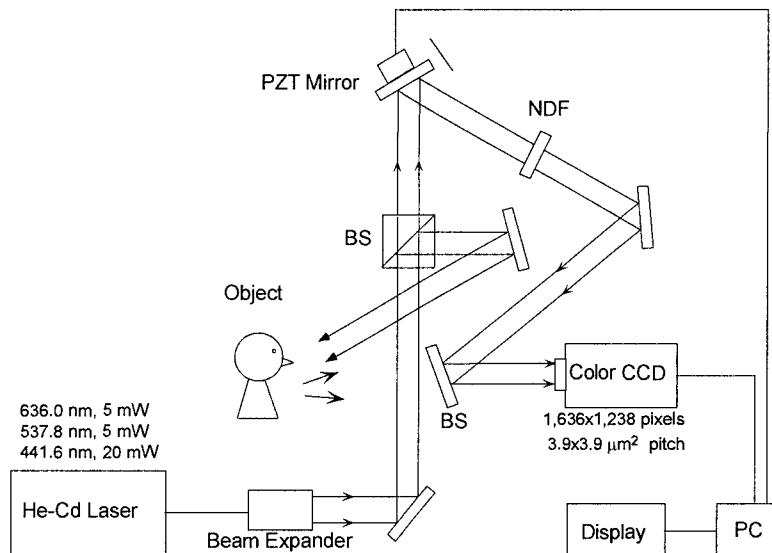


Figure 5-5. Setup for phase-shifting color digital holography

2.5 Data compression

For storage or transmission of enormous amounts of frames comprising the complex amplitude which are necessary for dynamic objects or movies

we need to compress this information for the following purposes. The first is to record holograms at high frame rate and to increase memory capacity. The second is to accelerate data transmission through internets for reconstructing 3-d images at remote places. The third is the real time reconstruction by a spatial light modulator whose dynamic range is generally lower than that of CCD. We have recently studied the effects of reduction of bit-depths in hologram recording by phase-shifting digital holography and found that only four bits, that is, sixteen intensity levels are enough for reconstruction of satisfactory images¹⁴. Here we describe the effects of information reduction of complex amplitude for image reconstruction by maintaining only phase. It has been known that phase of the object spectrum is essential for image formation of a scattering object. This property was first used for kinoforms that realized phase modulation corresponding to the computed Fourier spectrum of a virtual object by bleaching a photographic plate¹⁵. Now for real objects the same data compression can be easily carried out by phase-shifting digital holography. It has been demonstrated by experiments that an image of a diffusely reflecting object can be reconstructed only by phase data of the derived complex amplitude^{9, 10}.

The experimental results are shown in Fig.5-6. As expected we cannot reconstruct any image if we use only the amplitude distribution. It is seen that reduction of bit depth of the phase data does not seriously damage the image even down to 2 bits. We observed enhancement of halo in the image with low bit-depths. This halo results from ignorance of amplitude variation. This tendency was verified quantitatively by one-dimensional simulation where a diffusely reflecting object is represented by random phase variation¹⁶. Smoothing of the images reconstructed from the compressed data has proved to be effective for enhancing image quality. Figure 5-7 represents the dependence of the root-mean-square difference of the reconstructed image intensity on the bit-depth. The caption of phase & amplitude means the quantization of both real and imaginary parts of complex amplitude. The residual difference in the phase-only data can be attributed the halo component. The quantized phase data can also be supplied to a spatial phase modulator, for example, using liquid crystal cells for real time reconstruction of 3-d images¹⁰. If we use the spatial light modulator whose pixel pitch is equal to that of the CCD used for hologram recording, we could reconstruct the 3-dimensional image with unit magnifications both in the lateral and the axial directions and without any geometrical distortion.

Data compression for data transmission and encryption using phase-shifting digital holography were also discussed^{17, 18}. The method used mainly was the reduction of bit-depth of the real and imaginary parts of complex amplitude.

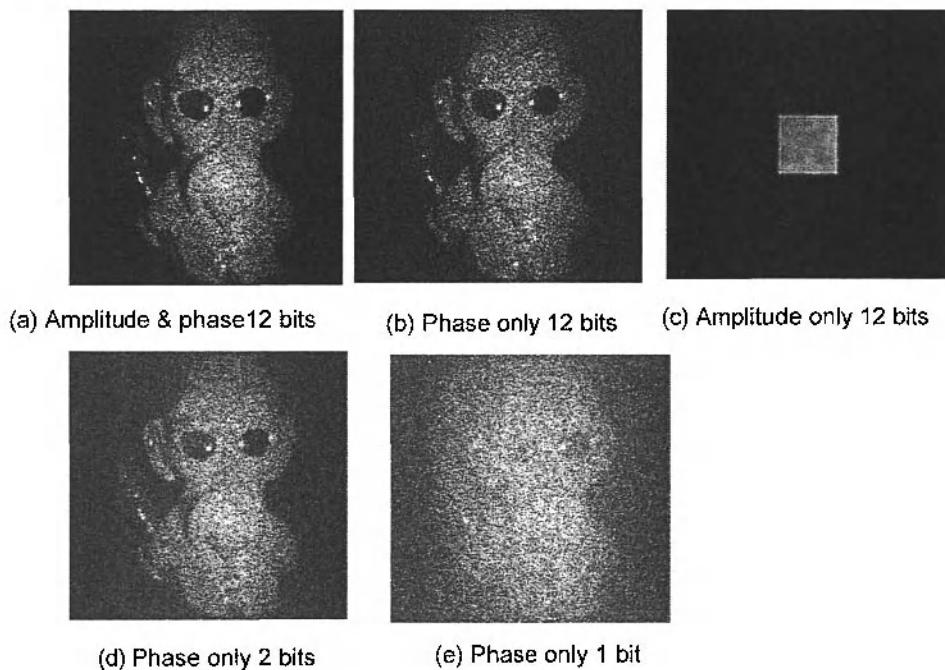


Figure 5-6. Images reconstructed by 12 bit amplitude and phase (a), only 12 bit phase (b), only 12 bit amplitude (c), with 2 bits phase (d), and 1 bit phase (e)

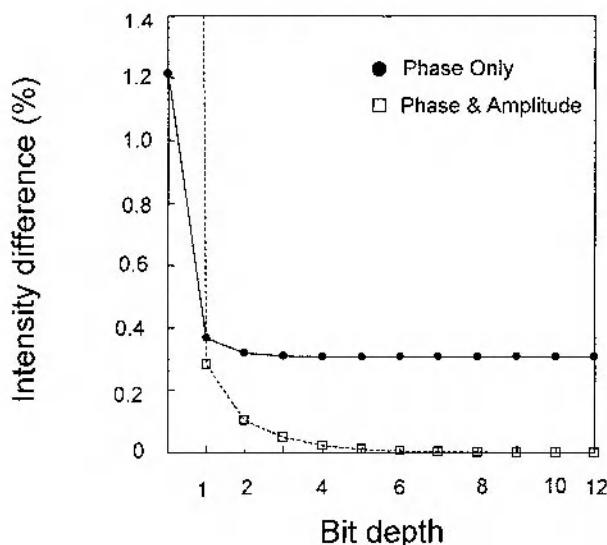


Figure 5-7. RMS image difference between the image reconstructed from full complex amplitude and those from complex amplitude and phase only with reduced bit-depth

3. MEASUREMENT OF SURFACE SHAPE AND DEFORMATION

3.1 General remarks and observed quantities

In holographic interferometry and electronic speckle pattern interferometry (ESPI) we can measure deformation of diffusely reflecting surfaces by superposing the complex amplitudes or the speckle intensities before and after surface deformation. Moreover, if we superpose the complex amplitudes before and after a change of object illumination, namely, that of the incident angle or the wavelength, we can obtain fringe patterns indicating surface contours. If the illumination is fixed and an object is deformed, the coherent superposition of the waves give rise to interference or correlation fringes which mean contour lines of displacement along a direction that depends on the optical system. These fringe patterns can be automatically analyzed either by phase-shifting analysis or Fourier transform method¹⁹, both of which provide phase changes that are wrapped between 0 and 2π . After phase-unwrapping we obtain distributions of unwrapped phase which are proportional to displacement or surface height distributions. In digital holography, on the other hand, the wrapped phase values can be directly calculated from the difference of the reconstructed phases before and after the changes of illumination or object deformation. Hence we only need phase unwrapping to derive displacement or surface height. The optical setup for digital holography is also simpler because we essentially need no imaging lens and thus can analyze three-dimensional objects by means of numerical focusing. We describe below the quantities to be detected in these methods by applying a theory developed previously^{20, 21}.

In holographic interferometry applied to deformation measurement images of the object before and after deformation that is represented by the point displacement distribution $a(r)$ are coherently superposed as shown in Fig.5-8. The effect of the lens aperture used for fringe observation is essential on the shape and contrast of fringe pattern that consists of speckle structure as shown in the figure. The mean size of speckle is given by the wavelength divided by the angular size of the lens aperture. We represent the observation point by the coordinate (R, Z) where R is the transverse coordinate and Z is the distance from the conjugate plane of the object as shown in Fig.5-9. If the complex amplitudes at the point before and after object deformation are denoted by $U_1(R, Z)$ and $U_2(R, Z)$, the intensity averaged over an area much larger than speckle size but smaller than fringe spacing is given by

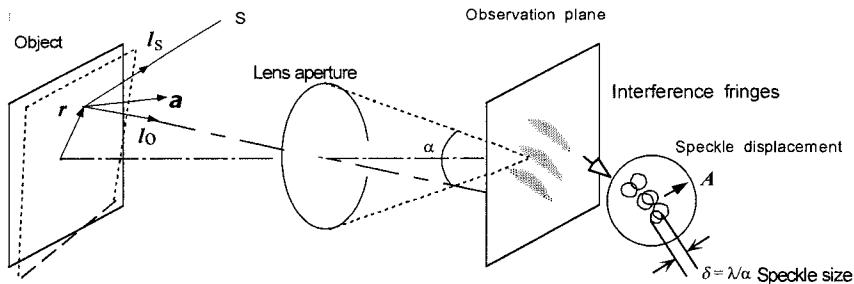


Figure 5-8. Role of speckle size and displacement in holographic interferometry

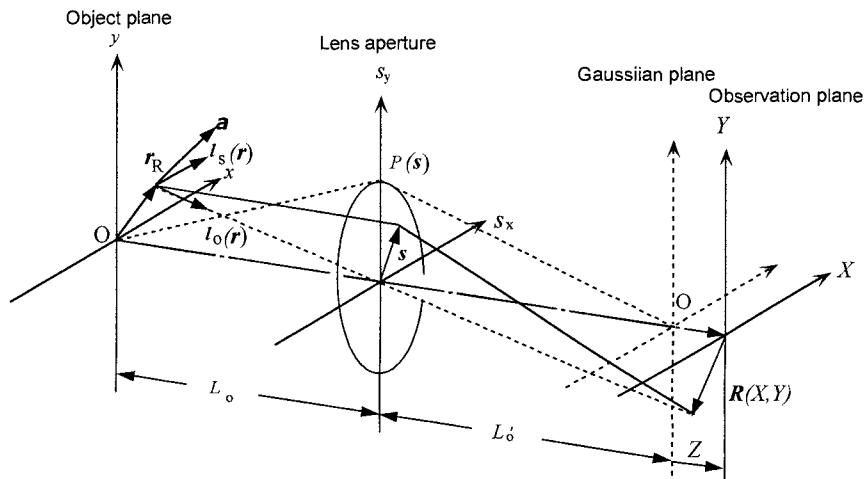


Figure 5-9. Coordinate systems for analysis of patterns observed in holographic interferometry and ESPI

$$\langle I(\mathbf{R}, Z) \rangle = \langle U_1(\mathbf{R}, Z) + U_2(\mathbf{R}, Z) \rangle^2 = \langle I_1(\mathbf{R}, Z) \rangle + \langle I_2(\mathbf{R}, Z) \rangle + 2\Re\langle U_1(\mathbf{R}, Z)U_2^*(\mathbf{R}, Z) \rangle \quad (5-12)$$

where the averaging is mathematically performed over a statistical ensemble of the microscopic structure of the object to discuss shape and contrast of fringe patterns for general object displacement and optical systems. This averaged intensity can be rewritten in terms of the unit vectors directed to the point source and along the direction of observation \$I_s\$ and \$I_o\$ in such a form as

$$\langle I \rangle = 2\langle I_o \rangle \left\{ 1 + \gamma \cos[k(I_s + I_o) \cdot \mathbf{a} + \alpha] \right\} \quad (5-13)$$

where we have assumed $\langle I_0 \rangle = \langle I_1 \rangle = \langle I_2 \rangle$. This equation represents superposition of interference intensity between identical points before and after object deformation. The fringe pattern means the contour lines of the displacement component along the bisector between the direction of the illumination source and that of observation. The fringe pattern has the complex contrast that represents the fringe contrast and phase and is given by

$$\gamma \exp(i\alpha) = \langle U_1(\mathbf{R}; Z)U_2^*(\mathbf{R}; Z) \rangle / \langle I_0(\mathbf{R}; Z) \rangle \quad (5-14)$$

that depends on the ratio of speckle displacement caused by object deformation to the mean speckle size as illustrated schematically in Fig.5-8. If the speckle displacement is smaller than the mean size, the contrast of the fringes is high. On the other hand, if the speckle displacement exceeds the mean size, the contrast vanishes. This is because the phase relationship is random between difference speckles. The speckle displacement depends on object deformation and the position of the observation plane in a complicated manner, where the mean speckle size is proportional to the distance of the observation plane from the lens aperture and inversely proportional to the size of observation aperture. Since speckle displacement does not depend on the aperture size, we can increase the fringe contrast by reducing the aperture. However, if the speckle size approaches the fringe spacing that depends on slope of displacement corresponding to rotation and strain, the above discussion on the fringe contrast based on the average intensity will cease to be valid. Anyway digital holography enables us to analyze phase distributions of scattered light quantitatively and hence this situation might be investigated in more detail.

In electronic speckle pattern interferometry (ESPI) speckle patterns formed by interferometric setups which combine at least one reflected by the object are recorded by a CCD camera, subtracted from the initial pattern, and squared by video circuits or a computer to display contour lines of object displacement. The displayed brightness is proportional to

$$V_S = \langle (I_{S1} - I_{S2})^2 \rangle = \langle I_{S1}^2 \rangle + \langle I_{S2}^2 \rangle - 2\langle I_{S1}I_{S2} \rangle, \quad (5-15)$$

where the third term on the right-hand side represents the correlation fringes.

In an arrangement that is used for measurement of out-of-plane displacement (Fig.5-10) the intensities recorded by the CCD camera are given by

$$I_{S1} = |U_1 + U_R|^2, \quad I_{S2} = |U_2 + U_R|^2 \quad (5-16)$$

with the object amplitudes U_1 and U_2 at the image plane $Z=0$ and the complex amplitude U_R of the reference light reflected from a fixed rough surface. The averaging contained in Eq.(5-15) is physically carried out by the low-pass filtering introduced electronically and/or visually. Substitution of Eq.(5-16) into Eq.(5-15) yields

$$V_S = 2\left(\langle I_o \rangle^2 + 2\langle I_o \rangle \langle I_R \rangle - 2\langle I_R \rangle \Re\langle U_1 U_2^* \rangle - |\langle U_1 U_2^* \rangle|^2\right) \quad (5-17)$$

where we considered the statistical independence between the reference amplitude U_R and the object amplitude U_1, U_2 . The third term in the right-hand side of this equation represents the fringe pattern that is the same as that derived in Eq.(5-12) for holographic interferometry where we set $Z=0$.

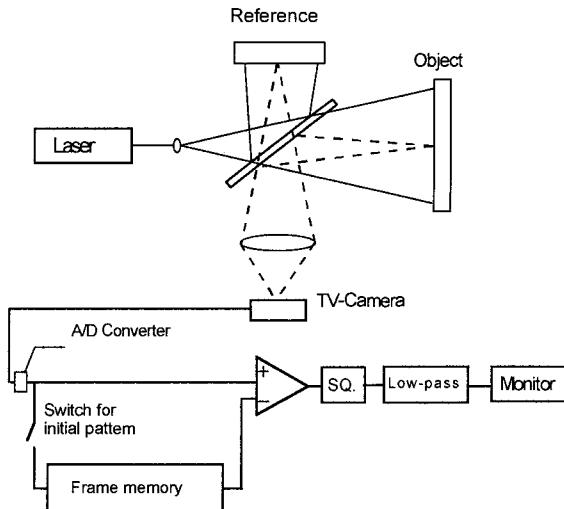


Figure 5-10. Arrangement of ESPI for measurement of in-plane displacement

In digital holography we can directly evaluate values of phases as $\arg(U_1)$ and $\arg(U_2)$ whose difference becomes

$$\Phi = \langle \arg(U_1) - \arg(U_2) \rangle = \frac{\langle U_1 U_2^* \rangle}{\sqrt{\langle I_1 \rangle \langle I_2 \rangle}} = -(\mathbf{k}_s - \mathbf{k}_o) \cdot \mathbf{a} + \alpha, \quad (5-18)$$

where $\mathbf{k}_s = -k\mathbf{l}_s$ and $\mathbf{k}_o = k\mathbf{l}_o$ mean the wave vectors of illumination and reflected light. The phase difference given by Eq.(5-18) cannot be determined when I_1 or I_2 vanish corresponding to dark speckles. Now we also consider the interference between the identical points for studying the sensitivity of the contouring and deformation measurement. In the following we derive the phase change corresponding to Eq.(5-18) in surface contouring using change of object illumination.

3.2 Surface contouring

For surface shape measurement we calculate the phase difference due to change of illumination, that is, changes of the incident angle or wavelength. The wave vectors of illumination beam are represented by \mathbf{k}_a and \mathbf{k}_b as shown in Fig.5-11. If we denote the vectors representing the observation direction by \mathbf{k}_{ao} and \mathbf{k}_{bo} , the difference of the reconstructed phase corresponding to the identical points is given by

$$\Phi(x, y) = -(k_{az} - k_{bz} - k_{aoz} + k_{boz})h(x, y) - (k_{ax} - k_{bx})x \quad (5-19)$$

where we have assumed the incident plane to be included in the x-z plane and the reference plane for the surface height to be the x-y plane. The first term on the right hand side of Eq.(5-19) means the phase difference proportional to the surface height and the second term stands for tilt components.

If we change the wavelength of the illumination from λ_a to λ_b with the incident angle θ_s and normal observation and reconstruct each of the hologram with the same wavelength as in hologram recording, the difference of the phases which are reconstructed with the same wavelength as in the recording is expressed by

$$\Phi(x, y) = -(1 + \cos \theta_s)(k_a - k_b)h(x, y) = -2\pi(1 + \cos \theta_s)h(x, y) / \Lambda, \quad (5-20)$$

which means the contours of object height with a sensitivity that is associated with the synthetic wavelength defined by

$$\Lambda = 1 / |1/\lambda_a - 1/\lambda_b| \quad (5-21)$$

When the incident angle is changed from θ to $\theta + \Delta\theta$, the phase difference becomes

$$\begin{aligned}
 \Phi(x, y) &= -k[\cos\theta - \cos(\theta + \Delta\theta)]h(x, y) - k[\sin\theta - \sin(\theta + \Delta\theta)]x \\
 &= 2kh(x, y)\sin\left(\theta + \frac{\Delta\theta}{2}\right)\sin\frac{\Delta\theta}{2} + 2kx\sin\frac{\Delta\theta}{2}\cos\left(\theta + \frac{\Delta\theta}{2}\right) \\
 &\approx kh(x, y)\Delta\theta\sin\theta + kx\Delta\theta\cos\theta
 \end{aligned} \tag{5-22}$$

where the bottom line results from an approximation for $\Delta\theta$ to be much smaller than unity. This phase distribution is just equal to the phase of the projected fringes produced by the two coexisting beams. For small $\Delta\theta$ the height sensitivity of the phase difference is given by $\lambda/\Delta\theta\sin\theta$.

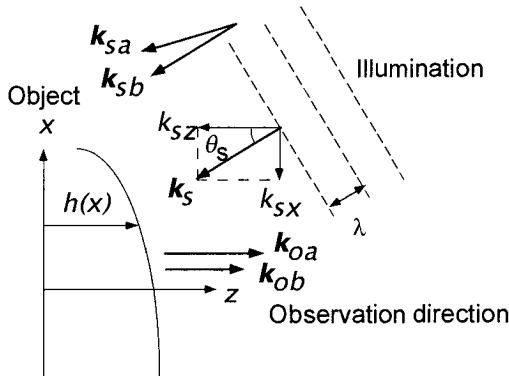


Figure 5-11. Principles of surface contouring by digital holography

An actual setup used in experiments²² is shown in Fig.5-12. The collimated laser beam is incident on a plane mirror that is rotated by a stepping motor under computer control. Three phase-shifted holograms were recorded before and after the rotation of the mirror. The resultant distributions of the reconstructed phase before and after the mirror rotation are subtracted from each other to produce the unwrapped phase. After phase-unwrapping we obtain phase distribution that is given by Eq.(5-22) from which the tilt component has to be subtracted to provide surface height distribution with respect to the reference plane. The presence of the tilt component adds to the fringe density to disturb phase-unwrapping more seriously than in its absence resulting from the dual wavelength method. Any way phase-shifting digital holography improves accuracy of surface contouring and deformation measurement than the off-axis method because of more pixels used for the reconstructed images.

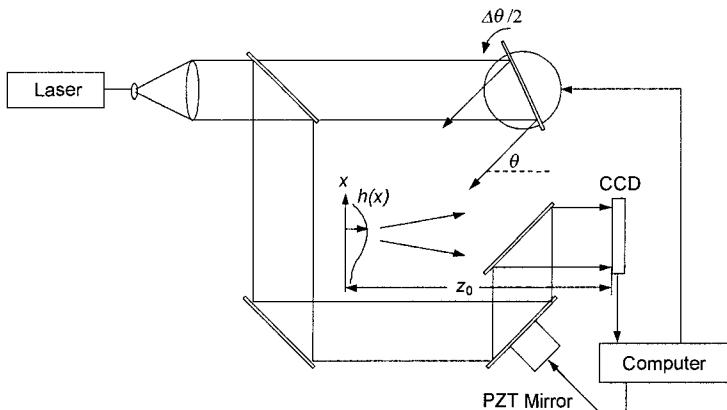


Figure 5-12. Arrangement for surface shape measurement

Figure 5-13 represents the result obtained from the dual incident angle method applied to a miniature bulb which was painted white. The distributions of the phase difference shown in (a) contain the carrier component corresponding to the second term of the right-hand side of Eq.(5-22). We subtract this component from the phase difference at 1,024x1,024 pixels before phase-unwrapping. This procedure aligns the reference plane parallel to the object plane as shown in Fig. 5-13 (b). The resultant distribution contains noise associated with speckles. We suppressed this noise by extracting one point from each 2x2 matrix where the modulus of the product $|U_1 U_2^*|$ becomes maximal. This filtering is based on the fact that the phase value is more reliable for higher amplitude. The compressed data are then smoothed by averaging over each 2x2 matrix with final data pixels of 256x256. Figure 5-14 (a) and (b) represent the 3-d maps and the cross-sections through the bulb axis before (a) and after the averaging (b). This method of surface shape measurement could be easily applied to small objects by employing a microscope system mentioned above.

In the dual wavelength method we can employ the normal incidence that is free from the shadowing effect as well as from correction of the reference plane²³. The wavelength shift is most easily realized by changing the injection current of a laser diode. We can use a mode hop that provides the shift of a several tenths of nanometers. The synthetic wavelength is given by $\Lambda = \lambda^2 / \Delta\lambda$ where we assume that the wavelength shift is much larger than the initial wavelength. This method was also used in conventional holography where reconstruction is performed with the same wavelength to introduce the chromatic aberration into the image recorded at the shifted wavelength. In digital holography is free from this aberration. Phase-shifting digital

holography is especially suited because of lack in the fringe carrier that depends on recording wavelength..

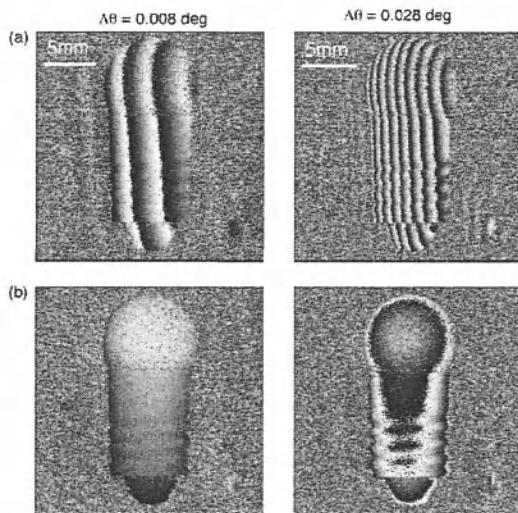


Figure 5-13. Phase difference before (a) and after (b) removal of tilt

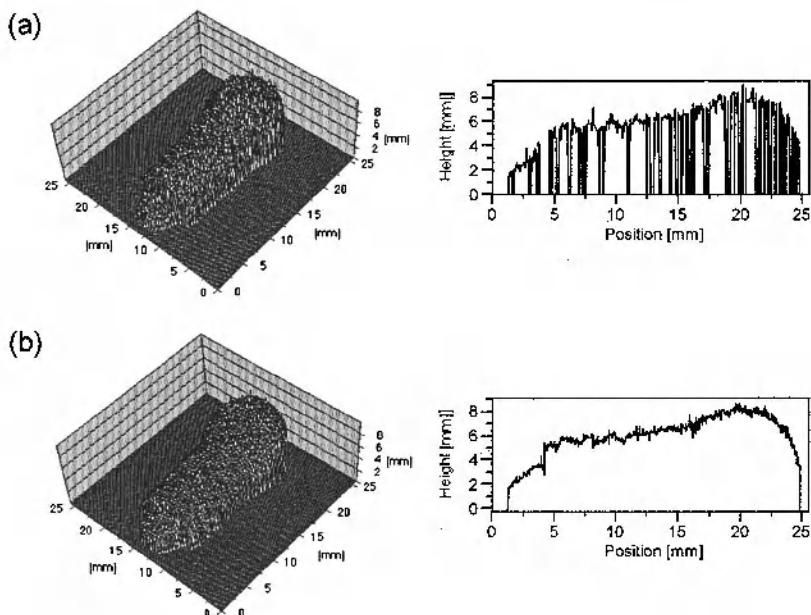


Figure 5-14. 3d-maps and cross-sections before (a) and after (b) nonlinear filtering

3.3 Deformation measurement

For deformation measurement the illumination is fixed and the phase difference of the reconstructed waves before and after object deformation depends on the vectors illustrated in Fig.5-15. If the speckle displacement is much smaller than the mean size, the phase difference becomes

$$\Phi(x,y) = -(\mathbf{k}_s - \mathbf{k}_o) \cdot \mathbf{a}(x,y) \quad (5-23)$$

This condition means that the systematic phase change can be observed as a result of cancellation of random phase variation between different speckles. It limits the measurement of in-plane deformation as illustrated in Fig.5-8.

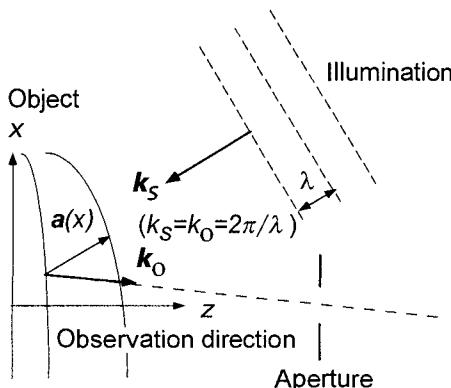


Figure 5-15. Principles of deformation measurement by digital holography

In order to measure both surface shape and deformation of diffusely reflecting surfaces we built the optical system shown in Fig.5-16. We employed a setup for image hologram so that limitation on the object size can be relaxed while ensuring high light flux incident on the CCD²⁴. Surface shape is measured by tilting a mirror for object illumination. This setup is the same as that of phase-shifting electronic speckle interferometry, but it can also be used for three-dimensional object by virtue of numerical focusing. We need not refocus the imaging lens on the position of interest and have only to record the hologram once before and after object deformation or mirror tilt. We also have much more freedom for suppression of speckle noise because phase difference is directly derived instead of fringe intensity detected in ESPI as mentioned above. Speckle displacement can be detected from the cross-correlation peak of the reconstructed intensities.

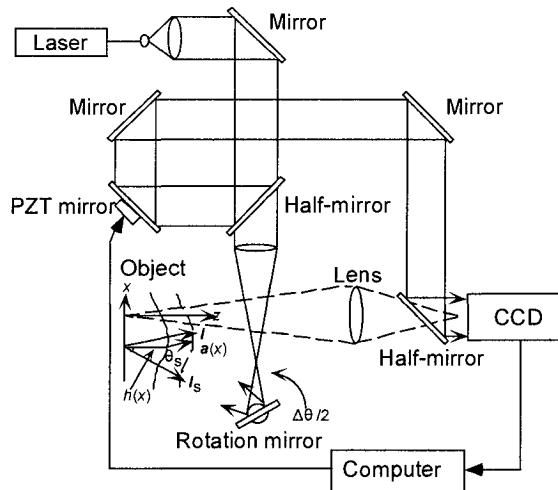


Figure 5-16. Setup of imaging digital holography for measurement of surface shape and deformation

Figure 5-17 shows the result of experiments conducted for out-of-plane deformation of a square plate of aluminum with the size of $50 \times 50 \text{ mm}^2$ and 1 mm thick. It was pushed at the center while being supported at the circular edge. The phase difference shown in Fig.5-17 (a) is unwrapped to deliver the distribution of out-of-plane displacement. The distribution of the displacement along the line indicated is displayed in Fig.5-17 (b). The resolution is estimated to be $0.01 \mu\text{m}$. The surface shape after total loading repeated 20 times is represented by Fig.5-18, where (a) is the phase difference arising from illumination tilt of $\Delta\theta=0.25$ degrees from $\theta_s=45$ degrees, (b) after removal of carrier component, and (c) is the cross-section along the indicated line. The fluctuation is about $10 \mu\text{m}$ that is an order of surface roughness.

If we use the two wavelength contouring, the setup becomes even simpler what would be very important for industrial applications. A unique capability of digital holography to record amplitude and phase simultaneously and to analyze them separately will provide new tools for shape and deformation measurements where digital speckle correlation technique for detecting in-plane displacement^{25, 26} can also be involved. This point will be extremely important because all the coherent-optical techniques consisting of holographic interferometry and speckle methods will be merged to digital holography that can treat three-dimensional information automatically and quantitatively with simple optical setups. Transmission of holographic data through internet and 3-dimensional reconstruction at remote locations is another incomparable feature of digital holography. Comparison with Finite Element Analysis will thus become easier.

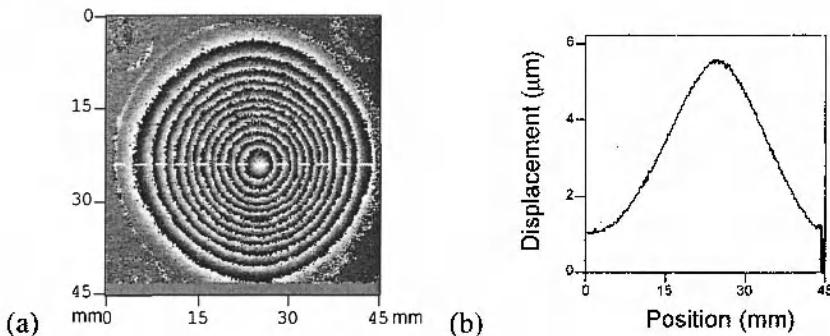


Figure 5-17. Phase difference due to out-of-plane displacement (a) and its cross-section after phase-unwrapping

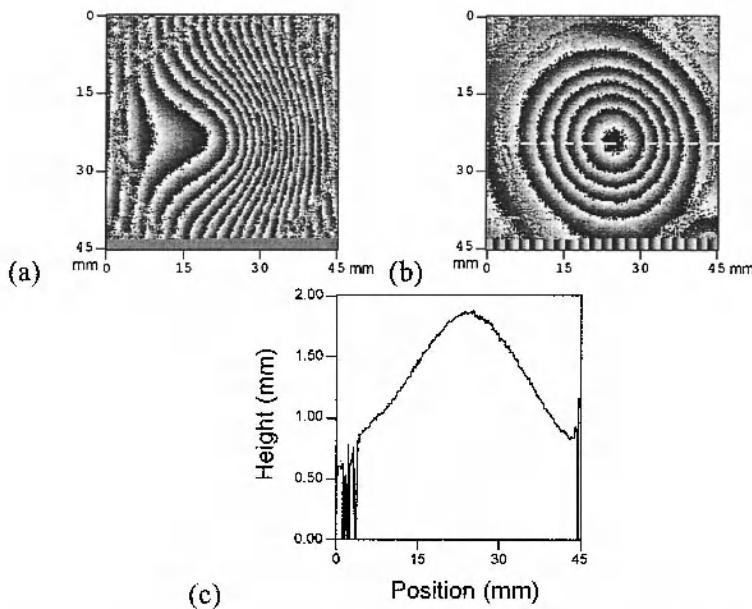


Figure 5-18. Phase difference due to tilt of mirror (a), that after removal of carrier (b), and a height profile along indicated the line (c)

3.4 Vibration analysis

Digital holography can also be applied to continuous deformation such as vibration²⁷. We used the setup shown in Fig.5-19 where single-mode optical fibers are used for guiding light from a single-mode He-Ne laser. We record time-averaged hologram intensity by keeping the vibration frequency much

higher than the frame rate of CCD that is equal to 30 Hz. Three time-averaged holograms are also recorded by shifting the reference phase. The object amplitude is given by

$$U_O(x', y', t) = U_E(x', y') \exp \left\{ k \left[a(x, y) \cos(\Omega t + \phi(x, y)) \right] \right\} \quad (5-24)$$

where U_E means the object amplitude representing the neutral position of vibration and Ω and ϕ denote angular frequency and phase of vibration, respectively. The resultant time-averaged intensity to be recorded by a CCD is given by

$$S_H(x, y; \delta) = \int I_H(x, y; \delta) dt = \int |U_R|^2 dt + \int |U(x, y, t)|^2 dt + 2 \Re \left[U_R \int U^*(x, y, t) dt \exp(i\delta) \right]. \quad (5-25)$$

This is substituted into Eq.(5-24) to derive time averaged object complex amplitude that is to be Fresnel transformed. The reconstructed intensity becomes

$$U_I(X, Y, -z_O) = U_{OA}(X, Y) = U_E(x', y') J_0 [ka(x, y)] \quad (5-26)$$

where J_0 means the 1st-kind Bessel function of the 1st order. This relationship suggests that we can observe the same fringe pattern as in conventional holography using photographic recording. The experimental intensity distributions obtained from a buzzer of 40 mm in diameter and vibrating at a resonant frequency of 5 kHz are displayed in Fig.5-20. We see that the pattern (a) resulting from the present method occupies the whole pixel number of the used CCD and is free from other additional components. However, the image (b) obtained from the off-axis setup and the same CCD contains the zero-th and the conjugate image together and occupies only one-nine-th of the whole pixels. Therefore, phase-shifting digital holography permits us to employ larger objects and provides much better image quality to improve the measurement accuracy and spatial resolution.

If we add AC phase modulation to the reference arm at the vibration frequency with a phase-modulator consisting of a PZT cylinder on which a single-mode fiber is wrapped, we can shift the position of the highest intensity from zero amplitude to an arbitrary amplitude corresponding to that of phase modulation²⁷. Thus we can extend the range of measurable amplitude. Although the reconstructed phase takes the values of either 0 or π , the distribution of phase might be utilized for identifying the positions of vibration nodes and antinodes with higher accuracy.

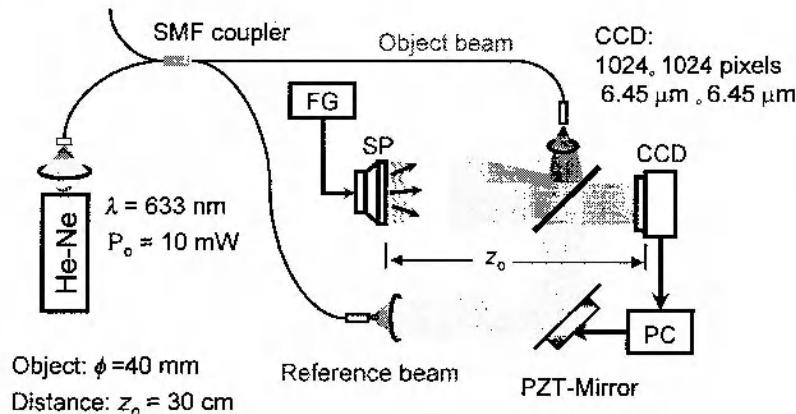


Figure 5-19. Setup for time-averaging vibration analysis using phase-shifting digital holography

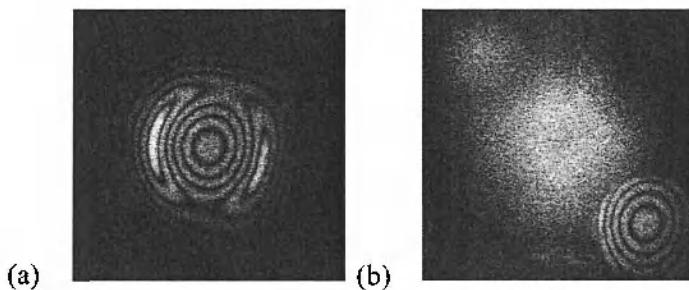


Figure 5-20. Reconstructed intensity of time averaged digital hologram. (a) Phase-shifting and in-line, (b) Off-axis

3.5 S/N ratio of the phase differences and its improvement

Accuracy of shape and deformation measurement depends on the dispersion of phase differences that is described by correlation properties of speckle. It is mostly affected by in-plane object displacement. At the conjugate plane of the object the speckle displacement is equal to in-plane displacement times imaging magnification, while at the defocused plane the gradient of Eq.(5-19) times an amount of defocusing adds to the above component. The plane where the speckle displacement becomes minimum correspond to the plane of fringe localization in holographic interferometry. In conventional holographic interferometry it was difficult to compensate for

the effect of in-plane displacement. It was only possible by real time interferometry where hologram position is adjusted and by a sandwich hologram²⁸ where the mutual position of the layered holograms was shifted. In digital holography, however, these mechanical adjustments are got rid of and mutual displacement between the reconstructed phase distributions can be delivered automatically and precisely from the peak position of the cross-correlations of phase or intensity distributions. In this way correlation between the laterally shifted wavefronts can be easily restored.

In the case of surface contouring there is always no lateral shift of the resultant wavefronts at the focused plane. At the defocused plane, however, the speckle displacement occurs owing to the change of illumination²⁹. This situation indicates that surface contouring of 3-dimensional object is limited by the speckle displacement at the defocused plane. In digital holography, however, this speckle displacement can be reduced to be zero by numerical refocusing at the desired plane. The refocusing also eliminates image blurring caused by defocusing. For these reasons image contouring by digital holography provides us with higher performance and larger measurement depth than the fringe projection method. The speckle noise due to random change of the amplitude can also be suppressed if we exclude the phase differences at the pixels where intensity is too low as mentioned in the experiments above. Since the procedure for deriving object phases is nonlinear with respect to intensity, a new method of speckle noise suppression might be expected that has been insufficient with linear averaging.

4. CONCLUSIONS

In this chapter we surveyed the principle and applications of phase-shifting digital holography that has raised performances of digital holography substantially. Since it delivers the values of complex amplitude at the CCD plane directly without any filtering process as required for the off-axis setup, both derivation and processing of information required for transmission, display, and measurement are straightforward and simple. The full number of CCD pixels is utilized for the entire image to provide image quality much better than in the off-axis setup. We discussed the image formation and applications to measurement of surface shape and deformation of diffusely reflecting objects.

In the section of image formation 3-dimensional microscopy has been explained that can applied for both amplitude and phase objects. Color holography using a laser emitting three-wavelength simultaneously and a

color CCD enables quick hologram recording and reconstruction with a satisfactory color balance.

For increasing memory capacity and transmission rate to remote places the derived complex amplitude can be compressed into phase-only data with a few bits without serious deterioration of image quality in the case of diffusely reflecting objects.

In measurements of surface shape and deformation the quantities to be evaluated in phase-shifting digital holography has been explained in comparison with those observed in conventional holographic interferometry and electronic speckle pattern interferometry (ESPI). We could say that digital holography, especially, phase-shifting digital holography has unified not only these methods but also will comprise speckle correlation technique together that detects speckle displacement and/or decorrelation due to surface deformation or illumination change.

The most serious issue in phase-shifting digital holography will be the recording time required for at least three phase-shifted patterns. In the experiments reported above it was 1 s with 1,024x1,024 pixels with 12 bits, while the reconstruction took a few seconds. The recording time could be shortened to less than 0.03 s by using a CCD with the frame rate of 120 per second for measuring the shape of liquid surface under convection³⁰. The limitation in recording time that needs at least three frames will be less serious when a high speed CCD is used. The simplicity in the reconstruction process also eases introduction of data compression for storage, transmission, and display of 3-d information based on 2-dimensional array of complex amplitude.

ACKNOWLEDGEMENT

The author thanks M. Yokota, F. Zhang, and G. Mills of Gunma University for helpful discussions. He also thanks J. Kato of RIKEN for providing us with the program for phase-unwrapping.

REFERENCES

1. J. W. Goodman and R. W. Lawrence, "Digital image formation from electronically detected holograms," *Appl. Phys. Lett.* **11**, pp.77-79, 1967.
2. M. A. Kronrod, N. S. Merzlyakov, and L. P. Yaroslavskii, "Reconstruction of a hologram with a computer," *Sov. Phys.-Tech. Phys.* **17**, pp. 333 - 334, 1972.
3. U. Schnars and W. Juptner, "Direct recording of holograms by a CCD target and numerical reconstruction," *Appl. Opt.* **33**, pp. 179-181, 1994.

4. G. Bruning, D. R. Herriott, J. E. Gallagher, D. P. Rosenfeld, A. D. White, and D. J. Brangaccio, "Digital wavefront measuring interferometer for testing optical surfaces and lenses," *Appl. Opt.* **13**, pp.2693-2703, 1974.
5. I. Yamaguchi and T. Zhang, "Phase-shifting digital holography," *Opt. Lett.* **22**, pp.1268-1270, 1997.
6. F. Zhang, I. Yamaguchi, and L. P. Yaroslavsky, "Algorithm for reconstruction of digital holograms with adjustable magnification," *Opt. Lett.* **29**, pp. 1668-1670, 2004.
7. I. Yamaguchi, T. Matsumura, and J. Kato, "Phase-shifting color digital holography," *Opt. Lett.* **27**, pp. 1108-1110, 2002.
8. J. Kato, I. Yamaguchi, and T. Matsumura, "Multicolor digital holography with an achromatic phase-shifter," *Opt. Lett.* **27**, pp. 1403-1405, 2002.
9. I. Yamaguchi, "Three-dimensional microscopy and measurement by phase-shifting digital holography," *Selected Papers from Fifth International Conference on Correlation Optics*, O. F. Angelsky, Editor, Proc. SPIE **4607**, pp.153-160, 2002.
10. O. Matoba, T. J. Naughton, Y. Frauel, N. Bertaux, and B. Javidi, "Real-time three-dimensional object reconstruction by use of a phase-encoded digital hologram," *Appl. Opt.* **41**, pp. 6187-6192, 2002.
11. I. Yamaguchi, J. Kato, S. Ohta, and J. Mizuno, "Image formation in phase-shifting digital holography and applications to microscopy", *Appl. Opt.* **40**, pp. 6177-6186, 2001.
12. T. Zhang and I. Yamaguchi, "Three-dimensional microscopy with phase-shifting digital holography," *Opt. Lett.* **23**, pp.1221-1223, 1998.
13. G. Indebetouw, A. E. Maghnouji, R. Foster, "Scaning holographic microscopy with transverse resolution exceeding the Rayleigh limit and extended depth of focuse," *J. Opt. Soc. Am.* **22**, pp. 892-898, 2005.
14. G. H. Mills and I. Yamaguchi, "Effects of quantization in phase-shifting digital holography," *Appl. Opt.* **44**, pp. 1216-1225, 2005.
15. L. B. Lesem, P. M. Hirsch, and J. A. Jordan, Jr., "The kinoform, a new wavefront reconstruction devices," *IBM J. Res. Dev.* **13**, pp.150-155, 1969.
16. I. Yamaguchi, K. Yamamoto, G. A. Mills, and M. Yokota, "Image reconstruction only by phase in phase-shifting digital holography", *Appl. Opt.* **44**, 2005, accepted for publication.
17. T. J. Naughton, Y. Frauel, B. Javidi, and E. Tajahuerce, Compression of digital holograms for three-dimensional object reconstruction and recognition," *Appl. Opt.* **41**, pp. 4124-4132, 2002.
18. T. J. Naughton, J. B. McDonald, and B. Javidi, "Efficient compression of Fresnel fields for Internet transmission of three-dimensional images," *Appl. Opt.* **42**, pp. 4758-4764, 2003.
19. M. Takeda, H. Ina, and S. Kobayashi, "Fourier transform methods of fringe pattern analysis for computer based tomography and interferometry," *J. Opt. Soc. Am.* **72**, pp. 156-170, 1984.
20. I. Yamaguchi, "Fringe formations in deformation and vibration measurements using laser light," in *Progress in Optics*, ed. E. Wolf (North-Holland Pub., Amsterdam, 1985) Vol.22, pp.271-340.
21. I. Yamaguchi, "Holography, speckle, and computers," *Opt. and Lasers in Eng.* **39**, pp. 411-429, 2003.
22. I. Yamaguchi, J. Kato, and S. Ohta, "Surface shape measurement by phase-shifting digital holography," *Opt. Rev.* **8**, pp. 85-89, 2001.
23. I. Yamaguchi, S. Yamashita, and M. Yokota, "Surface shape measurement by dual-wavelength phase-shifting digital holography, *Fringe 2005*, Proc. of 5th International Workshop on Automatic Processing of Fringe Patterns, ed. W. Osten and W. Juptner, pp.396-403, 2005.

24. I. Yamaguchi, J. Kato, and H. Matsuzaki, "Measurement of surface shape and deformation by phase-shifting image digital holography," Opt. Eng. **42**, pp. 1267-1271, 2003.
25. I. Yamaguchi: "Real-time measurement of in-plane translation and tilt by electronic speckle correlation", Jpn. J. Appl. Phys. **19**, L133-L136, 1980.
26. S. Noh and I. Yamaguchi: "Two-dimensional measurement of strain distribution by speckle correlation," Jpn. J. Appl. Phys. **31**, L1299-L1301, 1992.
27. F. Zhang, J. D. R. Valera, I. Yamaguchi, M. Yokota, and G. Mills, "Vibration analysis by phase-shifting digital holography," Opt. Rev. **11**, pp.297-299, 2004.
28. N. Abramson, "Sandwich hologram interferometry, a new dimension in holographic comparison," Appl. Opt. **13**, pp.2019-2025, 1974.
29. A. Yamamoto and I. Yamaguchi, "Profilometry of sloped plane surfaces by wavelength scanning interferometry," Opt. Rev. **9**, pp.112-121, 2002.
30. I. Yamaguchi, O. Inomoto, and J. Kato, "Surface shape measurement by phase-shifting digital holography," *Fringe 2001*, Proc. of 4th International Workshop on Automatic Processing of Fringe Patterns, ed. W. Osten, pp.365-372, 2001.

Chapter 6

CONTROLLING IMAGE RECONSTRUCTION PROCESS IN DIGITAL HOLOGRAPHY

Pietro Ferraro¹, Sergio De Nicola² and Giuseppe Coppola³

¹ *Istituto Nazionale di Ottica Applicata del CNR Via Campi Flegrei, 34 80072 Pozzuoli (Na), Italy*

² *Istituto di Cibernetica "E.Caianiello" del CNR, Via Campi Flegrei, 34 80072 Pozzuoli (Na), Italy*

³ *Istituto per la Microelettronica e Microsistemi del CNR Istituto per la Microelettronica e Microsistemi del CNR Via P. Castellino, 111 - 80131 Napoli (Italy)*

Abstract: The reconstruction of digital holograms is a full numeric process. Such peculiarity offers the possibility to control several parameters during the reconstruction process. In this chapter we will describe some recent advancements of Digital Holography and how it is possible to control the numerical reconstruction process by optimizing or regulating different parameters. By controlling the reconstruction process it is possible to overcome some problems arising during the optical tests of materials and devices or to improve the imaging capability of DH for example for color 3D imaging.

Key words: Digital holography; holographic interferometry; interference microscopy; aberration compensation; phase-measurement; phase contrast; multiwavelength display;

1. INTRODUCTION

Digital Holography has some special attributes making it unique in the field of interferometric methods. Digital Holography has of course the very same characteristics of classical film holography for imaging a 3D amplitude field but at same time it has the ability to manage the phase of the reconstructed optical field. In fact while in classical film holography the reconstructed optical phase allows to image the amplitude field along different direction (responsible of the three-dimensional prospective/perception for an observer), in DH the phase information of the reconstructed wave-front is available numerically and can be easily handled giving great opportunities either in the scientific and engineering playgrounds of 3D imaging and high sensitivity interferometry. DH has been successfully applied in fields such as deformation analysis, object contouring, measurement of particle position, and encryption [1-4].

In DH phase distribution can be obtained and used quantitatively for phase-contrast imaging in microscopy, by using an off-axis holographic optical recording set-up to perform metrological measurement. The possibility to manage the phase is very attractive in digital holography because aberrations can be removed. In fact, it has been proved that wave front curvature introduced by microscope objective and lenses can be successfully removed and/or compensated as well as other type of aberrations like spherical type, astigmatism and anamorphism, or even longitudinal image shifting introduced by cube beam-splitter.

The chapter will be divided in 5 sections. In section 2, we will describe how it is possible to compensate some common optical aberration to get the correct phase map of the object under investigation. Section 3 will be devoted to illustrate how it is possible to control the pixel size in the reconstructed image plane independently form the reconstruction distance and/or the wavelength exploiting all the potentialities of DH as tool either for metrological application as well as for 3D full color imaging. In section 4 it will be demonstrated that combining the DH in multiwavelength operation with aberration compensation it is possible to retrieve the correct phase map of phase objects. Finally in section 5 it will be discussed some how to overcome some difficulties that could arise in using DH for inspection of micro-objects. Examples will be given of how it is possible to track automatically and in real time the focus for objects moving during the recording process. Then an example of how to recover the spatial resolution in the reconstructed image plane will be discussed.

2. COMPENSATING ABERRATIONS IN DIGITAL HOLOGRAPHY

Complete compensation of aberrations is of fundamental importance when the quantitative phase determination is needed for metrological applications. In fact compensation of aberrations is a way for controlling the shape of the reconstructed wave front to recover the correct information useful for physical measurements. For this reason in this section it will be described some procedures which have been developed recently to compensate some typical optical aberrations. In fact when optical elements are adopted in an interferometric set-up is unavoidable the occurrence of wave front aberrations.

2.1 Compensation of defocus aberration

One of the most common optical wave front aberrations is the defocus aberration. It consists in a parabolic phase factor whose curvature depends on the magnification power of the optical system. It is introduced naturally every time a magnification is obtained by means of lenses. This is especially true in microscope configurations as shown in Fig.6-1. In fact, natural wave front curvature is introduced by the microscope objective hindering the phase contrast imaging at the reconstruction planes.

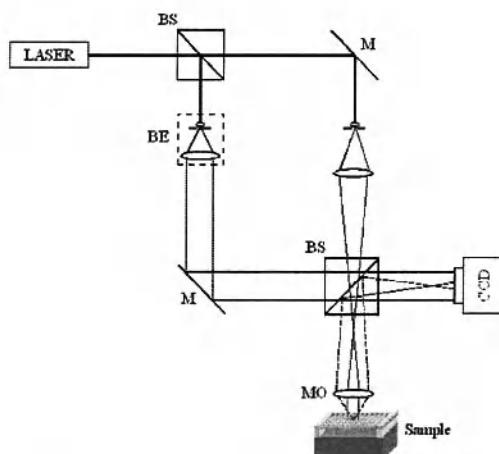


Figure 6-1. Experimental set-up for recording digital holograms; BS – beam splitter; M – mirror; MO – microscope objective; BE – Beam expander

For removing defocusing aberrations of the wave front emerging from a microscope objective different approaches are possible. In one approach, as demonstrated in Ref. [5], it is required an accurate knowledge of several optical parameters. The parameters are: focal lengths of the optical components used for imaging the object, the wave vector components K_x and K_y of the reference wave front for off-axis DH configurations, and finally distances lens-object and lens-CCD. Equations for different experimental configurations have been derived for different usual optical configurations both in-line and off-axis. Obviously, the main difficulty of the abovementioned approach is to evaluate involved distances with accuracy. As consequence, the employed compensation phase factor will remove the undesired spherical phase factor partially. As has been verified in different papers, fine digital adjusting is necessary to completely remove the curvature before using the phase contrast imaging as quantitative information [6,8]. To illustrate the problem it can be considered the original intrinsic phase of the object beam (see Fig. 6-1) consisting of a parabolic phase factor superimposed to the phase distribution $\Delta\phi(x, y)$ of the object wave front caused by the imaging microscope objective. Due to the high focal power of the microscope objective the recorded hologram at the CCD plane array appears basically as an ensemble of circular fringes slightly disturbed by the characteristic object phase distribution. In Fig. 6-2a is shown a typical hologram of a silicon MEMS, 10 μm wide and 20 μm long. From this hologram, circular fringes are clearly visible. While amplitude numerical reconstruction at distance d provides good contrast image (Fig. 6-2b), the presence of the high curvature inherent into the reconstructed phase wave front, hinders the phase information content of the reconstructed phase (Fig. 6-2c).

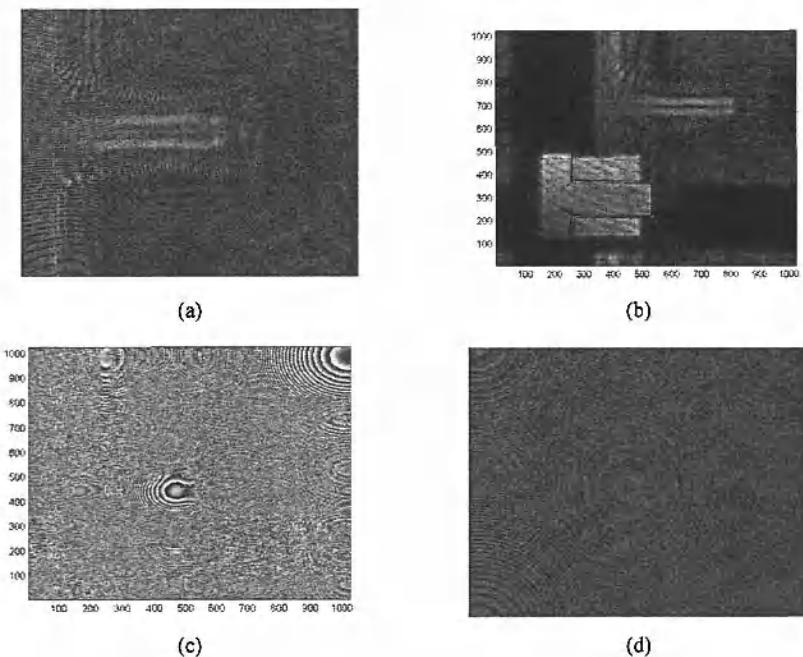


Figure 6-2. Example of spherical aberration compensation: (a) recorded hologram of the cantilever microstructure; (b) amplitude reconstruction with spherical aberration compensation; (c) numerical reconstruction of the wrapped phase map without the removal of spherical aberration; (d) wrapped phase map reconstructed from the reference hologram.

The phase could be corrected by means of a phase factor introduced into the holographic numerical reconstruction process and determined on the basis of the geometrical parameters of the optical imaging set-up at the reconstruction plane [6, 7]. But, an alternative way to proceed consists in calculating or estimating the unwanted wave front curvature in the hologram plane (the plane of the CCD array) instead that in the reconstruction plane.

Once that the parabolic phase correction factor has been calculated in the hologram plane it is straightforward to demonstrate that compensation of the curvature can be easily obtained at all reconstruction planes, since the correction factor is numerically reconstructed in exactly the same way as is made for the hologram.

The aforementioned approach consists essentially in three steps. i) a phase correction wave front is calculated using one of the three methods described below; ii) the numerical holographic phase reconstruction at distance d is performed of both the hologram and correction wave front; iii) the two reconstructed phase fields are subtracted at the reconstruction plane at distance d for obtaining the contrast phase image of the object with compensation of the inherent curvature. Three methods, can be used to

accomplish the first step. One way to do it is to use the same hologram to recover the curvature correction factor. In fact, if there is a portion of the hologram area, as shown in Fig. 6-2a, in which the circular fringes are not affected by the phase content of the object under investigation, then the phase correction factor can be calculated in that area and extrapolated to all the hologram.

A second method can be applied if the object is made of plane mirror-like surface. For example, the micro structures, that we have investigated, are realized on a silicon surface, which can be accounted as a good reflecting face. In this case the correction wave front can be evaluated by a sort of holographic double-exposure procedure. A first hologram (or multiple phase shifted holograms if the case) of an area with the micro structures is recorded. Then, simply translating the sample transversely, in respect to the optical axis, a second hologram (or multiple phase shifted holograms if the case) of a flat area in proximity of previous one, but without the microstructures, is recorded. The phase obtained by this last hologram is itself the curvature correction wave front. A third approach is to generate a synthetic phase distribution similar to those causing circular fringes at the CCD plane by observing the fringe pattern. Fine but tedious digital adjusting to obtain the correct phase factor at the hologram plane can be performed. Independently by the way adopted for evaluating the correction phase wave front, the above steps ii) and iii) are implemented for removing the unwanted curvature at whichever reconstruction plane.

Here will be shown an example of application using the second method that essentially consists in a sort of a double-exposure digital holographic interferometry. The first exposure is made of the object under investigation, whereas the second one is made of a flat reference surface in proximity of the object. This is a typical case meet, for example, when MEMS structures are inspected, because the micro-machined parts are fabricated on a flat silicon substrate. In that case the area around the micro-machined structure offers a very good plane surface that can be used as reference. In Fig. 6-3a is shown a recorded hologram of a cantilever realized on a silicon wafer. The micro-machined beam is 10 μm wide 20 μm long. If the wafer is slightly translated transversally, an additional hologram (Fig. 6-3b) can be recorded and used as reference.

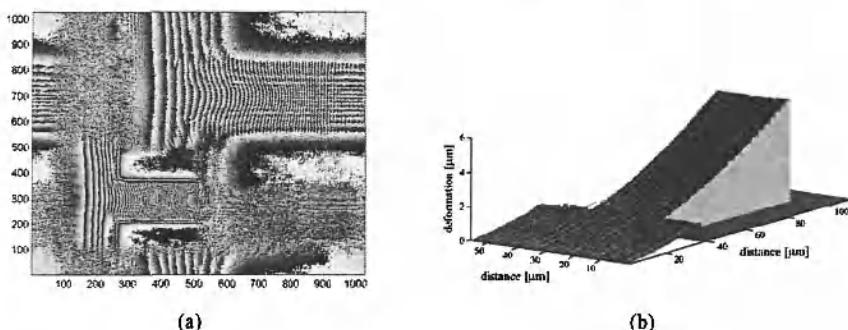


Figure 6-3. (a) Reconstruction of the wrapped phase map with removal of spherical aberration; (b) 3D profile of the cantilever microstructure obtained by unwrapping the phase map shown in (a)

Both the holograms are numerically reconstructed at distance $d = 100\text{ mm}$ by means of the Eq.(6-1). Thus, at this distance, i.e. in the reconstructed image plane, it is possible calculate the phase difference between the two holograms and, as consequence, to obtain the phase map of the object cleared of the wave front curvature introduce by the imaging microscope objective as shown in Fig. 6-3c and Fig. 6-3d wrapped and unwrapped phase, already converted in microns, obtained at distance $d=100\text{mm}$ are shown. This method is easy to apply but at the expense of a double image recording. Moreover the phase map could be affected by some errors due to the presence of eventual local imperfection on the reference surface. The displacement operation need to record the reference hologram can also introduce systematic errors if, for example, some amount of tilt is introduced during the translation.

As said above two more approaches exist to remove the defocus aberration for retrieving the correct phase map based on modelling a parabolic phase factor to subtract to the phase map obtained from one single digital hologram details about those methods can be find in [7].

2.2 Compensation of anamorphism and astigmatism

One other type of aberration that instead can affect both amplitude and phase is anamorphism. Usually such kind of aberration is produced for example by a diffraction grating [8]. In Fig. 6-4a is described for example a simple optical setup made with white light source a lamp.

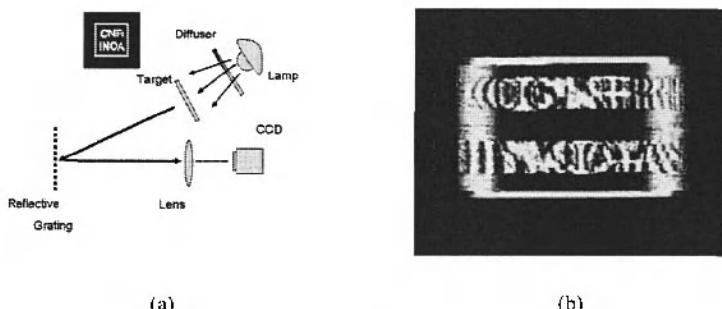


Figure 6-4. (a) Experimental setup for direct imaging through reflective grating G in white light; (b) test object imaged through the reflective grating showing strong anamorphism

If a target (like that shown in the inset of Fig. 6-4a is imaged through a reflective diffraction grating, its image results totally aberrated due to the severe anamorphism introduced along the direction perpendicular to the lines of the grating (Fig. 6-4b). In the same way if a digital hologram is recorded by using a set-up like that shown in Fig. 6-5 in which the object beam has been obtained through a diffraction grating, then again the numerical reconstruction will give an aberrated amplitude image of the object as shown in Fig. 6-6.

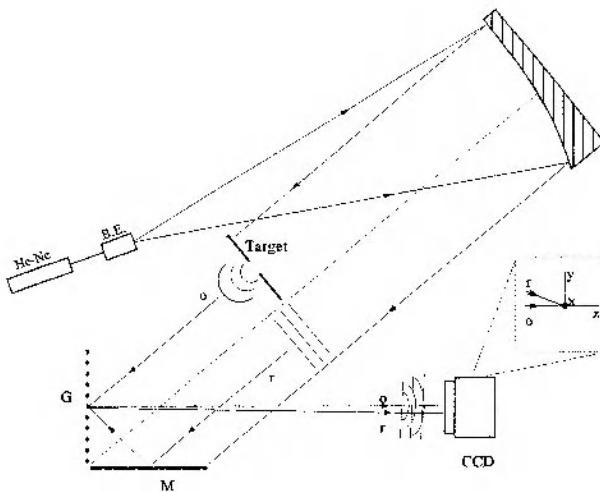


Figure 6-5. RGI in a holographic setup; BE, beam expander; M, mirror; G, grating; α , object beam; r , reference beam

Differently from the incoherent case, now the aberration can be removed since a 3D holographic numerical reconstruction process is involved.

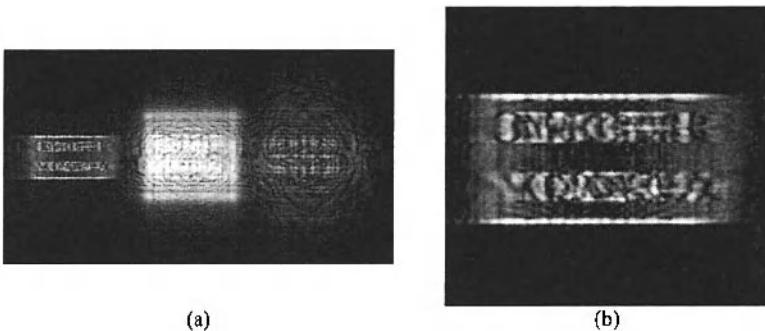


Figure 6-6. Incorrect-image holographic reconstruction of the test object; (b) magnified image

In fact it can be easily demonstrated that by modifying the reconstructing diffraction integral adopted in the Fresnel Transformation Method (FTM) usually given by

$$o(x, y) \propto \iint h(x', y') r(x', y') \exp\left\{\frac{i\pi}{d\lambda} (x'^2 + y'^2)\right\} \exp\{-2i\pi(x'v + y'\mu)\} dx' dy' \quad (6-1)$$

in the following form:

$$o(x, y) \propto \iint h(x', y') r(x', y') \exp\left\{\frac{i\pi}{d\lambda} (x'^2 + (y' \cos \alpha)^2)\right\} \exp\{-2i\pi(x'v + y'\hat{\mu})\} dx' dy' \quad (6-2)$$

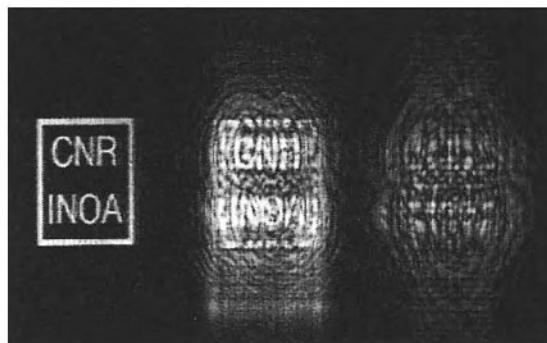


Figure 6-7. Reconstruction of the test object at distance $d=240$ mm with anamorphic correction and compensation for astigmatism aberrations

In which the following transformation $y' \rightarrow y'\cos(\alpha)$ along the y-axis has been adopted. As shown in Fig. 6-7 the aberration has been completely removed and the letter of the reconstructed target is clearly readable [9].

It is important to note that it's the special attribute of DH to use the reconstruction of complex wave front scattered by the object that allows to compensate the anamorphism while in the incoherent case such type of anamorphism cannot be removed. Either reconstructions of Fig. 6-6 and Fig. 6-7 were obtained at the same distance $d=240$ mm.

However while the letters are clearly legible in Fig. 6-7 where the aberration was removed a deeper analysis of the reconstructed amplitude image show the presence of another kind of aberration, the astigmatism, as predicted by the theoretical analysis of the optical system [10]. In fact looking at an enlarged image (see Fig. 6-8a) of the reconstructed target at distance $d=240$ mm of Fig. 6-7 it is possible to note that a blurring effect appears on vertical lines of letters.

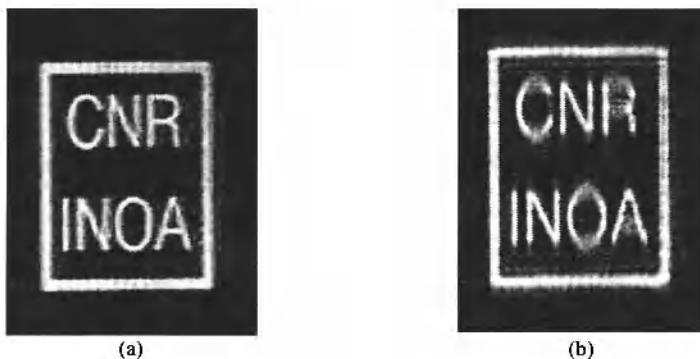


Figure 6-8. (a) Image reconstruction of the test object at distance $d=240$ mm with anamorphic correction and without compensation for astigmatism aberrations; (b) Image reconstruction of the test object at distance $d=260$ mm with anamorphic correction and without compensation for astigmatism aberrations

For example the letter "I" appears to be out of focus. If one more numerical reconstruction is performed at different distance $d=260$ mm the letters appears to be legible but now the horizontal parts of the letter results to be out of focus. By comparing the Fig. 6-8a and Fig. 6-8b it is clear that does not exist a single image plane in which all the letter are in focus: it is clear that a sagittal and a tangential focuses are present due the presence of the astigmatism. Such aberration can be removed by adopting inside the diffraction integral of Eq.(6-1) the following expression

$$\exp\left\{\frac{i\pi}{\lambda}\left(\frac{x'^2}{d_x} + \frac{y'^2}{d_y}\right)\right\} \quad (6-3)$$

where two distance have been adopted for each coordinate axis ($d_x=240$ and $d_y=260$ mm). The result of the reconstruction is shown in Fig. 6-9 in which is noticeable that letters appears all in good focus.



Figure 6-9. Image reconstruction of the test object at distance $d=240$ mm with anamorphic correction and compensation for astigmatism aberrations

2.3 Compensation of spherical aberrations

Other kind of aberrations can be removed in DH as demonstrated in ref. [11] by modeling a pupil function to use in conjunction with the diffraction integral in the FTM reconstructing mode. They demonstrated that spherical aberration can be compensated for improving the accuracy in the metrology of microstructures. One more example was can be found in the paper in ref. [12]

3. CONTROLLING NUMERICAL RECONSTRUCTIONS IN DIGITAL HOLOGRAPHY

In digital holography (DH), holograms are digitized by means of a solid-state array light detector, and the numerical reconstruction procedure is based either on the calculation of the diffraction integral by a Fresnel transformation method (FTM) or a Convolution Method (CM) [13,14,15]. In FTM, the reconstruction pixel (RP) increases with the reconstruction distance or the recording wavelength so that the size of the reconstructed image of the object, in terms of number of pixels, is reduced for longer

distances or greater wavelengths. By contrast, in the CM the RP does not change but remains equal to the pixel size of the recording array. The CM is more appropriate for reconstruction at small distances whereas the FTM is useful for longer distances according to the paraxial approximation necessary to apply it.

In FTM the object wave field is reconstructed numerically by using the discrete form of the Rayleigh-Sommerfield diffraction formula in the Fresnel approximation, namely

$$\begin{aligned} \psi(r\Delta\nu_\xi, s\Delta\nu_\eta) = & \frac{1}{i\lambda d} \exp\left(i\frac{2\pi}{\lambda}d\right) \exp[i\pi\lambda d(r^2\Delta\nu_\xi^2 + s^2\Delta\nu_\eta^2)] \times \\ & \times \Delta x \Delta y \sum_{n=-N/2}^{N/2-1} \sum_{m=-M/2}^{M/2-1} I(n\Delta x, m\Delta y) R(n\Delta x, m\Delta y) w(n\Delta x, m\Delta y) \times \\ & \times \exp\left[-i2\pi\left(\frac{rn}{N} + \frac{sm}{M}\right)\right] \end{aligned} \quad (6-4)$$

where $I(n\Delta x, m\Delta y)$ is the two-dimensional spatial sampling of the hologram intensity distribution $I(x, y)$ on a rectangular raster of $N \times M$ points and $\psi(r\Delta\nu_\xi, s\Delta\nu_\eta)$ are the sampled complex values of the reconstructed field $\psi(\xi, \eta)$ at discrete points $(r\Delta\nu_\xi, s\Delta\nu_\eta)$ with and r, s, m and n are integer numbers.

Eq.(6-4) allows to compute a matrix of $N \times M$ complex numbers of the reconstructed field via the discrete two-dimensional fast Fourier transform algorithm. According to the theory of discrete Fourier transform, the sampling frequencies interval are given $\Delta\nu_\xi = 1/N\Delta x$ and $\Delta\nu_\eta = 1/M\Delta y$ where Δx and Δy are the distances between the neighbouring pixels on the CCD array in the horizontal and vertical directions, respectively. Since the spatial frequencies ν_ξ and ν_η corresponding to the spatial variables ξ and η in the reconstruction plane are related to the reconstruction wavelength λ and to the reconstruction distance d by the relation

$$\nu_\xi = \frac{\xi}{\lambda d}, \quad \nu_\eta = \frac{\eta}{\lambda d} \quad (6-5)$$

we have that the dimensions of the RP are given by

$$\Delta\xi \times \Delta\eta = \lambda d / N\Delta x \times \lambda d / M\Delta y \quad (6-6)$$

3.1 Controlling the pixel size in the reconstructed image plane as function of the distance

From Eq.(6-6) it is clear that the reconstructed image is enlarged or contracted according to the reconstruction distance d measured backward from the hologram plane to the image plane (ξ, η) . The RP depends on the number of pixel $N \times M$ in the recording pixel array. Since the size of reconstructed objects becomes smaller the larger the reconstruction distance, it is not possible to compare directly two well-focused amplitude images reconstructed at different distances.

Therefore, when using FTM for hologram reconstruction, it is necessary to use a resizing operation on the reconstructed images [16] at end of the reconstruction process or a scaling operation on the hologram.

Recently, it has been proposed the zero-padding method (ZPM) that exploits the FTM itself to control the size of the reconstructed image through enlargement of the number of the pixels of the recorded digital holograms.

Considering two holograms recorded at different distances d_1 and d_2 with array size $N_1 \times M_1$ then according to the ZPM, the image size can be controlled changing the RP by using a larger number of pixels in the reconstruction process. N_1 and M_1 can be augmented by padding the matrix of the hologram recorded at the greater distance d_2 ($d_1 < d_2$) with zeros in both the horizontal and vertical directions such that

$$\begin{aligned} N_2 &= N_1(d_2 / d_1) \\ M_2 &= M_1(d_2 / d_1) \end{aligned} \quad (6-7)$$

From Eqs. (6-6) and (6-7) the dimensions of the RP at the two distances d_1 and d_2 are equalized and are given by

$$\begin{aligned} \Delta\xi_1 = \Delta\xi_2 &= \frac{d_1\lambda}{N_1\Delta x} = \frac{d_2\lambda}{N_2\Delta x} \\ \Delta\eta_1 = \Delta\eta_2 &= \frac{d_1\lambda}{M_1\Delta x} = \frac{d_2\lambda}{M_2\Delta x} \end{aligned} \quad (6-8)$$

Thus, in order to obtain two reconstructed images with the same size, the matrix $N_2 \times M_2$ of the hologram recorded at distance d_2 is padded with $(N_2 - N_1)$ zeros along the x direction and with $(M_2 - M_1)$ zeros along the y direction.

An example of application of the ZPM for controlling the size of the amplitude reconstruction by FTM is shown in Fig. 6-10.

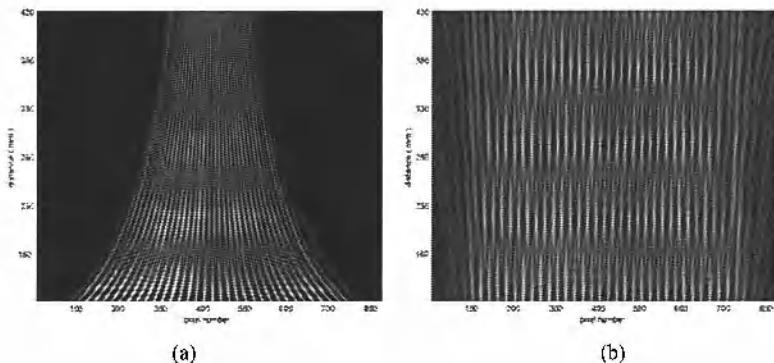


Figure 6-10. Ronchi grating reconstructed at different distances: (a) pitch of the grating decreases for longer distances; (b) size is kept unchanged with padding operation

A portion of the amplitude reconstruction along the longitudinal z-axis for a Ronchi grating without the padding operation is shown in Fig. 6-10a. The initial reconstruction distance is $d_1 = 295 \text{ mm}$, the final was $d_2 = 695 \text{ mm}$ and the number of pixels for the first reconstruction is $N_1=M_1=1024$. In the left figure, it is clearly noticeable that the pitch of the grating decreases for longer distances. On the right the amplitude reconstructions at different distances are calculated by using the results of the same reconstruction when the padding operation applied; in particular final reconstruction was performed with padding of zeros up to $N_2=M_2=2412$. In Fig. 6-10b it can be noted that the size of the grating is kept constant [17].

3.2 Controlling the pixel size in the reconstructed image plane as function of wavelength

Important applications of digital holography rely on the possibility to use multiple wavelengths digital holography (MWDH) to record and reconstruct coloured objects [3,4]. However, use of MWDH for metrological applications demands some caution when combining reconstructions at different wavelengths. In MWDH for each wavelength the width of the RP increases with the reconstruction wavelength for a fixed reconstruction distance. Consequently, holograms recorded with different wavelengths produce images with different sizes, when numerically reconstructed by means of the FTM. Colour DH display requires simultaneous reconstruction

of images recorded with different wavelengths (colours) and the resulting reconstructed images must be perfectly superimposed to get a correct colour display. A proper resizing operation of the reconstructed images can be applied at the end of the reconstruction process. The ZPM discussed in the previous section, being intrinsically embedded in the reconstruction process, does not require image scaling at the end of the reconstruction process and it can be applied to MWDH when the object size has to be kept constant for each wavelength, or in the general case, when both the recording distances and wavelengths can be varied independently.

In MWDH, if the first hologram has been recorded with wavelength λ_1 and the second with wavelength λ_2 with $\lambda_2 > \lambda_1$ while the distance remained unchanged then the number of pixels of the second hologram must be changed such that $N_2 = N_1(\lambda_2/\lambda_1)$ to satisfy the following relation

$$\Delta x_1' = \Delta x_2' = \frac{d' \lambda_1}{N_1 \Delta \xi} = \frac{d' \lambda_2}{N_2 \Delta \xi} \quad (6-9)$$

in order to obtain the same width for the reconstruction pixel as required in multi-wavelength DH holography used for colour display and/or for applications in metrology. In multi-wavelength DH different wavelengths are used and for each wavelength the width of the reconstruction pixel increases with the reconstruction wavelength, for a fixed reconstruction distance. Consequently, digital holograms recorded with different wavelengths produce images with different sizes, when numerically reconstructed by means of the FTM.

Colour DH display needs simultaneous reconstruction of images recorded with different wavelengths (colours) and the resulting reconstructed images must be perfectly superimposed to get a correct colour display [3]. The diverse size for each image inhibits a correct superposition in all cases in which multi-wavelength operations is required for display application or for direct phase comparison required for holographic interferometry [18, 19].

To avoid the above problems, the convolution approach could be adopted in which the reconstruction pixel remains constant and equal to the size of the pixel of the CCD array; however for large reconstruction distance this approach does not work properly.

Otherwise, as proposed in literature, it is necessary a resizing operation of the reconstructed images at end of the reconstruction process, or a scaling operation of the hologram has to be done for taking into account such problem [3, 18, 19].

In order to demonstrate the size can be kept constant also in multi-wavelengths operation two holograms can be recorded with two different wavelengths at $\lambda_1=532\text{ nm}$ and $\lambda_2=632.8\text{ nm}$, respectively by means of the set-up of Fig. 6-11.

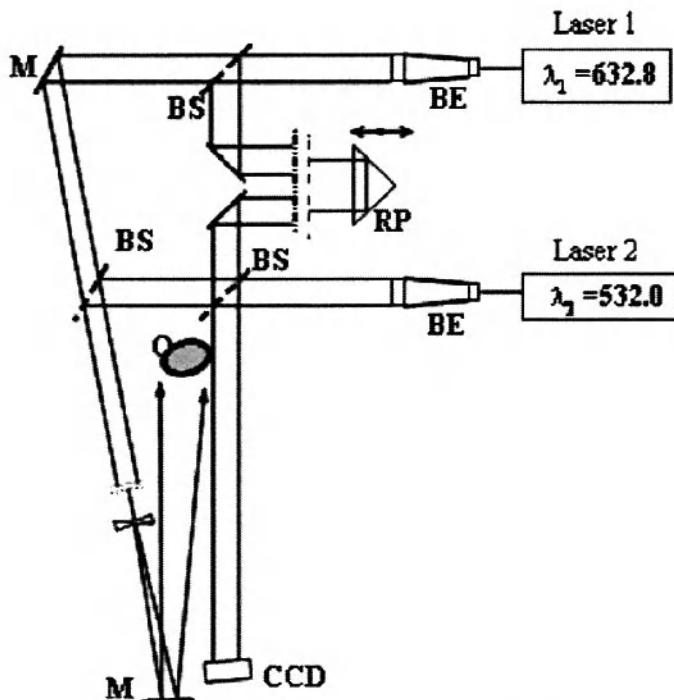


Figure 6-11. Digital Holography in reflection configuration for multi-wavelength operation

In Fig. 6-12a is shown the reconstructed amplitude image of a toy at λ_2 . In Fig. 12b is shown the amplitude image reconstruction with ZPM operation image of the toy. The holograms were initially recorded with $N_1 \times N_1 = 1024 \times 1024$ pixels. The red hologram (that at λ_2) was reconstructed in according to Eq.(6-8) adding a number of zeros around the hologram $N_2 = 1218$ pixels such that $N_2 - N_1 = (632.8/532)/1024 = 197$. In Fig. 6-12c is shown the amplitude image reconstruction at λ_1 (green hologram) of the toy. It is clear the image in this case has equal size in respect to that of Fig. 6-12b.

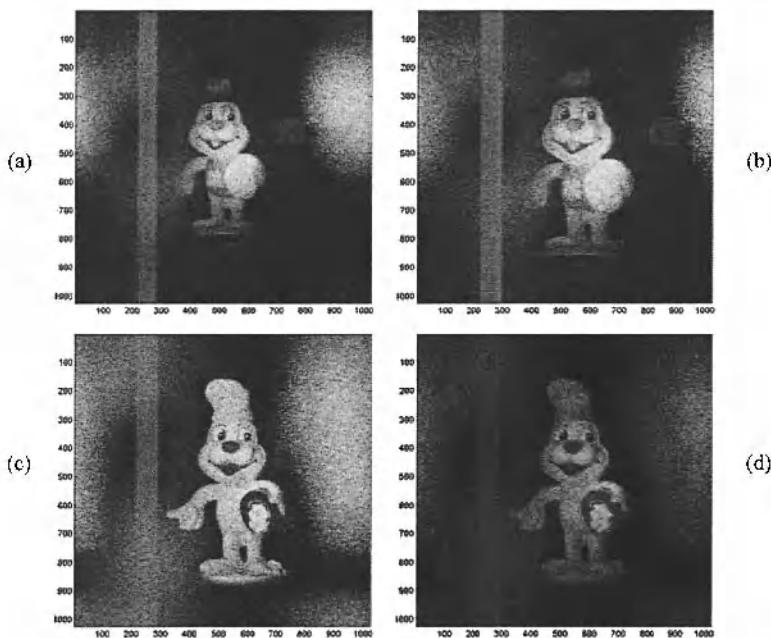


Figure 6-12. Multi-wavelength operation demonstration. Amplitude reconstruction of digital hologram recorded with $\lambda_2=632.8\text{ nm}$, (a) without padding operation and (b) with padding operation as given by Eq.(6-9); (c) reconstruction of digital hologram recorded at $\lambda_1=532\text{ nm}$; (d) RGB combination of images red and green respectively of (b) and (c)

In Fig. 6-12d is reported the RGB combination of the red and green image reconstruction with perfect superimposition and new colour (yellow) due to the mixture for red and green [20].

3.3 Controlling numerical reconstructions as function of distance and wavelength simultaneously

The very same method described in the previous sub-section 3.2 is also appropriate for multi-wavelength microscopy and/or recognition of an object in movement, since in this case, in focus numerical reconstruction of the moving object at various wavelengths results in a sequence of reconstructed images with different sizes which cannot be compared directly to each other, owing to the dependence of the reconstruction pixel (RP) on the recording distance and wavelength.

Fig. 6-13 illustrates a typical set-up for recording MWDH holograms. In this application, two lasers with different wavelengths were used, a

laser emitting in red region at $\lambda_1 = 632 \text{ nm}$ and the other in green at $\lambda_2 = 532 \text{ nm}$.

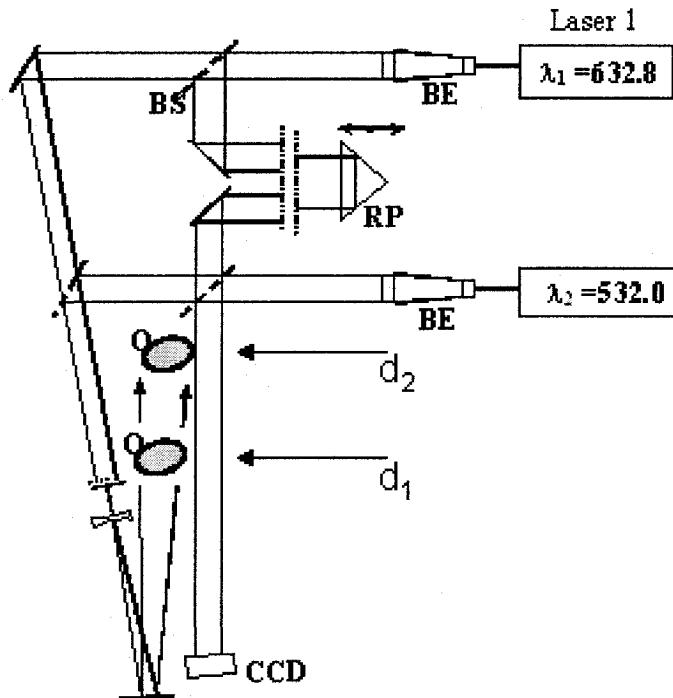


Figure 6-13. Digital holography in reflection configuration for multiwavelength operation

The optical configuration is arranged to allow the two lasers to propagate along the same paths either for the reference and object beams. The reflecting prism in the path of the red laser beam permits the matching of the optical paths of the two interfering beams inside the optical coherent length of the laser.

The object under investigation was a Pinocchio toy and four holograms of a Pinocchio toy were recorded at two wavelengths $\lambda_r = 632 \text{ nm}$ and $\lambda_g = 532 \text{ nm}$. For each wavelength the Pinocchio toy was inserted into the optical set-up at two considerable different distances $d_1 = 725 \text{ mm}$ and $d_2 = 850 \text{ mm}$ from the CCD array and the corresponding hologram was recorded.

The holograms were digitized into a square array $N \times N = 1024 \times 1024$ pixel, $6.7\mu\text{m} \times 6.7\mu\text{m}$ pixel width. The size of the reconstruction pixel at red

wavelength λ_r and reconstruction distance d_1 was $RP_{r1} = \lambda_r d_1 / N_r \Delta x = 66.8 \mu\text{m}$ with $N_r = N = 1024$; at the green line wavelength λ_g and distance $d_2 = 850 \text{ mm}$ the size of the reconstruction pixel decreases and $RP_{g2} = \lambda_g d_2 / N_g \Delta x = 65.9 \mu\text{m}$ with $N_g = N = 1024$. For the numerical reconstructions corresponding to the other two recording conditions we have $RP_{r2} = \lambda_r d_2 / N_r \Delta x = 78.3 \mu\text{m}$ and $RP_{g1} = \lambda_g d_1 / N_g \Delta x = 56.2 \mu\text{m}$.

Since RP_{r1} is greater than RP_{g2} , according to the ZPM we pad of the hologram recorded at red line wavelength λ_r and reconstruction distances d_1 with zeros up to $N_r^p = N_g (\lambda_r d_1 / \lambda_g d_2) = 1038$ allows for the same size of reconstruction pixel to be obtained for both reconstructions, namely $RP_{r1}^p = \lambda_r d_1 / N_r^p \Delta x = RP_{g2}$.

Similarly, we obtain $RP_{r2}^p = \lambda_r d_2 / N_r^p \Delta x = RP_{g1}$ by padding the hologram recorded at red line wavelength λ_r and reconstruction distance d_2 with zeros up to $N_r^p = N_g (\lambda_r d_2 / \lambda_g d_1) = 1426$. Generally, when the two holograms to be compared are recorded with the same wavelength and different distances d_1 and d_2 , the hologram recorded at a greater distance is padded with zeros. On the other side, when the holograms are recorded with same distance and different wavelength λ_1 and λ_2 , the hologram with the greater wavelength is padded. In the case of holograms recorded with different distance and different wavelengths, the hologram having the greater product $d * \lambda$ is padded. Fig. 6-14 illustrates the application of ZPM for the superimposition of the numerical reconstructions of the hologram of the Pinocchio toy recorded at distance d_1 and wavelength $\lambda_r = 632 \text{ nm}$ with that recorded at distance d_2 and wavelength $\lambda_g = 532 \text{ nm}$.

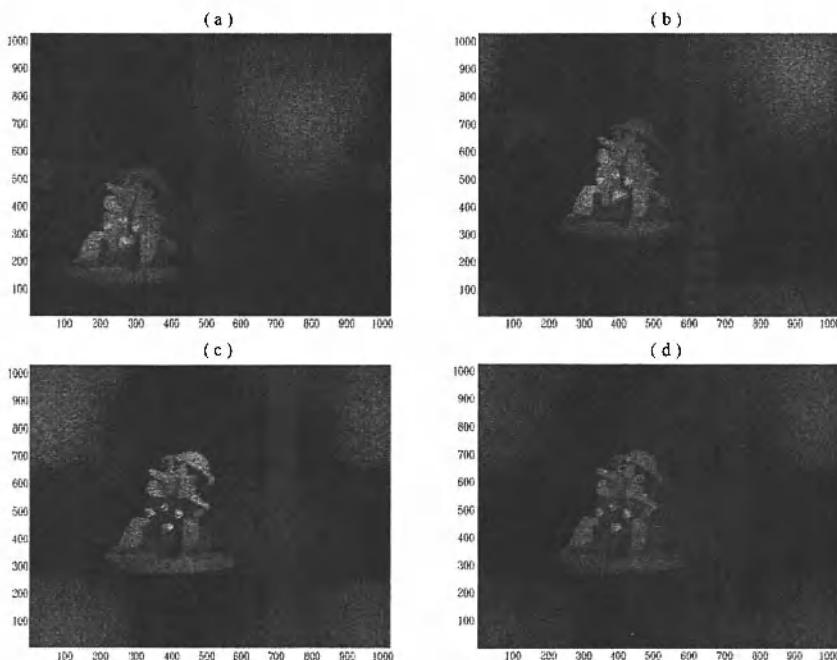


Figure 6-14. Amplitude reconstruction of holograms of Pinocchio toy. (a) Amplitude reconstruction of red hologram reconstructed at distance $d_1 = 725$ mm with $N_r = 1024$ pixel, (b) amplitude reconstruction of red hologram padded with zeros up to $N_{rp} = 1038$ pixel, (c) amplitude reconstruction of green hologram at distance $d_2 = 850$ mm with $N_g = 1024$ pixel, (d) colour superimposition of red padded reconstruction and green reconstruction

The reconstruction of the hologram of the Pinocchio toy recorded at distance d_1 and wavelength $\lambda_r = 632 \text{ nm}$ is shown in Fig. 6-14a; no padding operation was applied to the $N_r \times N_r = 1024 \times 1024$ digitized matrix.

The reconstruction shown in Fig. 6-14b was obtained by applying the FTM method to the original hologram padded with zeros up to 1038 pixels to obtain a reconstruction pixel RP_{r1}^P equal to RP_{g2} , as discussed above. Only the central portion of 1024×1024 pixels is shown in the reconstructed amplitude map. Fig. 6-14a shows the amplitude reconstruction of the hologram recorded at distance d_2 and green wavelength $\lambda_g = 532 \text{ nm}$.

The result of the superimposition between the two amplitude reconstructions of Fig. 6-14b and Fig. 6-14c is shown in Fig. 6-14d. The red + green combination of the two reconstructions gives perfect superimposition and the new yellow colour as a result of the matching of the two reconstructions pixel.

4. **CONTROLLING RECONSTRUCTIONS IN MWDH COMPENSATING CHROMATIC ABERRATIONS**

It has previously examined how DH can be used to compensate several kinds of aberrations and presented examples of application of the technique for metrological applications such as MEMS inspection. DH is also very suitable for measuring optical phase retardation in phase objects.

Of course, complete compensation of aberrations plays a crucial role when quantitative phase information is required and this is comes to be true in MWDH for microscopic applications, where further complications arise, owing to the need of removing the defocusing and chromatic aberrations of the wavefront emerging from the microscope objective, as well as imperfect achromaticity of the other optical elements, such as wave-plates, polarizers and beam-splitters.

Speckle noise is also a limitation in multiwavelength phase imaging-based applications, such as DH phase imaging technique to remove 2π ambiguity, where reconstructed phase maps at wavelength λ_1 and λ_2 are subtracted to obtain a new phase map equivalent to that of a longer beat wave $\lambda_{12} = \lambda_1\lambda_2/|\lambda_1 - \lambda_2|$ [21].

A direct subtraction of two single-wavelength phase maps cannot be made, since the corresponding RP are different.

It is possible to combine the flexibility offered by DH for compensating defocusing and chromatic aberrations with the capability of controlling the size of the RP for each wavelength by ZPM. To illustrate the method it is considered the off-axis set-up for recording MWDH hologram shown in Fig. 6-15.

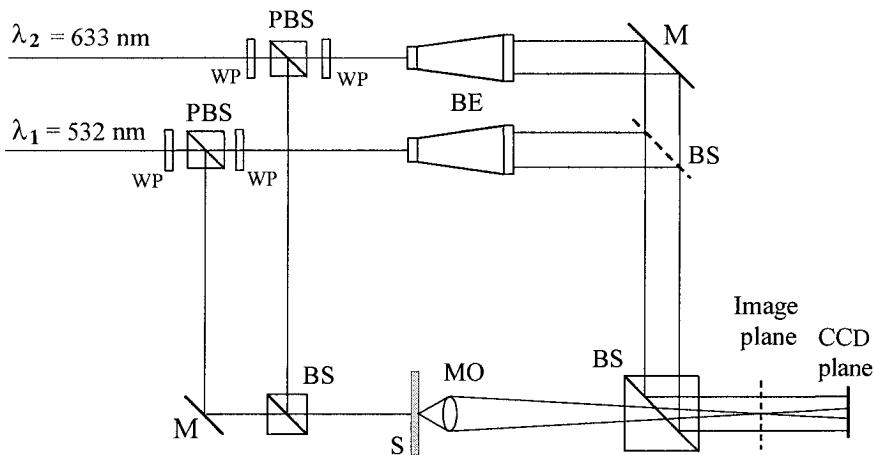


Figure 6-15. Off-axis set-up for recording MWDH holograms for phase objects

The off-axis geometry allows to avoid the problem of the twin image and to eliminate the zero diffraction order. Two lasers with different wavelengths were used, a laser emitting in green region at $\lambda_1 = 532 \text{ nm}$ and the other in red region at $\lambda_2 = 632.8 \text{ nm}$. The optical configuration is arranged to allow the two lasers to propagate almost along the same paths either for the reference and object beams. The reference and object beams are plane wave fronts obtained by a beam expander (BE). The first beam splitter is a cube polarizing beam splitter (BS) and a $\lambda/2$ wave plate is inserted in the reference beam to obtain equal polarization direction for the two beams. The microscope objective is an aspheric lens with focal length $f = 9 \text{ mm}$ and $N.A. = 0.4$. The phase object was a photoresist pattern recorded on a glass substrate with $20\mu\text{m}$ wide resist stripes. MWDH was used to measure the shape and depth of about $1.3\mu\text{m}$ of recorded stripes

The pattern was set at distance $d = 110 \text{ mm}$ from the CCD array. The CCD array has 1024×1024 square pixels, each having a $\Delta\xi = 6.7 \mu\text{m}$ side.

Fig. 6-16a shows digital off-axis hologram of the phase object recorded at the red wavelength $\lambda_2 = 623 \text{ nm}$; Fig. 6-16b shows amplitude reconstruction by FTM of the digital hologram in (a) after application of ZPM, to obtain a phase image with size equal to that of the reconstructed hologram at the green wavelength $\lambda_1 = 532 \text{ nm}$ and to compensate for the different size of the RP.

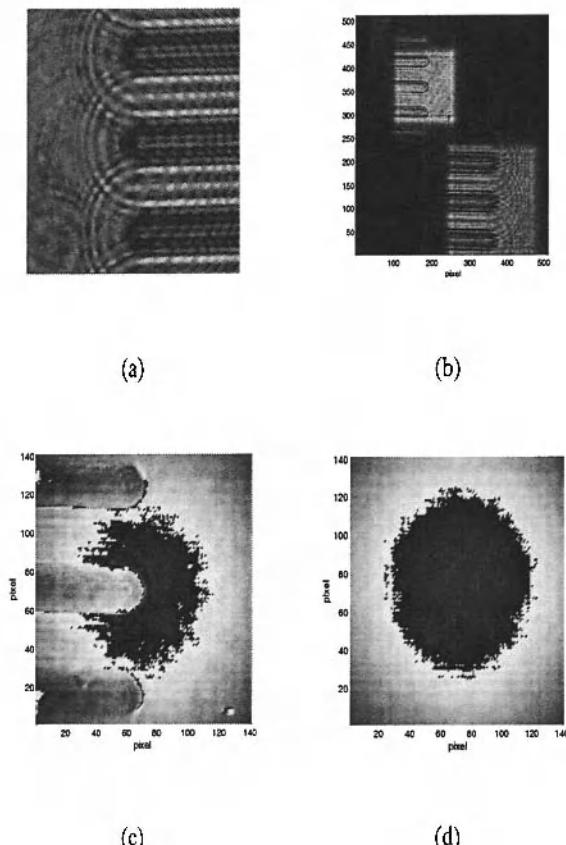


Figure 6-16. (a). Digital off-axis hologram recorded @ 632nm; (b) Amplitude reconstruction of the digital hologram in (a) after padding operation; (c) Wrapped phase map in presence of chromatic aberration at coarse equivalent wavelength;(d) Wrapped phase map of chromatic aberration obtained by subtracting the two reference respectively red and green holograms

Fig. 6-16c shows the map of the optical phase retardation of the object obtained subtracting each other the reconstructed phase images at wavelength λ_1 and λ_2 by ZPM.

The result is a new phase map equivalent to that of a longer beat wave $\lambda_{12} = \lambda_1\lambda_2/|\lambda_1 - \lambda_2| = 3.34\mu m$.

In Fig. 6-16c it is clearly visible that the phase map is affected by a large spurious circular fringe that perturbs the phase retardation map. The perturbing phase is due to the chromatic aberration introduced by the optical imaging system in the object beam of the interferometer.

With the aim to remove chromatic aberration, two holograms, the reference holograms, were recorded without the test object for each wavelength λ_1 and λ_2 with $N_1 = 1024$.

Reconstructed wrapped phase map for the red hologram $\phi_R^2(x, y)$ and the phase map of the reference green hologram $\phi_G^1(x, y)$ allow to obtain a sort of calibration of the holographic set-up to compensate for the chromatic aberration once upon a time.

From the reconstructed complex amplitudes $R_1(x, y)$ and $R_2(x, y)$ of the two reference holograms, the relative phase $\phi_R^{(12)}(x, y) = \text{Arg}(R_1(x, y)/R_2(x, y)) = \phi_G^1(x, y) - \phi_R^2(x, y)$ between the two complex wave fields is computed and show in Fig. 6-16d. We can observe circular-like fringes caused mainly by residual chromatic aberration in the microscope objective and the recombining beam-splitter cube and unavoidable slight misalignment of the two reference beams. The reference phase map of Fig. 6-16d can be used to subtract the disturbing aberration and to obtain a correct phase map at the beat wavelength of the phase object under investigation [22]. Fig. 6-17a and Fig. 6-17b show the wrapped phase images $\phi_O^1(x, y) = \phi_1(x, y) - \phi_G^1(x, y)$ and $\phi_O^2(x, y) = \phi_2(x, y) - \phi_R^2(x, y)$ reconstructed at distance $d = 110$ mm from the red hologram and the green hologram, respectively, with the target inserted into the object beam of the interferometer.

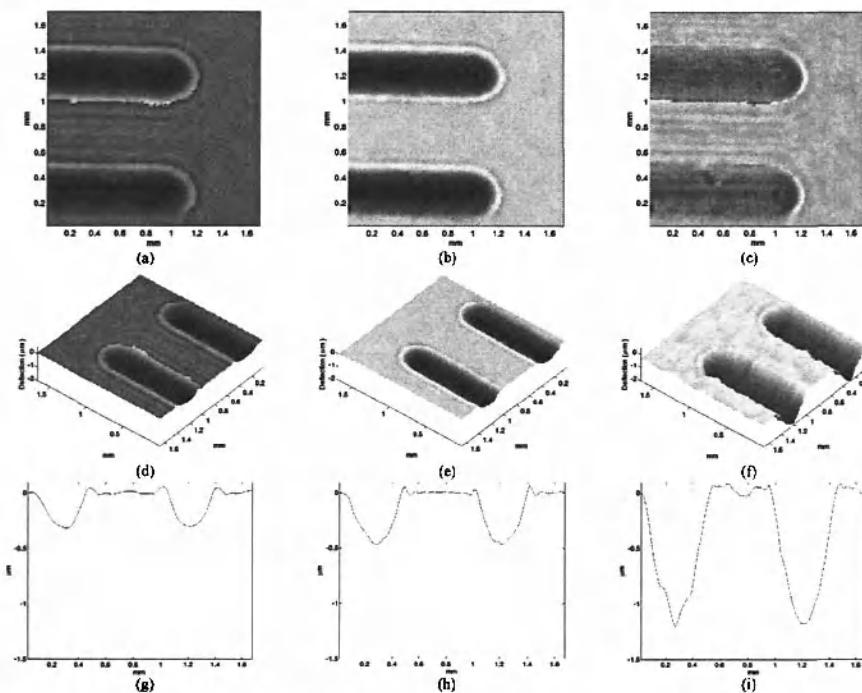


Figure 6-17. (a). Unwrapped phase map using the only red wavelength; (b) unwrapped phase map using the only green wavelength; (c) Unwrapped phase map using the equivalent wavelength; (d) 3D depth map using the only red wavelength; (e) 3D depth map using the only green wavelength; (f). 3D depth map using the equivalent wavelength; (g). detail of depth profile along a single line across the map using only red wavelength; (h). detail of depth profile along a single line across the map using only green wavelength; (i) Detail of depth profile along a single line across the map using beat equivalent wavelength.

The red hologram was reconstructed after application of ZPM, as explained before. Since reconstruction with two wavelengths produces in general different spatial frequency for each case, the two phase images were centered by adjusting numerically the reference beam and were directly subtracted from each other, to generate (Fig. 6-17c) the difference phase map

$$\phi_O^1(x, y) - \phi_O^2(x, y)$$

Numerous discontinuities, caused by the noise of the individual phase maps, can be clearly observable in the difference phase map. The difference phase map $\Delta\varphi(x, y)$ can be evaluated according to the following equation

$$\Delta\varphi(x, y) = \text{Arg}(\phi_1(x, y)R_2(x, y)/\phi_2(x, y)R_1(x, y)) \quad (6.10)$$

and Eq.(6-10) can be written as $\Delta\varphi(x, y) = \phi_1(x, y) - \phi_2(x, y) - \phi_R^{(12)}$ to make it clear that we are actually accounting for the chromatic aberration by subtracting the corresponding phase contribution $\phi_R^{(12)}$ from the resized difference phase $\phi_1(x, y) - \phi_2(x, y)$ and the result is shown in Fig. 6-17c.

Fig. 6-17d and Fig. 6-17e show 3D plots of the depth profile calculated from the wrapped phase maps of Fig. 6-17a and Fig. 6-17b, respectively, by using standard unwrapping procedure.

Usually, the unwrapping procedure removes 2π ambiguities between the adjacent pixels in the wrapped phase map, but it is unable to remove phase discontinuities when the depth of the object is greater than the wavelength. Nevertheless, in this case the depth profile of the resist layer is greater than both the red and the green wavelength and the phase profile is incorrectly recovered. Fig. 6-17f shows the depth map corresponding to the equivalent beat wavelength λ_{12} . The optical thickness $z(x, y) = \lambda_{12}\Delta\varphi(x, y)/2\pi$ of the resist is computed by the wrapped difference phase map $\Delta\varphi(x, y)$, according to the resist thickness.

The longer beat wavelength of the aberration-compensated “coarse” phase map $\Delta\varphi(x, y)$ allows for substantial removal of many phase discontinuities and 2π ambiguities which are present in the single-wavelength phase maps. The profile shown in Fig. 6-17i gives the depth variation along a line of the 3D depth map shown in Fig. 6-17f to be compared with the incorrect profiles shown in Fig. 6-17g and Fig. 6-17h, that have been calculated by means of single wavelength phase maps. This example shows clearly that the ability to control the reconstruction pixel independently of the wavelength and fully exploit the flexibility offered by digital holography for numerical compensation of chromatic aberration allowed to retrieve correctly the depth profile of the resist pattern

5. APPLICATION OF DH FOR INVESTIGATION AND TESTING OF MEMS STRUCTURES

In today’s global marketplace to increase productivity, companies seek to lower overhead, such as labor and research and development costs, while increasing high-quality yields. The combined result is increased profitability. One of threads that ties together labor, R&D, production yields, and profitability, is test and measurement. Automated metrology reduces labor costs for quality inspection and enables the manufacture of complex systems like electronics and optical systems, where tolerances exceed human sensory abilities. Automated metrology turns the one-off experiments of an R&D

laboratory into efficient batch operations, while similar metrology systems feed production data back into manufacturing lines for real-time yield improvements. The Digital Holography Microscopy described in the previous section seems to have all the features to be a fully automated inspection technique. In fact, the reconstruction method based on the FTM allows the controlling of different parameters and the variation of these parameters can be optimized by a software procedure in order to achieve the best image reconstruction of the object under inspection. Moreover, by means of the controlling parameters procedures it is possible to obtain most of information about the characterized object by a single recorder image. In other words, thanks to the unique property of DH to make the phase information of the reconstructed wave front available, the 3D shape of the object under investigation can be obtained by a single hologram [23].

If the sample experiences even a very small displacement along the optical axis, a very large change occurs in the distance to the imaging plane and the focus can be lost. The tedious search for new focal planes can be intolerable, especially if there is the need to visualize the phenomena in anything approximating real-time.

In next sections the methods to overcome these problems are described; obviously the methods are based on the possibility to control the parameters of the reconstruction procedure based on the FTM.

5.1 Automatic Focus Tracking

The experimental setup illustrated in Fig. 6-18 allows the possibility to employ different microscope objective in order to obtain images of sample object with different magnifications. However, a higher magnification implies shorter focal length. So, if the sample experiences even a very small displacement along the optical axis, a very large change occurs in the distance to the imaging plane and the focus can be lost as depicted in Fig. 6-18.

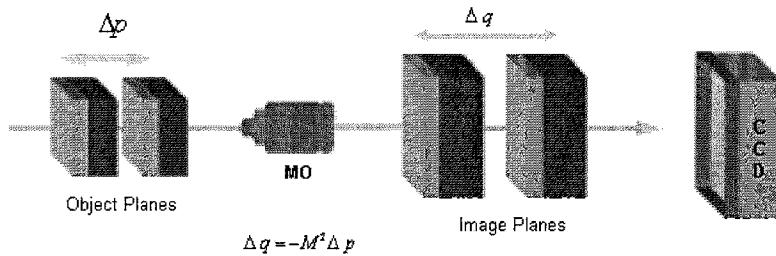


Figure 6-18. Out-of-focus caused by a longitudinal displacement of the imaged object

If the distance from the lens to the object is p , q is the distance of the image plane from the lens, then any axial displacement Δp of the object results in a translation of the imaging plane in front of the CCD given by:

$$\Delta q = -M^2 \Delta p \quad (6-11)$$

where M is the magnification of the imaging system. For example, with an imaging system of a magnification $M = 50$, a sample displacement of $1\mu m$, translates the image plane by $\Delta q=2.5mm$. Thus, an unpredictable displacement of the object entails a displacement of the image plane and the focus can be lost. The effect of defocus influences the phase-contrast reconstruction, thereby affecting the quantitative information. In order to recover the right information, the distance d used in the numerical reconstruction method must be changed to get in-focus amplitude and phase-contrast image. Displacement of the object may occur for different reasons and it can not be often neglected; for example in thermal characterization of objects. Temperature changes can cause unpredictable expansion or contraction of the object under study and/or its mechanical support, and the tedious search for new focal planes can be intolerable, especially if there is the need to visualize the phenomena in anything approximating real-time.

The research of the new reconstruction distance can be simplified and automated detecting the axial displacement of the object by measuring the phase-shift of the hologram fringes [24]. In fact, with reference to the configuration reported in Fig. 6-18, axial displacement of the object causes a shift in the fringe pattern of the hologram at time t given by:

$$\Delta\varphi(t) = 4\pi \frac{\Delta p}{\lambda} \quad (6-12)$$

By recording the phase-shift in a small flat portion of the object (in order to speed up the elaboration time), it was possible to determine the

displacement and so the incremental change of the distance to be used in the numerical reconstruction for each recorded hologram. In particular the updating at time t of the reconstruction distance $d(t)$ is given by:

$$d(t) = d_0 + M^2 \lambda \frac{\Delta\phi(t)}{4\pi} \quad (6-13)$$

where d_0 is the initial distance between the sample and the object plane.

By means of the described method, we have obtained the reconstruction images illustrated in Fig. 6-19.

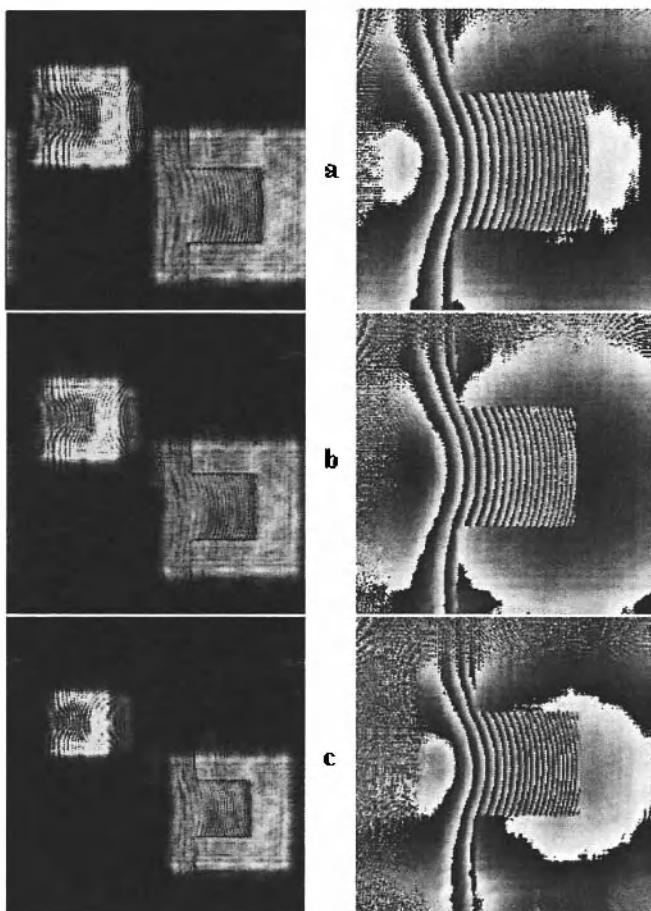


Figure 6-19. In focus amplitude and phase-map for the cantilever beam from three holograms recorded at three different distances (a) $d=100\text{mm}$; (b) $d=117.3\text{mm}$; (c) $d=140.8\text{mm}$

In fact, the reconstruction distance variation, due to the overall thermal expansion of the silicon, the metallic plate and the translation stage, was evaluated by measuring the average intensity change in a group of 4x4 pixels (see Fig. 6-20).

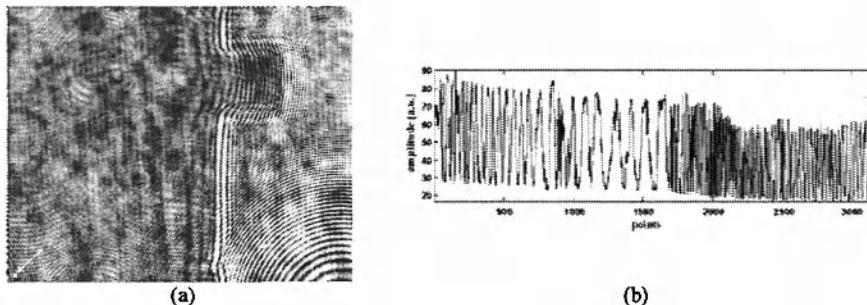


Figure 6-20. (a) hologram of the MEMS with indication of the point in which the phase shift signal is recorded to track the longitudinal displacement; (b) Intensity phase shift signal measured at a group of 4x4 pixels

This signal was analyzed by applying a Fourier-transform method, and for each point recorded the wrapped and unwrapped phase-shift $\Delta\phi(t)$ of the fringes was calculated. Then, this value is employed to upgrade the reconstruction distance by means of Eq. 6-13. The reconstruction distance variation is shown in Fig. 6-21 as a function of the sampled point.

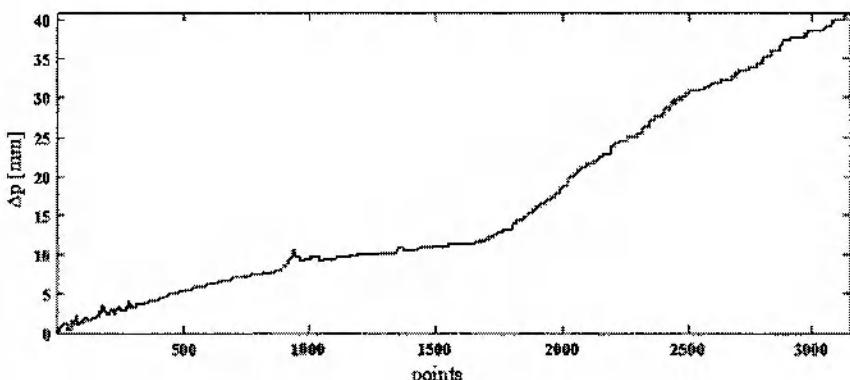


Figure 6-21). Displacement of the sample measured in real time by analyzing the phase shift of hologram fringes

As shown, the final hologram reconstruction distance differs from the initial distance by about 40mm. By means this signal is possible to obtain both the amplitude and phase reconstructions of the cantilever always in-focus as illustrated in Fig. 6-19.

5.2 ZPM for subtracting reconstructed phase maps

As shown above DH is an ideal technique for retrieving the phase distribution of the object wave field for quantitative phase contrast imaging in microscopy, meaning that the reconstructed phase distribution that cannot be observed in optical reconstruction of film holography, can be easily computed and displayed quantitatively.

For this reason DH offers interesting prospects and advantages in different applications, such as in the characterization of microcomponents, microelectromechanical systems (MEMS) [23].

To use DH for quantitative microscopy, reliable methods need to be developed for accurate determination and comparison of sequences of reconstructed amplitude and phase images, and the dependence of the reconstruction pixel on the recording distance has to be taken into account when comparing two phase images reconstructed by FTM.

As described in the previous paragraph DH has been applied in a quasi-real-time inspection of silicon MEMS structures subjected to a thermal load. amplitude and phase images of the cantilever. There the automatic procedure for taking into account the axial displacement of the sample has been discussed. Here we want to show the effectiveness of the ZPM to account of the dependence of the size of the phase images on the reconstruction distance for all the holograms of the sequence. Fig. 6-19 shows three amplitude(left column) and phase (right column) reconstructions of the cantilever beam at different distances Fig. 6-19a shows the reconstruction of the first hologram at $d=100$ mm; Fig. 6-19b shows that of the second hologram at $d=117.3$ mm while Fig. 6-19c shows that of the third hologram at $d=140.8$ mm. The amplitude and phase reconstructions in Fig. 6-19 are all in focus but it can be noted that the size of the images changes with reconstruction distance since the three holograms were recorded and reconstructed at different distances. In order to obtain resized amplitude and phase reconstructions the holograms reconstructed at distances $d_1 = 117.3$ mm and $d_2 = 140.8$ mm have been padded with zeros up to 614x614 pixels and 718 x 718 pixels, respectively by using Eq. (6-8).

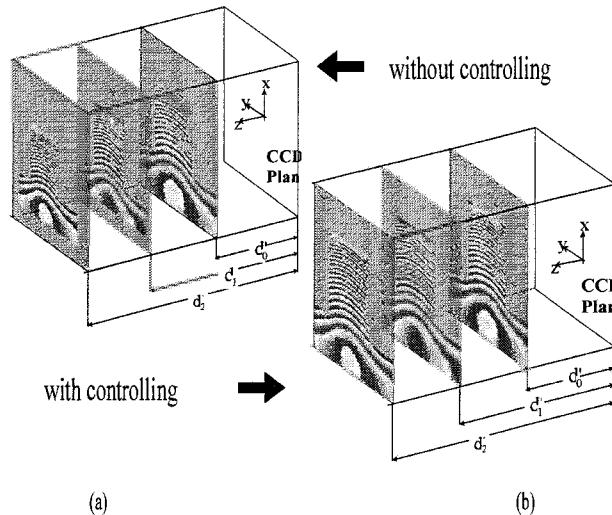


Figure 6-22. Wrapped image phases reconstructed at different distances without (a) and with (b) application of padding operation]

Fig. 6-22a shows the wrapped phase map without application of ZPM. The reconstructed phase images after application of ZPM are shown in Fig. 6-22b. The phase image size is now independent on the reconstruction distance, so that direct phase subtraction can be performed and possible difference phase differences corresponding to different thermal states of the cantilever can be emphasized. Comparison between reconstructed phase maps after application of ZPM is shown in Fig. 6-23. Fig. 6-23a is the unwrapped phase image of the hologram recorded at distance d_1 ; Fig. 6-23b is the unwrapped phase image of the MEMS at $d_2=140.8\text{ mm}$ without ZPM; Fig. 6-23c is the unwrapped phase of the MEMS at d_2 obtained by ZPM. Fig. 6-23d shows the calculated difference between the resized unwrapped phase maps indicating the small deformation caused by thermal load.

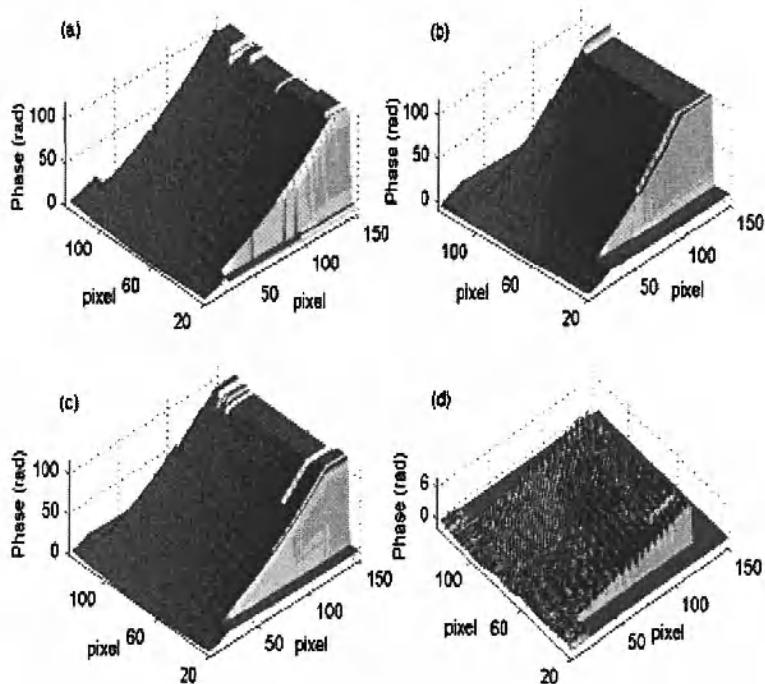


Figure 6-23. Unwrapped phase of MEMS at different distance with padding operation and phase map subtraction between reconstructions at different distance

5.3 DHM for real-time inspection

The focus tracking procedure described above allows to obtain both amplitude and phase reconstruction in focus updating the reconstruction distance. However, although well-focused images were obtained, it was not possible to compare two of them directly since they had different sizes owing to the different width of the reconstruction pixel which depends on the reconstruction distance. This problem can be overcome as mentioned in a previous section by means of ZPM. In this way the size of the reconstructed image is independent of the reconstruction distance so that on these reconstructed images direct phase subtraction can be performed and phase difference can be obtained. This approach is instrumental if an anything approximating real-time inspection of an object has to be carried out.

As example of this approach, the quasi real-time inspection of a Silicon micro-heater is reported. The micro-heater has been realized integrating an heater resistors on top of micro-machined suspended membranes. By means of DHM it is possible to analyze the behavior of the micro-hot plate during

its operation mode. In fact, when the temperature of the micro-heater increases, a warpage of the membrane is induced and the investigation of this behaviour can be carried out by the method above-described. In particular, the dynamic analysis is possible by tracking the focus change induced by unwanted longitudinal displacement occurring during the test and simultaneously controlling the phase-map image size in the reconstruction process. Thus, the inspection has been performed by a continuous recording of the structure hologram and referring each hologram to a reference hologram (i.e. the hologram recorded at room temperature). This subtraction allows on one hand to evaluate the variation with respect to the reference position; on the other hand employing the reference hologram the curvature phase compensation is automatically carried out, because this aberration is the same for each state of the object. In this way a continuous sequence of the quantitative 3D full-field deformation experienced by the micro-heater, can be easily obtained. A typical phase map reconstructed and its respective 3D full-field of the deformation, corresponding to two different temperatures of the operating conditions of the micro-heater, has been extracted from the sequence and illustrated in Fig. 6-24.

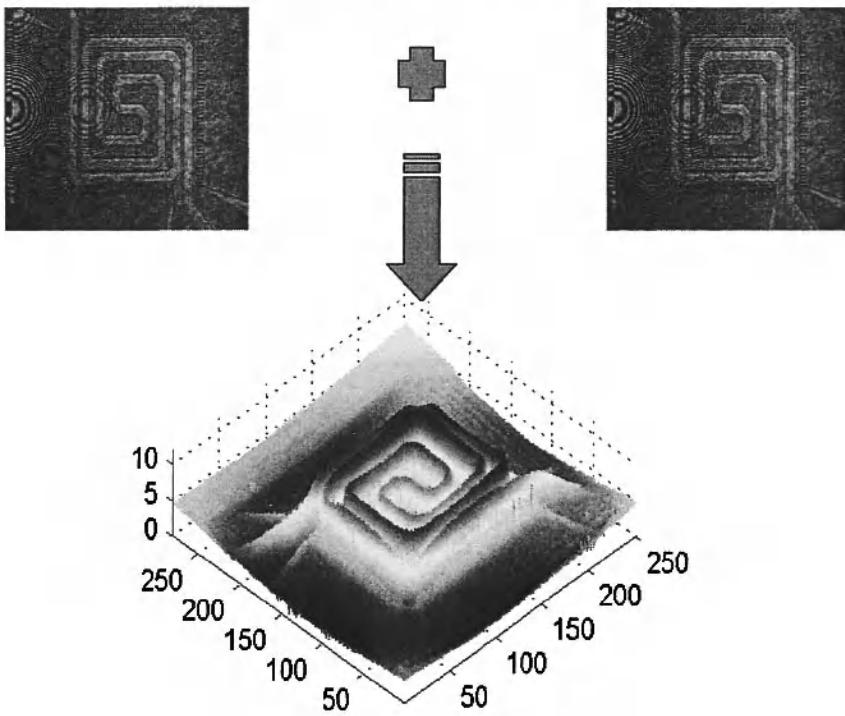


Figure 6-24. Phase map reconstruction in 3D full-field of the deformation obtained by two holograms recorded at two different temperatures of operating conditions of the micro-heater ($\Delta T = 700 \text{ }^{\circ}\text{C}$). Axis unit length in microns

5.4 Recovering of the resolution by ZPM

In Digital holographic microscopy it is usual to put the object in the Fresnel region of the image sensor, because use of a single Fourier transform for numerical reconstruction reduces the calculation time by half. However, this limits resolution because when light passes through an object that contains high-spatial frequency components, the diffraction occurs at large angles. This limitation is particularly severe when high resolution with visible wavelength light is desired. In fact, elementary pixel width of CCD or CMOS sensors is at least of several microns, leading to spatial cut-off frequencies lower than 100 mm^{-1} . Consequently, standard CCD/CMOS sensors limit incident beams to being quasi-parallel so that only small objects placed at a large distance from the sensor can be recorded with digital holography [25, 26]. Obviously, if paraxial approximation is satisfied,

the FTM can be used in the reconstruction process. However, when the image reconstruction is performed by FTM, the spatial frequencies displayable in reconstructed images are band limited by the size of the reconstruction pixel, which represents the sampling gauge in the image plane. From Eq. (6-2) it is clear that the reconstruction pixel, and consequently resolution, depends on the wavelength, the distance, number of the pixels N and M of the sensor array, and their physical size (Δx). In other words, in the image reconstruction plane spatial frequencies higher than the Nyquist limit are under-sampled. Depending on the objects, under-sampling can affect the correctness of reconstructed phase map. Thus, in order to improve the resolution, the sampling gauge in the image plane (i.e. the reconstruction pixel) has to be decreased. This decreasing can be carried out by the ZPM, that is through artificial enlargement of the number of the pixels in recorded digital holograms. In particular, the dimension of the padding is set to make the minimum phase increment that can be resolved less than π .

As example of this approach, the characterization of MEMS structure that grows too rapidly is reported (see Fig. 6-25).

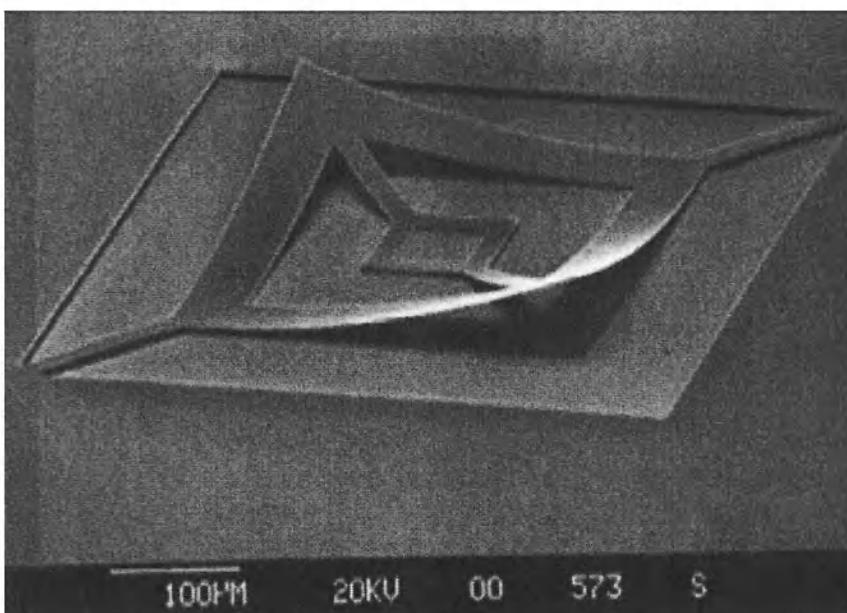


Figure 6-25. SEM picture of a MEMS with a large amount of deformation

In particular, the ringed corner of the inspected MEMS has a so warped shape that induces an undersampling for a fixed distance, wavelength, number and size of pixels of the CCD.

In order to recover the lost resolution the same hologram employed to reconstruct the field shown in Fig. 6-26a has been padded with zeros up to 1024x1024. In Fig. 6-26b is shown the 3D full-field deformation from the padded hologram. Although the obtained shape is consistent with the expected deformation, a small under sampling at the extremity of the corner still occurs .To carry out the complete correct profile of the corner of the MEMS, the original hologram has padded again up to 2048x2048, and the resulting 3D full-field is shown in Fig. 6-26c. In this case no under sampling are present and the MEMS shape is correctly reconstructed.

It is important to note that real content of information in Fig. 6-26c is exactly the same as that obtained reconstructing the MEMS shape from an original hologram of 2048x2048 pixel [27]. That means the required information about phase map can be extracted even from a reduced hologram.

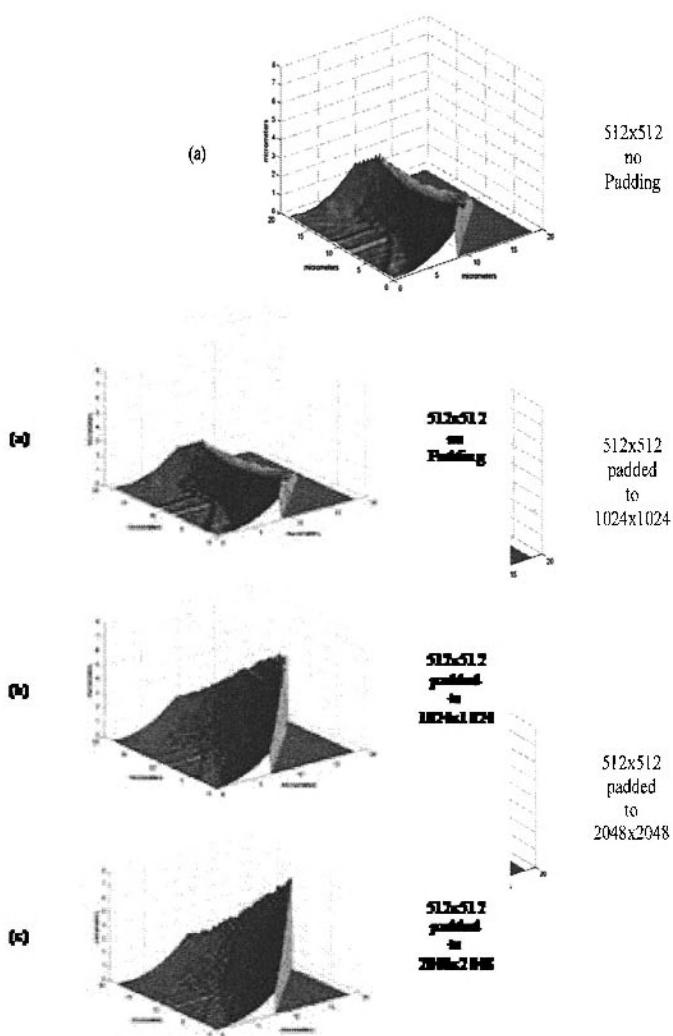


Figure 6-26. Profile of the MEMS obtained by: (a) central portion of 512x512 pixel hologram; (b) same hologram of (a) but padded with zeros up to 1024x1024pixels; (c) padded up to 2048x2048 pixels

REFERENCES

1. U. Schnars, W. Juptner, Digital recording and numerical reconstruction of holograms, *Meas. Sci. Technol.*, 13, R85-R101, (2002).
2. T.M. Kreis, W. Jüptner, Principles of Digital Holography. In: Jüptner, Osten, editors. *Fringe 97*, Academic Verlag, 253–363, (1997).
3. I. Yamaguchi, T. Matsumura, and J. Kato, Phase-shifting colour digital holography, *Opt. Lett.* 27, 1108-1110, (2002).
4. O. Matoba and B. Javidi, Optical retrieval of encrypted digital holograms for secure real-time display, *Opt. Lett.*, 27, 321-323, (2002).
5. E. Cuche, F. Bevilacqua, and C. Depeursinge, Digital Holography for quantitative phase-contrast imaging, *Opt. Lett.* 24, 291-293 (1999).
6. E. Cuche, P. Marquet, and C. Depeursinge, Simultaneous amplitude-contrast and quantitative phase-contrast microscopy by numerical reconstruction of Fresnel off-axis holograms, *Appl. Opt.* 38, 6994-7001, (1999).
7. P. Ferraro, S. DeNicola, A. Finizio, G. Coppola, S. Grilli, C. Magro, G. Pierattini, Compensation of the inherent wave front curvature in digital holographic coherent microscopy for quantitative phase contrast imaging, *Appl. Opt.*, 42, 1936-1946, (2003)
8. P. Ferraro, S. De Nicola, A. Finizio, and G. Pierattini, "Reflective grating interferometer in a non collimated configuration", *Appl. Opt.* 39, 2116-2121 (2000).
9. S. De Nicola, P. Ferraro, A. Finizio, and G. Pierattini, Correct-image reconstruction in the presence of severe anamorphism by means of digital holography, *Opt. Lett.* 26, 974-977 (2001)
10. S. De Nicola, P. Ferraro, A. Finizio and G. Pierattini, "Reflective grating interferometer: a folded reversal wave-front interferometer," *Appl. Opt.* 38, 4845-4849 (1999).
11. A. Stadelmaier and J. H. Massig, "Compensation of lens aberrations in digital holography," *Opt. Lett.*, 25, 1630-1633 (2000).
12. G. Pedrini, S. Schedin, and H. J. Tiziani, "Aberration compensation in digital holographic reconstruction of microscopic objects," *J. Mod. Opt.* 48, 1035-1041 (2001).
13. S. Grilli, P. Ferraro, S. De Nicola, A. Finizio, G. Pierattini, R. Meucci Whole optical wavefields reconstruction by digital holography, *Opt. Exp.* 9, 294-302 (2001)
14. T. M. Kreis, Frequency analysis of digital holography, *Opt. Eng.* 41, 771-778, (2002)
15. T. M. Kreis, Frequency analysis of digital holography with reconstruction by convolution, *Opt. Eng.* 41, 1829-1839(2002)
16. P. Ferraro, G. Coppola, S. DeNicola, A. Finizio, G. Pierattini, and D. Alfieri, Controlling image size as a function of distance and wavelength in Fresnel transform reconstruction of digital holograms, *Opt. Lett.* 29, 854-856 (2004)
17. S. DeNicola, P. Ferraro, G. Coppola, A. Finizio, G. Pierattini, and S. Grilli, Talbot self-image effect in digital holography and its application to spectrometry, *Opt. Lett.* 29, 104-106 (2004)
18. M. Kim, Tomographic three-dimensional imaging of a biological specimen using wavelength-scanning digital interference holography, *Opt. Exp.* 7, 305-310 (2000)
19. N. Demoli, D. Vukicevic, and M. Torzynski, Dynamic digital holographic interferometry with three wavelengths, *Opt. Express* 11, 767-774, (2003)
20. P. Ferraro, G. Coppola, D. Alfieri, S. De Nicola, A. Finizio, and G. Pierattini, Recent advancements in digital holographic microscopy and its applications in "Optical Metrology in Production Engineering", vol. 5457 of "SPIE Proceedings" W. Osten and M. Takeda eds., , pp. 481-491 (2004)
21. P.J. de Groot, Extending the unambiguous range of two-color interferometers, *Appl. Opt.* 33, 5948-5950 (1994).
22. S. De Nicola, A. Finizio, G. Pierattini , D. Alfieri, S. Grilli, L. Sansone, P. Ferraro "Recovering correct phase information in multiwavelengths digital holographic microscopy by compensation of chromatic aberrations", in press *Opt. Lett.* (2005).
23. G. Coppola, P. Ferraro, M. Iodice, S. De Nicola, A. Finizio, and S. Grilli, "A digital holographic microscope for complete characterization of microelectromechanical systems", *Meas. Sci. and Tech.* 15, 529-539 (2004).
- 24 P.Ferraro, G. Coppola, S. DeNicola, A.Finizio, G. Pierattini, "Digital holographic microscope with automatic focus tracking by detecting sample displacement in real time", *Opt. Lett.*, 28, 1257-1259 (2003).
- 25 J. H. Massig, Digital off-axis holography with a synthetic aperture, *Opt. Lett.* 27, 2179-2181 (2002)

- 26 C. Liu, Z. Liu, F. Bo, Y. Wang, and J. Zhu, Super-resolution digital holographic imaging method, *Appl. Phys. Lett.* 81, 3143-3145 (2002)
- 27 P. Ferraro, S. DeNicola, A. Finizio, G. Pierattini, and G. Coppola, Recovering image resolution in reconstructing digital off-axis holograms by Fresnel-transform method, *Appl. Phys. Lett.* 85, 2709-20711(2004)

Chapter 7

TELE-METROLOGY BASED ON DIGITAL HOLOGRAPHY

Thomas Kreis¹, Werner Jüptner¹, Ulf Schnars²

¹*BIAS - Bremer Institut für Angewandte Strahltechnik, 28359 Bremen
Airbus, 28195 Bremen*

Abstract: Digital holography is a versatile tool for remote science or tele-metrology. The data of the hologram, stored in the computer as a matrix of intensity values, can be transferred by electronic transmission lines to any other point. The hologram itself can be reconstructed optically and serve as the reference in coherent-optical measurements. Furthermore the acquired holograms of a remote measurement can be transferred to the evaluation place in the same way. However, it is also possible to transfer the pre-evaluated data, which are in common Interferograms. In all cases a high amount of data has to be transferred requiring a significant compression of the data. Different methods are compared.

Key words: Digital holography, tele-metrology, data compression, remote control, comparative holography

1. INTRODUCTION

Interferometric methods are characterized by their high resolution and accuracy in measuring the optical path difference between two states of an object. After the invention of the laser a lot of restrictions could be overcome, e.g. the limitations by the short coherent length in white light interferometry. Furthermore, holographic interferometry enabled the comparison of two states of a complex object with a rough surface at two different times but still the restriction of location remained. *Digital Holographic Interferometry* including the related methods *Digital Speckle Interferometry*, *Digital Shearing Interferometry* and *Digital Speckle Photography* as

computer based methods open a new window of coherent-optical metrology namely tele-metrology [1]: Interferometric experiments may be performed in harsh and/or remote environments like underground geological measurements, airborne tests, underwater experiments, etc. In this case it is favorable and often necessary to disconnect the data acquisition unit from the data processing and evaluation equipment. A prominent example is the Fluid Science Laboratory (FSL) on board of the International Space Station (ISS), where fluid science experiments shall be performed comprising interferometric measurements with the different methods mentioned before [2].

Stored holographic data may also be captured in one place, transmitted electronically and used at another place for measurements. One example for this way of tele-metrology is the shape control by comparative Digital Holography. In this case the hologram of the master state (object) is reconstructed optically, and the reconstructed object wave is applied to the interferometer.

The electronic imaging used in the interferometric methods results in high digital data rates, which must be recorded and transmitted in telescience experiments: A data acquisition based on a CCD target with only 1k x 1k pixels, a pixel data depth of 8 bits and a frame rate of 30 fps results in a data rate of 240 Mbps. For the FSL example – serving as an example for telescience - the communication links between space and earth form a serious bottleneck. The standard downlink capacity given for this double rack is limited to 2Mbps, which for short periods can be raised to 6 – 8 Mbps on the basis of a requested extra service. The solution of today is the data storage and off-line evaluation after the return to earth and/or the transmission of selected data during the flight.

In all cases of tele-metrology it is necessary to transmit the acquired data in nearly real-time to another place. This needs a combination of data reduction and data compression/decompression, in common. While the data reduction may be obtained by lowering the frame rate, or the quantization depth, or the selection of an area of interest, a high potential is in compression. The achievable compression factor depends on the image content and on the compression method. Therefore, compression is a significant part of tele-metrology.

2. DIGITAL HOLOGRAPHY AND RELATED METHODS

Holography is a method to store image data including intensity and phase by the superposition of an object wave – reflected by an opaque surface or refracted by a transparent object – with a reference wave [3]. The resulting

micro-interference is stored by a light sensitive medium, which is in Digital Holography and related methods a CCD target or CMOS array. The acquired data are digitized into a number of discrete pixels, quantized into a finite number of gray-values, and stored in a computer [4, 5]. The reconstruction and evaluation of the wave fields is performed numerically in the computer, where in most of the cases two reconstructed wave fields are compared interferometrically to produce the desired measurement result [6]. In most fluid science experiments these results indicate the integrated change of the refractive along the line of sight due to a 3D temperature and/or concentration field. In the case of Comparative Digital Holography shape and/or deformation can be compared to a master object.

2.1 Digital Holographic Interferometry

Digital Holographic Interferometry is based on the digital recording and storage of two or more holograms in different states of the object (where the illumination with different laser light colors may be regarded as different states of the object). Since no imaging optics is needed, the optical far field is recorded. Depending from being in the Fresnel-region or the Fraunhofer-region of the far field, the numerical reconstruction of the wave fields is performed by the Fresnel-transform or the Fourier-transform. The reconstruction gives direct access to the phase as the interesting quantity for the measurement. In Digital Holographic Interferometry one hologram serves as the reference state. The interference phase distribution to be measured is calculated as the difference between the separately calculated phase distributions of the two wave fields to be compared [6].

2.2 Digital Speckle Pattern Interferometry

Digital Speckle Pattern Interferometry (DSPI), also called *Electronic Speckle Pattern Interferometry (ESPI)*, originally was developed to overcome the difficulties with stability and resolution requirements, and the time consuming handling of silver halide emulsions, and to use electronic cameras instead [7]. The imaging system focuses onto the surface of the object illuminated by coherent light. A reference wave is superposed on the focal plane of the camera, i.e. the CCD target, and the resulting speckle pattern is recorded. Thus the method can be regarded to be a digital image plane holography. To adapt the speckle size to the resolution of the camera, object and reference waves have to be almost collinear. The speckle size, which is controlled by the aperture of the imaging system, is responsible for the noisy appearance of the DSPI fringe patterns – not only that of the raw images, but that of the correlation fringes too. According to the theory of

speckles, the primary intensity distribution shows a fully noise modulated pattern, Fig. 7-1a. During measurements the changes of the phase distribution of the object wave created by the physical phenomena are of interest. Those happen because of the deformation of a diffusely reflecting surface, or due to the change of the refractive index distribution of the transparent media the object wave is passing through, and results in an altered speckle pattern, Fig. 7-1b. If the second pattern is subtracted from the first one, which was stored in the computer, one obtains the so-called correlation fringes, which uniquely depend on the phase changes to be measured, Fig. 7-1c: Low frequency correlation fringes disturbed by high frequency high contrast speckle are the result of the subtraction. Quantitative evaluation of these patterns consists in the numerical determination of the interference phase distribution from these fringe patterns.

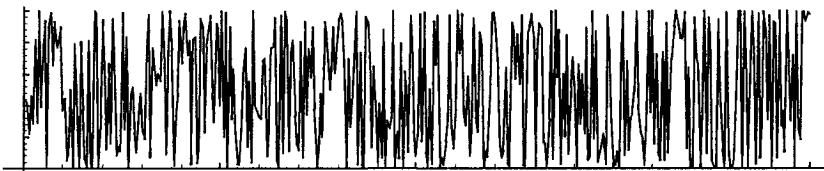


Fig. 7-1a: Primary speckle Pattern $I_A(x)$

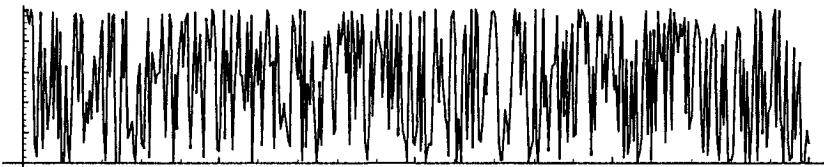


Fig. 7-1b: Primary speckle Pattern $I_B(x)$

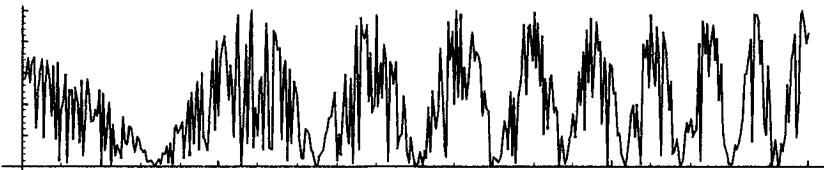


Fig. 7-1c: Positive intensity difference $|I_A(x) - I_B(x)|$

A refined DSPI method records and stores four primary speckle patterns before and four after the phase change, all with mutual phase shifts in the reference wave. The point wise solution of the two systems of equations for the intensity at a given point yields two phase distributions, whose difference is the desired interference phase distribution [8]. By this method the speckle noise is strongly reduced and one gets access to the correct sign of the interference phase.

2.3 Shearing Interferometry

Shearing interferometry, also called *differential DSPI*, is performed by superposing to the object wave not an arbitrary collinear reference wave field but a physically shifted (=sheared) version of the object wave field [9]. This is produced by a shearing element, which can be a glass wedge in front of one half of the imaging lens, two tilted glass plates, a Wollaston prism, or a Michelson interferometer with one mirror slightly tilted. Performing the same procedure as with DSPI, one obtains correlation fringes, too. However, these fringes do not depend on the displacement of the surface points but on the difference quotient in the direction of the shear. Thus only the spatial variation of the optical phase in the shearing direction is measured, while the method is insensitive to rigid body motions or spatially constant phase changes. Therefore this method does not pose the same strong requirements on alignment and vibration isolation as holography or DSPI are doing.

3. DISTANT MEASUREMENTS BY DIGITAL HOLOGRAPHY

Distant measurements may be performed by recording a holographic image of one object before the transportation to another place and observing changes at another place. However, more challenging is the measurement of the shape or deformation in comparison of two different objects at different locations. It requires transmitting the stored hologram to the second place with the difficulties of transmission bandwidth in a number of cases. About the solution to compress the data and decompress at the work place will be reported in the next section.

The hologram has to be reconstructed optically for the interferometric measurement. Different method of reconstruction can be applied, e.g. LCD displays or Digital Mirror Devices. The different methods are described in the next subsection.

A big challenge is the interferometric comparison of two different objects with rough surfaces. In common, the two object waves will not interfere in a visible way. A way out of this problem has been invented by Neumann [10] and further developed by Füzessey et.al. [11]. The method will be described.

3.1 Optical Reconstruction of Digital Holograms

The optical reconstruction of the hologram can be performed by intensity and/or phase modulation of a light wave comparable to the reference wave. Three methods may be applied to this task in principle:

- Liquid Crystal Displays (LCD) are able to modulate the intensity as well as the phase,
- Digital Mirror Devices can modulate the intensity, and
- Illuminated and developed light sensitive photo materials may modulate the intensity as well as the phase.

However, the pixel size of the CCD target does not match the size of the spatial light modulators. A software fit must be done before reconstruction.

Liquid Crystal Displays

Liquid Crystal Displays (LCD's) are electro-optical devices used to modulate optical waves. An LCD is an array of cells from liquid crystal material. An individual LCD cell changes its transmittance or its refractive index depending on the applied voltage. It is therefore possible to modulate the brightness or phase of light, which passes the device, i.e. an LCD can be applied to the reconstruction of an optical hologram, Fig. 7-2 [1]. At first a digital hologram is recorded with a CCD-sensor, Fig. 7-2a. The hologram is stored and transmitted to the measurement site and optically reconstructed by the LCD, which modulates the optical wave corresponding to the reference wave with the hologram function and generates the original object wave, Fig. 7-2b. The virtual image can be observed at the position of the original object. Alternatively it is possible to reconstruct the real image by illuminating the LCD with the conjugate reference wave.

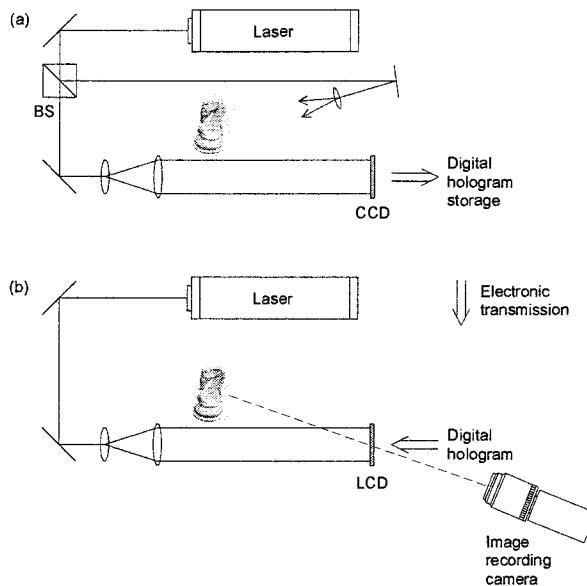


Fig. 7-2: a) Digital hologram recording with a CCD
b) Optical reconstruction with an LCD [1]

The digitally recorded hologram of a chess knight has been optically reconstructed by an LCD as an example, Fig. 7-3a. Furthermore, the optical reconstruction of two superimposed holograms, which are recorded in different object states results in a holographic interferogram, Fig. 7-3b.

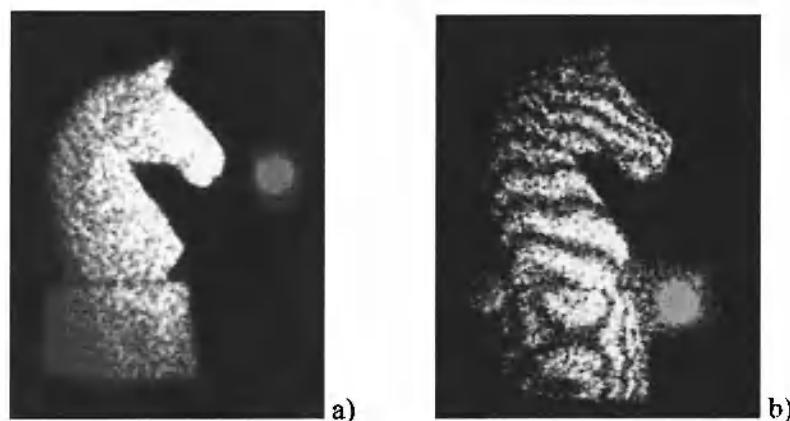


Fig. 7-3: a) Optical reconstruction of a Digital Hologram by means of an LCD, b) Optically reconstructed holographic interferogram [1]

Digital Mirror Devices

Digital Mirror Devices are electro-optical components with an array of addressable micro-mirrors. The tilttable aluminium mirrors are mounted on hinges over a CMOS static random access memory (SRAM). DMD's are available with up to 1280×1024 mirror elements today. The individually addressable mirrors can be tilted binary either -10 deg (on) or $+10 \text{ deg}$ (off) along an axis diagonal to the micro-mirror. A DMD can be used as spatial light modulator, too. The application of a DMD to reconstruct a digital hologram optically has been published by Kreis, Aswendt und Höfling [12]. In an optical reconstruction set-up DMD's are therefore operated in reflection. In contrast to LCD's, which absorb up to 90% of the available light, a DMD is a reflective device yielding much more light. Consequently the diffraction efficiency in the hologram reconstruction is better compared to LCD's.

The disadvantage of hologram reconstruction is the intensity modulation by the binary mirror deflection, which leads only to two states on or off. An on-time synchronisation is required, a difficulty solved by modern electronics¹. Furthermore, the DMD on-time must be controlled by pulse-width modulation for the intensity variation. This means different diffraction states throughout one cycle.

¹ Electronics for DMD's with excellent synchronization properties a delivered by ViaLux, Germany.

Photographic Reconstruction

The transmitted hologram can be transformed into a photographic hologram by a computer printer like a computer generated hologram. The high resolution of standard ink-jet or laser printers with up to 3000 dots per inch makes it possible to print digital holograms directly on a transparent film. The hologram is then reconstructed by illuminating this film with the reconstruction (=reference) wave. The size of each pixel may be adapted by the printing program.

3.2 Comparative Digital Holography

Interferometry is the comparison of the two optical waves. In holographic interferometry at least one of these waves is stored by a hologram. This enables to compare two states of an object with a rough surface in contrary to conventional interferometry where one wave is generated by wave front division. By the interference the phase differences between the two wave fields can be measured. In this way it is possible to measure the change of the refractive index in transmission objects or the shape and deformation of a technical object. However, a severe restriction in conventional hologram interferometry is the requirement that the microstructures of the surfaces to be compared are identical otherwise the two reflected speckle fields are decorrelated and no interference occurs: Standard hologram interferometry is restricted to the comparison of two states of the *same* object.

Comparative interferometry is a method to overcome this restriction [10, 11]. The method is based on the illumination of the two states of the test component with the corresponding conjugated object wave of the master object. The object wave of the master component acts as a coherent mask for the adaptive illumination of the test component. Therefore comparative interferometry is performed in the following way: First, two holograms of the master object are taken in the two states according to a specific load. The simultaneous reconstruction of these holograms generates an interferogram. The interference phase depends on the displacement by:

$$\Delta\varphi_1(x, y) = \frac{2\pi}{\lambda} \vec{d}_1(x, y, z) (\vec{b}_1 - \vec{s}_1) \quad (7-1)$$

Afterwards, the test object is illuminated in the original observation direction \vec{b}_1 by the reconstructed, conjugated wave front of the master object: the real image of the master object is projected onto the test object, which is then observed in the original illumination direction \vec{s}_1 . This leads to the following relations

$$\vec{s}_2 = -\vec{b}_1 \quad \vec{b}_2 = -\vec{s}_1 \quad (7-2)$$

$$\Delta\varphi_2(x, y) = \frac{2\pi}{\lambda} \vec{d}_2(x, y, z) (\vec{b}_2 - \vec{s}_2) = \frac{2\pi}{\lambda} \vec{d}_2(x, y, z) (\vec{b}_1 - \vec{s}_1) \quad (7-3)$$

The test object is illuminated by the conjugated wave fronts of the master. Therefore, the interferogram indicates the difference of the displacements between the two objects:

$$\Delta\varphi(x, y) = \Delta\varphi_1(x, y) - \Delta\varphi_2(x, y) = \frac{2\pi}{\lambda} (\vec{d}_1(x, y, z) - \vec{d}_2(x, y, z)) (\vec{b}_1 - \vec{s}_1) \quad (7-4)$$

The comparative holography can be combined with Digital Holography to the so-called Comparative Digital Holography [13, 14]. Comparative Digital Holography is performed in a Digital Holography set-up and adapts the comparative method, Fig. 7-4. The digital hologram of the master object is recorded at a location A, Fig. 7-4a. The transmission of this digital hologram to a test location B can be done by any data transfer medium, e.g. by the internet. At location B a Liquid Crystal Display is programmed by the hologram data to serve as spatial light modulator to reconstruct the hologram. Comparative holography is based on the illumination of the second object by the conjugated wave fronts of the master object as coherent masks, Fig. 7-4b. The observation is done in the original illumination direction.

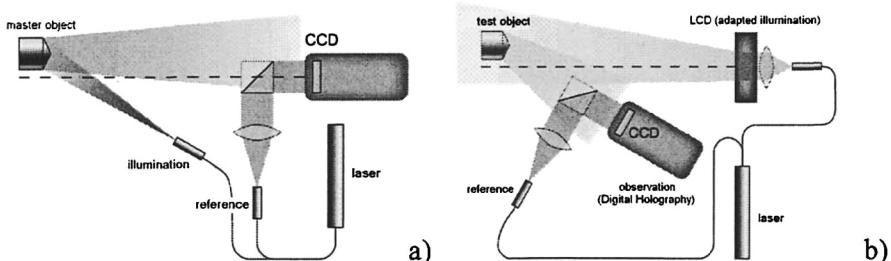


Fig. 7-4: Comparative Digital Holography [14], a) recording of the master object as coherent mask, b) illumination of the test object with the conjugated wave front of the master

The advantage of Comparative Digital Holography compared to conventional comparative holographic interferometry is the chance to store all states separately and to reconstruct them later independently, i.e. reducing the technical requirements for comparative measurements significantly.

4. DATA COMPRESSION AND DECOMPRESSION

4.1 General Remarks

Tele-metrology means the real time observation and control of holographic experiments from a remote place, too. This offers the possibility of an earth-bound quantitative pre-evaluation of interferometric fluid science experiments performed in space. As described in the introduction, the main problem is to reduce the data rate by a factor of 50-100 (e.g. 120 Mbps to 2 Mbps). Another problem can be the economic storage of the scientific data in space. This can be especially important for space experiments using small carriers like sounding rockets, capsules or satellites offering less space and power for storage devices.

Data compression in general consists of taking a stream of symbols and transforming them into codes. If the compression is effective, the resulting stream of codes will be smaller than the original symbols. The decision to output a certain code for a certain symbol or set of symbols is based on a model. The model is simply a collection of data and rules used to process input symbols and determine which codes to output [15]. As an example let the symbols be the alphanumeric characters. In a statistical model, Fig. 7-5, each character is designated its probability of occurrence. Then a good strategy is to assign short codes to symbols with high probability and longer codes to those with low probability. This in most cases will result in average code lengths which are shorter than an assignment of equal length codes to all characters. Huffman coding is based on this concept [16].

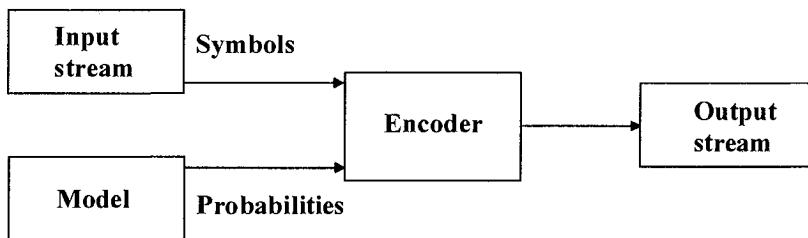


Fig. 7-5: Coding based on statistical model

Data compression techniques are divided into two major families: lossless and lossy compression. Lossless compression consists of those techniques guaranteed to generate an exact duplicate of the input data stream after a compression/decompression cycle. Common coding schemes performing lossless compression are Huffman coding [16], the LZ77 and LZ78 algorithms of A. Lempel and J. Ziv [17, 18], the PKZIP routine of Phil Katz, or the PCX, GIF, BMP formats for computer images. On the other hand

lossy data compression concedes a certain loss of accuracy in exchange for increased compression. Most lossy compression techniques can be adjusted to different quality levels, gaining higher accuracy in exchange for less effective compression. Frequently used lossy techniques are ADPCM (Adaptive Differential Pulse Code Modulation) or LPC (Linear Predictive Coding) for sampled audio signals, and the JPEG (Joint Photographic Experts Group) for computer images. Recent image compression is based on fractal theory.

For tele-metrology application now we have to look for efficient compression techniques for patterns produced by ESPI/DSPI, shearography, and digital holography. The primary speckle patterns of ESPI/DSPI, of digital shearography as well as the digital holograms exhibit high contrast and high spatial frequencies. Thus they have high information theoretic entropy which implies low compression rates. Therefore the common lossless coding schemes for images will result in compression factors of less than 2. On the other hand lossy compression, even with moderate compression factors in the range of 10, will destroy the high frequency information in the patterns, so the subsequent subtraction (ESPI/DSPI or digital shearography) or Fresnel-/Fourier-transform (digital holograms) will give no correlation fringes or useful wave fields any more.

However the speckle difference patterns of ESPI/DSPI and digital shearography - like the one in Fig. 7-1c – present themselves as low spatial frequency correlation fringes which are corrupted by high frequency speckle noise. Now lossy compression destroys the speckle noise but maintains the correlation fringes. A similar behavior one finds in digital holographic interferometry. The interference patterns we obtain in deformation analysis or refractive index measurements in fluid science experiments both exhibit low frequency fringes overlaid with a speckle noise, which in the case of digital holographic interferometry is less severe than in e. g. ESPI/DSPI.

Now there are two directions for compression/decompression: the first is to perform the image subtraction or numerical reconstruction at the sensor site and to transmit only the resulting interference patterns, the second is a compression of the primary speckle patterns or digital holograms, which are then transmitted, and where the decompressed patterns or holograms are then used to calculate the interferograms. While in the first mentioned case a lossy compression may alter the speckles, but maintains the correlation fringes - this may even result in an image enhancement optimizing the interference fringes - in the second case the high frequency structure must not be changed to allow numerical reconstruction. Nevertheless also in this case we aim at an efficient compression.

4.2 Comparison of Different Methods

The objective of the comparison of data compression is to analyze the lossy compression performance of such compression algorithms which are able to recover the image information which is required for tele-metrology operation. A well known international standard algorithm for visual image quality-oriented lossy compression is the JPEG algorithm. It is known that due to the 8×8 pixel-block structure, the recovered image information is more or less reduced to the mean gray level values of such a structure resulting in compression factors (CF) between $CF=20$ and $CF=40$. On the other hand, studies in different fields, e.g. with particle tracking images [19], have demonstrated the advantage of wavelet-based algorithms for providing higher spatial accuracy and higher compression factors ($CF>60$). In particular, wavelet-based algorithms which use a coefficient-tree-based embedded encoding method gave the best results. These algorithms are: the Shapiro 'Embedded Zerotree Wavelet' approach (EZW) [20], and the Said/Pearlman algorithm called 'Set Partitioning in Hierarchical Trees' (SPIHT) [21]. The high performance for spatial accuracy of these algorithms is directly based on two particular features. a) The discrete wavelet transform decorrelates the image information into multi-resolution layers of 'coefficients images' which represent the true spatial information of the image. b) The coefficient-tree-based encoding process is able to encode the high resolution information already at coarse quantization levels (high compression levels) as each of the coefficient trees (which can be seen as multi-resolution coefficient vectors) incorporates all the spatial position-relevant information out of the actual local position addressed by the tree-vector in the multi-resolution layers of wavelet coefficients. The SPIHT algorithm was selected for the actual compression analysis, as it also provides control for constant image quality and for constant data rate output. The JPEG 2000 algorithm [22] was found not useful, because even as it is a wavelet-based approach, it does not provide the advantages of tree-based embedded encoding for delivering spatial accuracy at high compression factors. JPEG 2000 also requires full frame image buffering and the compression process is based on image segmentation, which would lead to complex implementations for encoding of time-related telemetrology images out of the many image segments.

4.2.1 Lossy Compression of Interferograms

The performance of lossy compression by the SPIHT-algorithm is shown by treating interferometric fringe patterns of different fringe densities with different compression factors. The patterns are DSPI-patterns (with 36, 17, and 10 pixels/fringe) as well as shearographic interferograms (between 80

and 6 pixels/fringe in a single pattern) and holographic interferograms (with 57, 16, and 6.4 pixels/fringe). A typical speckle difference pattern shows a high frequency noise visible in an enlarged detail, Fig. 7-6a. Fig. 7-6b gives the same pattern after compression/decompression with the compression factor CF set to 40 - meaning the transmitted data was 1/40 of the data amount of the original pattern of Fig. 7-6a. Fig. 7-6c displays the results for a compression factor CF=100. The dark fringes remain dark, the bright ones bright.

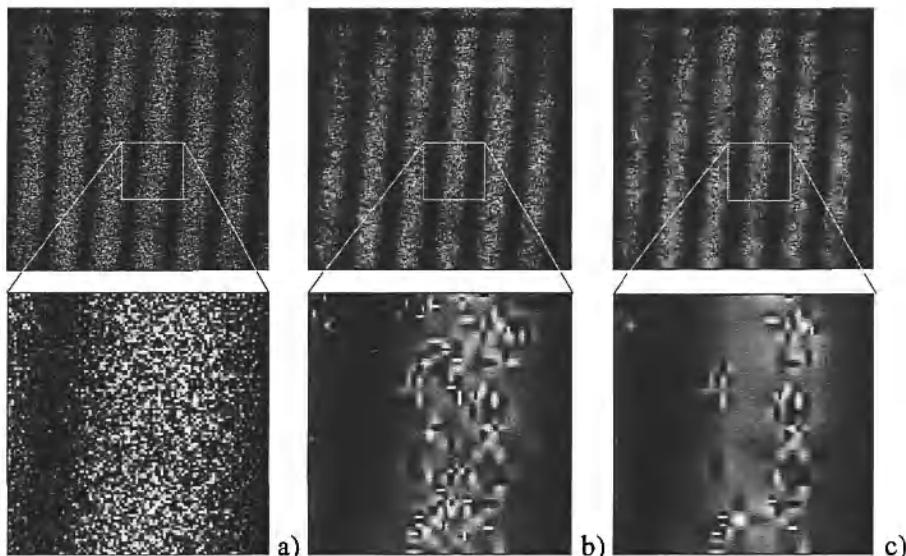


Fig. 7-6: Speckle difference pattern, a) original, b) wavelet compression, CF=40, c) wavelet compression, CF=100

However, the stochastic speckle pattern is changed to an irregular pattern of artifacts. Nevertheless, in all cases the overall behavior of the speckle difference pattern remains unchanged, so that a reliable qualitative judgment remains possible. Even if one performs a quantitative evaluation [6], the artifacts stemming from compression/decompression will pose no problem. This is due to the fact that most quantitative evaluation methods inherently perform a low-pass filtering which eliminates the artifacts nearly totally. The interference phase distribution modulo 2π of the original DSPI-Patterns, Fig. 7-6a, can be obtained by the Fourier transform method [6], Fig. 7-7a. The irregularities in the interference phase distribution in Fig. 7-7a are due to the speckle noise. For a quantitative comparison the interference phase distributions are calculated from compressed/decompressed patterns, Figs. 7-7b, -11. They are nearly identical to that of the uncompressed pattern in Fig. 7-7a.

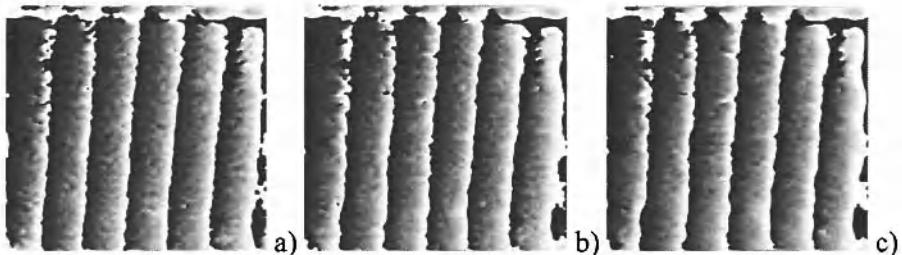


Fig. 7-7: *Interference pattern, a) no compression, b) compression factor $CF=40$, c) compression factor $CF=100$*

Compression works well as long as the interference fringe pattern is of low spatial frequency while the speckle noise is of high frequency, Figs. 7-8 to 7-9c. The speckle difference pattern without compression is depicted in Fig. 7-8. The average spatial frequency of the interference fringes is higher than in Fig. 7-6a. The interference phase modulo 2π of Fig. 7-9a without compression shows how speckle noise is transformed into phase noise. The interference phase distributions of compressed and decompressed speckle difference patterns, Figs. 7-9b and 7-9c, still allow for a qualitative interpretation of the phase patterns, but they are corrupted by noise and a merging of dark regions: A limit of the spatial frequencies is approached where for higher frequencies a reliable quantitative evaluation is not possible any more. Also a qualitative judgment becomes difficult due to a vanishing of the interference fringes in the speckle noise of nearly the same frequency.

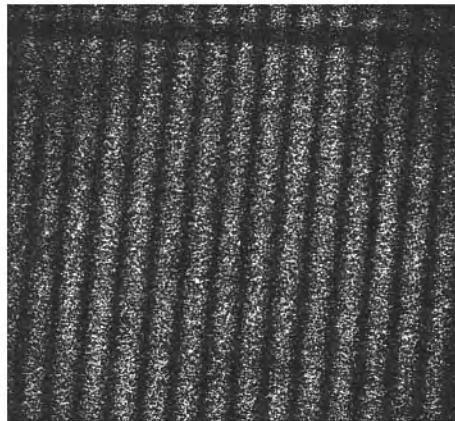


Fig. 7-8: *Speckle difference pattern, no compression*

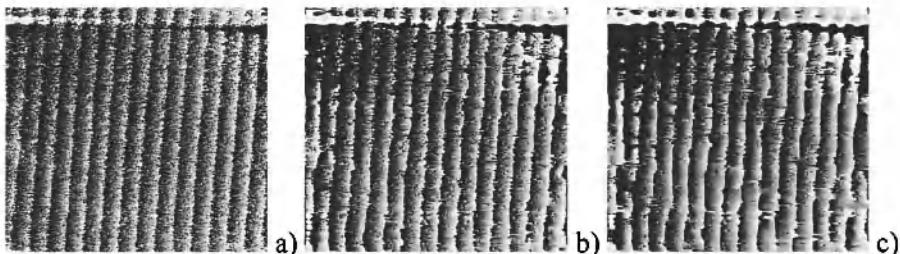


Fig. 7-9: Interference phase, a) no compression, b) compression factor $CF=40$, c) compression factor $CF=100$

The consequences are that high spatial frequencies of the fringes in the difference patterns have to be avoided. This can be done by taking a new reference pattern for the subtraction process before the optical phase change due to the experiment takes too high values and thus produces too much fringes per frame. Today this becomes possible with high-speed cameras. The presented speckle difference patterns exhibited nearly equidistant parallel fringes. But in most applications the fringe density and fringe orientation in a frame varies as demonstrated by the example of shearographic pattern. Figs. 7-10 to 7-11b. Fig. 7-10 gives the uncompressed pattern together with an enlarged detail in which the spatial frequencies of the fringes increase. Fig. 7-11a displays the same detail after compression/decompression by a compression factor of 40 using the Said/Pearlman algorithm. The fringes can still be recognized; a qualitative evaluation is possible. For comparison the same pattern is compressed/decompressed by the same CF but using the JPEG procedure, Fig. 7-11b. The block structure typical for JPEG arises, and the fringe pattern begins to vanish. Holographic interference patterns look like the shearographic patterns and in the experiments led to comparable results. This example proves the superiority of the Said/Pearlman algorithm over JPEG for the purpose of tele-metrology.

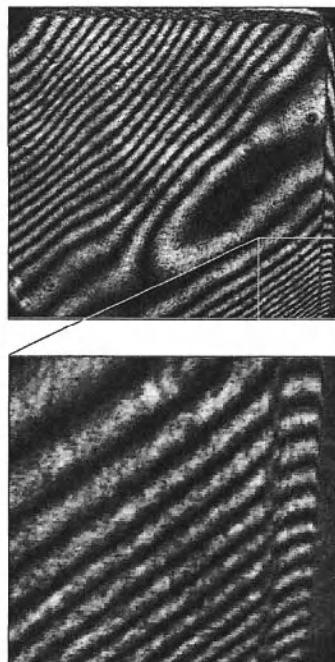


Fig. 7-10: Shearographic interference pattern, no compression

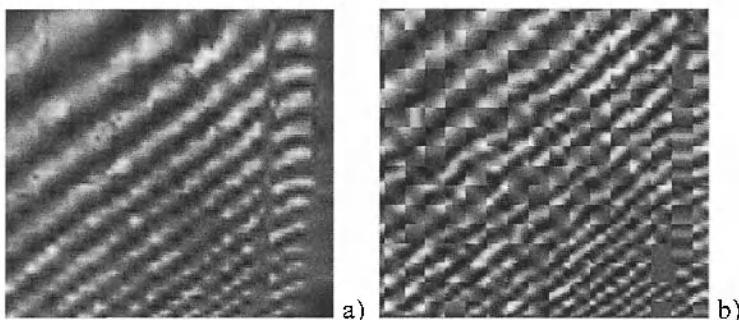


Fig. 7-11: Shearographic interference pattern, a) Said/Pearlmann compression, CF=40,
b) JPEG compression, CF=40

4.2.2 Compression of Primary Speckle Patterns and Digital Holograms

The compression of the primary speckle patterns in DSPI/ESPI or shearograms, which by nature consist of objective speckles, is more difficult than the compression of interferograms. This is true for Digital Holograms, too, since their information is coded in high frequency micro-interferences. Compression methods like JPEG, which are great in compressing common

photographs by retaining the visual content, fail because they alter the high frequency speckle pattern which seems stochastic but in its microstructure contains the information about the wave field. First attempts to compress digital holograms were performed by Naughton et al [23, 24], but their background was in pattern recognition, so they are only interested in a visually acceptable reconstruction of the intensity of the recorded wave field. On the other hand in tele-metrology interferometric experiments are the essentials, so the main interest is in the reconstructed phase of the wave field. This poses higher demands on the reliability of the compressed and decompressed speckle fields. The search for an efficient compression method has to recognize the special nature of the speckle fields, i. e. the electromagnetic field properties of the holographic data [25].

A new approach of Kayser is based on the relations of the electric field in space [25]. The spatial properties of an optical wave field are ruled by the Helmholtz equation, the complex amplitude $u(x,y,z)$ of any monochromatic light field propagating in vacuum or in a homogeneous dielectric medium must obey [26]. For low speckle densities, i.e. in the far field, the Helmholtz equation can be treated as a two-dimensional differential equation with a disturbance in the direction of light propagation. This disturbance $\rho(x,y)$ in the plane z_0 is given as the difference to a plane wave's propagation

$$\left. \frac{\delta^2 u(x,y,z)}{\delta z^2} \right|_{z=z_0} \approx -k^2 u(x,y,z_0) + \rho(x,y) \quad (7-5)$$

with the wave number k . Here the Helmholtz equation is reduced to a 2D differential equation

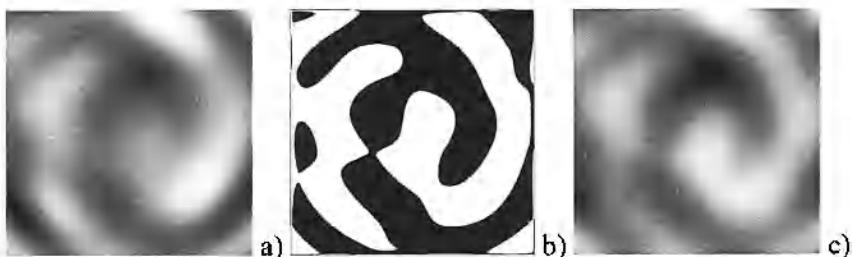
$$0 = \Delta u(x,y,z) + k^2 u(x,y,z_0) \approx \frac{\delta^2 u(x,y)}{\delta x^2} + \frac{\delta^2 u(x,y)}{\delta y^2} + \rho(x,y) \quad (7-6)$$

which is of elliptic type and can be calculated by 2D boundary conditions. Using the discrete cosine transform (DCT) for von-Neumann boundary conditions, the discrete data can be calculated by

$$\rho_{m,n} = \frac{2}{\Delta^2} \left(\cos \frac{\pi m}{J} + \cos \frac{\pi n}{L} - 2 \right) u_{m,n} \quad (7-7)$$

with pixel indices m and n in x - and y -direction, pixel pitch Δ , and numbers of pixels J in x - and L in y -direction. Nearly all information of $\rho_{m,n}$ comes from its boundary and only very little data is needed from the interior

region: Already a 1-bit quantization of the interior of $\rho_{m,n}$ leads to good accordance of the decompressed data with the original, Fig. 7-12.



*Fig. 7-12: a) Real part of a holographic e-field with low speckle density,
b) 1-bit quantized real part of $\rho_{m,n}$ which leads to the c) decompressed real
part with normalized cross correlation peak height of 0.85*

The disturbances in Eq. (7-6) increase for increasing speckle densities, because the validity of the approximation of Eq. (7-5) is lost. The information contents are spread over the whole data set, because the boundary is now of 3D type. For this reason the transformation will be omitted in Eq. (7-7) and transferred directly to the DCT domain. Looking to the distribution of the data in the DCT domain for positive and negative frequencies separately, the values of the constituting coefficients show an exponential behavior. The idea consists in changing the simple quantization to a quantization of the logarithms of positive and negative values. This includes the possibility to introduce a certain threshold as well as a better approximation of the quantized data.

The general complex electromagnetic fields will be restricted to the intensity distributions for the purpose of compressing the raw speckle fields and digital holograms in tele-metrology, which is the fact for this application: The strategies explained for electromagnetic data with high space-bandwidth products can now be used for holograms with their high speckle densities. To allow comparability between digital holograms of different origins, a two-step compression scheme is applied:

- First the data are down-sampled to a resolution making the data comparable in their space-bandwidth products.
- Second the data are quantized in the DCT-domain by taking the logarithms of their positive and negative parts and quantizing it. To guarantee a reproducible quantization, the input data has to be scaled to a maximum sampling range (8 bit or 10 bit camera resolution). For a good decompression quality, the energy of the data is conserved.

The resulting data now can be further compressed by any lossless compression technique. But to compare the different compression ratios,

only the compression results from the second step are used. A reasonable quality measure in this comparison is the normalized cross-correlation.

The described compression algorithm has been applied to various simulated as well as real digital holograms. The first simulated holograms have been from a plane with randomly distributed scattering points. The number of these points and the wavelength of the assumed laser light could be changed. The holograms are calculated in different distances from the object plane, Fig. 7-13. A further simulated object was a black and white pattern of variable size rectangles. A lensless Fourier-transform hologram was calculated by using a reference wave diverging from the plane of the object, Fig. 7-14a. The reconstructed intensity, Fig. 7-14b, shows real and virtual image both focused, which is a typical property of Fourier-transform holograms. The really existent object was a chess knight, its digital hologram, the reconstructed intensity and the reconstructed phase are given in Fig. 7-15.

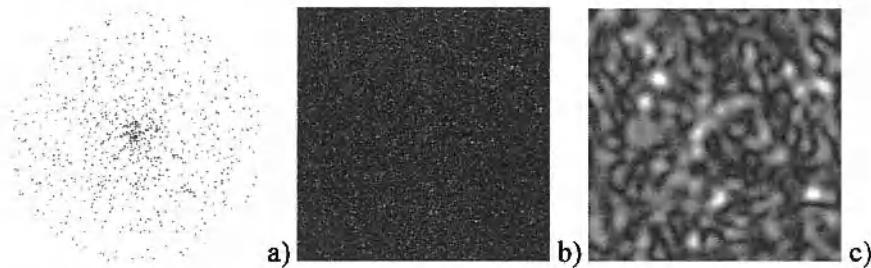


Fig. 7-13: a) Scattering plane with roughness $0.6 \mu\text{m}$, 1000 scattering centers, b) hologram in distance 100 mm, c) hologram in distance 1000 mm

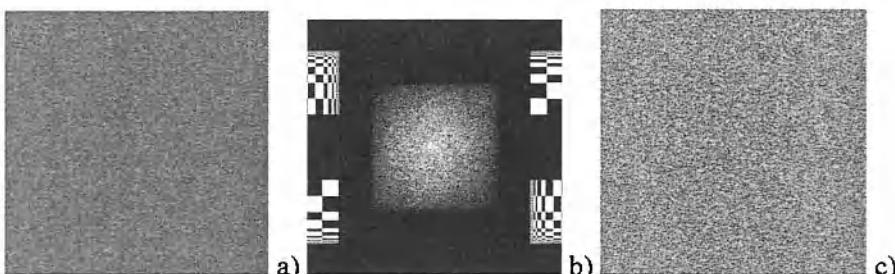


Fig. 7-14: a) Simulated lensless Fourier-transform hologram, b) reconstructed intensity distribution, c) reconstructed phase distribution

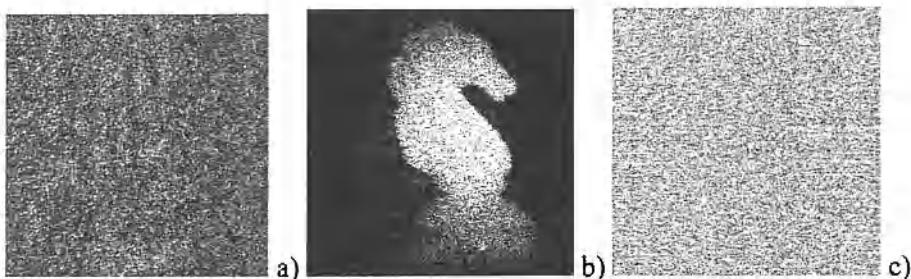


Fig. 7-15: a) Experimentally recorded Digital Hologram, b) reconstructed intensity distribution, c) reconstructed phase distribution

The tests with the simulated data have been performed with different compression rates. The cross-correlation peak height (CP) was used as a measure of matching. The scattering centers produced a $CP=0.94$ for 4-bit compression, a $CP=0.87$ for 3-bit compression, and a $CP=0.75$ for 2-bit compression. So even with 2 bit data representation there is a good agreement indicated by the CP . The rectangles-pattern showed nearly the same results. With increasing compression rate the sharp borders of the rectangles loose their contours and the noise in the black and white areas increases.

The digital hologram of the real chess knight was compressed using standard JPEG and JPEG2000 algorithms for comparison. The achievable CP -values of JPEG ($CP=0.90$) and JPEG2000 ($CP=0.85$) are reasonable. However, the wavelet compression of JPEG2000 generates artifacts in the form of a double image, which originate from sub-sampling in the compressor. The compression with the wave field algorithm shows nearly the same results for JPEG and 4-bit compression ($CP=0.93$) concerning the quality measure and the data size. The reconstructed image of 3-bit wave field compression ($CP=0.82$) algorithm is comparable to JPEG2000, but it does not show the artifacts. Even a 2-bit compression by the wave field algorithm generates acceptable results ($CP=0.69$).

Comparable results also have been obtained with the reconstructed phase distributions. This is since the phase field is a comparable matrix of values.

5. CONCLUSIONS

Tele-metrology by coherent-optical methods can be performed in two different ways:

- In a first principle measurement set-up the data of a hologram including one image of Digital Speckle Pattern Interferometry can be transferred to another place in order to serve as a reference for the measurement at another object. By using the Comparative Digital Holography nearly no restrictions are given concerning the objects or the locations. However, a

bottleneck may be the bandwidth of the transmitting channel. The compression of the original data may overcome the difficulties.

- The second way is to perform the data acquisition in one remote location and transmit the data to the evaluation place. Two different chances to perform these measurements are given in this case: The holograms or the interferograms may be transmitted. In any way high data rates with high frequency contents may occur. A compression of data is required, too.

Tele-metrology based on digital holography and related methods requires efficient data compression as mentioned before. If the interference patterns are to be transmitted, the SPIHT-algorithm is optimal in obtaining high compression factors. If the primary speckle patterns or digital holograms must be transmitted, a special purpose algorithm taking into account the electromagnetic field properties of the patterns is advantageous. Developments are going on to make these algorithms a standard in future.

ACKNOWLEDGEMENTS

The authors thank A. Henrichs of Astrium GmbH, J. Becker of ESA/ESTEC and E. Kolenovic and D. Kayser, both of BIAS, for their contribution to the topics reported here. Part of the research was supported by ESA/ESTEC under contract 16577/002/NL/PA, a further part was sponsored by the Deutsche Forschungsgemeinschaft (DFG) under contract Kr953/20-1. This financial support is gratefully acknowledged.

REFERENCES

- [1] U. Schnars, W. Jüptner, *Digital Holography*, Springer Berlin Heidelberg New York, 2004
- [2] Th. Kreis, W. Jüptner, J. Becker, A. Henrichs, *Telescience and interferometric metrology on the International Space Station*, Acta Astronautica, 2005
- [3] D. Gabor, *A new microscopic principle*, Nature 161, 777-778, 1948
- [4] U. Schnars. Direct phase determination in hologram interferometry with use of digitally recorded holograms, J. Opt. Soc. Amer. A, 11, 2011-2015, 1994
- [5] U. Schnars, Th. M. Kreis, and W. P. O. Jüptner. Digital recording and numerical reconstruction of holograms: reduction of the spatial frequency spectrum, Opt. Eng., 35(4), 977-982, 1996
- [6] Th. Kreis. *Handbook of Holographic Interferometry*, Wiley-VCH, Weinheim, 2005
- [7] A. E. Ennos. *Speckle interferometry*, in J. C. Dainty (Ed.), *Laser Speckle and Related Phenomena*, Topics in Applied Physics, vol. 9, Springer, Berlin, 203-253, 1975
- [8] K. Creath. *Phase shifting speckle interferometry*, Appl. Opt., vol. 24, 3053-3058, 1985
- [9] W. Steinchen, L. Yang. *Digital Shearography: Theory and Application of Digital Speckle Pattern Shearing Interferometry*, SPIE Press, Bellingham, WA, 2003

- [10] D.B. Neumann. *Comparative Holography*, In: Tech. Digest. Topical Meeting on Hologram Interferometry and Speckle Metrology, paper MB2-1. Opt. Soc. Am.
- [11] Z. Füzesi, F. Gyimesi. Difference holographic interferometry: displacement Measurement, Opt. Eng. 23, 780-783, 1984
- [12] Th. Kreis, P. Aswendt, R. Höfling. *Hologram reconstruction using a digital micromirror device*, Opt. Eng. Vol. 40, 926 – 933, 2003
- [13] W. Osten, T. Baumbach, S. Seebacher, W. Jüptner. *Remote shape control by comparative digital holography*, In: Jüptner W, Osten W (eds) Proc 4th International Workshop on Automatic Processing of Fringe Patterns. Akademie Verlag, Berlin, 373-382, 2001
- [14] W. Osten, T. Baumbach, W. Jüptner. *Comparative digital holography*. Opt. Lett. 27(20), p. 1764-1766, 2001
- [15] M. Nelson, J.-L. Gailly. *The Data Compression Book*, M&T Books, New York, 2nd Ed., 1996
- [16] D. A. Huffman. *A method for the construction of minimum-redundancy codes*, Proc. of the IRE, 49(9), 1098-1101, 1952
- [17] J. Ziv, A. Lempel. *A universal algorithm for sequential data compression*, IEEE Trans. Inform. Theory, 23(3), 337-343, 1977
- [18] J. Ziv, A. Lempel. *Compression of individual sequences via variable-rate coding*, IEEE Trans. Inform. Theory, 24(5), 530-536, 1978
- [19] H.-G. Maas, M. Virant, J. Becker, W. Bösemann, L. Gatti, A. Henrichs. *Photogrammetric methods for measurements in fluid physics experiments in space*, Acta Astronautica, 50(4), 225-231, 2002
- [20] J. Shapiro. Embedded Image Coding using Zerotrees of Wavelet Coefficients, IEEE Trans. Signal Proc., 41(12), 1993
- [21] A. Said, W. A. Pearlman. *A New Fast and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees*, IEEE Trans. on Circuits and Systems for Video Technology, 6, 243-250, 1996
- [22] ISO/IEC Standard 15444-1, (JPEG 2000)
- [23] Th. J. Naughton, Y. Frauel, B. Javidi, E. Tajahuerce. Compression of digital holograms for three-dimensional object reconstruction and recognition, Appl. Opt., vol. 41, no. 20, 4124-4132, 2002
- [24] Th. J. Naughton, J. B. McDonald, B. Javidi. Efficient compression of Fresnel fields for Internet transmission of for three-dimensional images, Appl. Opt., vol. 42, no. 23, 4758-4764, 2003
- [25] D. Kayser, Th. Kreis, W. Jüptner. Compression of digital holographic data using its electromagnetic field properties, Proc. SPIE, vol. 5908, 2005
- [26] J. W. Goodman. *Introduction to Fourier Optics*, McGraw-Hill Companies, Inc., New York, Second Edition 1996

Chapter 8

COMPUTER-GENERATED HOLOGRAMS FOR WHITE LIGHT RECONSTRUCTION

Hiroshi Yoshikawa

Nihon University, 7-24-1 Narashino-dai, Funabashi, Chiba 2748501, Japan

Abstract: The white light reconstruction holograms are very practical to display 3D images because they don't require the laser for illumination. This chapter describes computer-generated holograms that can reconstruct 3D images with white light. In this chapter, we discuss a simplified model to calculate the rainbow hologram quickly. Full color holograms are also realized. We also apply stereographic approach to calculate rainbow hologram from an array of 2D images. In addition, the hybrid hologram, an optical hologram made from a computer-generated hologram, and output devices are also described.

Key words: computer-generated hologram; rainbow hologram; fast calculation; white light reconstruction; 3D display; holographic printer; color hologram;

1. BASICS OF COMPUTER-GENERATED HOLOGRAMS

Computer-generated hologram (CGH) is usually defined as the computer simulation of interference pattern formed by the object wave and the reference wave. Since there is no material that can record phase information of the light wave directly, it is necessary to use the reference wave to encode phase information as fringe pattern. There are two major ways to calculate the object wave. One is Fourier hologram and the other is Fresnel hologram. The main difference of those two holograms is the distance between the object and the hologram. In the Fourier hologram, it is assumed that the object is located at infinite distance from the hologram. In this case, the object wave on the hologram plane can be calculated as the Fourier transform of the wave front at the object plane. Since the Fourier transform is two-dimensional transform, some modifications are required to calculate the object wave from 3D object. One idea is the object is approximated by

the sum of a number of equally spaced planes (Waters, 1968; Lesem et al., 1969; Brown and Lohmann, 1969). In contrast, the object is located near the hologram in Fresnel hologram. In this chapter, Fresnel hologram is mainly described because it is quite suitable for 3D display and white light reconstruction.

2. CALCULATION METHODS OF SIMPLE HOLOGRAMS

This section describes basics of hologram calculation procedure.

2.1 Rigorous calculation

The first step in calculating CGH is calculation of the complex amplitude formed by the object beam. Calculation geometry is shown in Fig. 8-1. The hologram plane is located at $z = 0$ and perpendicular to the z -axis. The object to be recorded is approximated as a collection of self-illuminated points. The location of the i -th point is specified as (x_i, y_i, z_i) . Note that the object can be put in front of the hologram or even overlapped with the hologram, and thus is known as an image hologram. Each point has real-valued amplitude a_i and relative phase ϕ_i . The complex amplitude $O(x, y)$ on the hologram is determined from the superposition of the object wavefronts from each point by

$$O(x, y) = \sum_{i=1}^N \frac{a_i}{r_i} \exp[i(kr_i + \phi_i)], \quad (8-1)$$

where N is the number of object points. The wave number is $k = 2\pi/\lambda$, where λ is the free-space wavelength of the light. The oblique distance r_i between the object point and the point on the hologram is defined as:

$$r_i = \sqrt{(x - x_i)^2 + (y - y_i)^2 + z_i^2}. \quad (8-2)$$

The second step is adding reference beam $R(x, y)$. The total complex amplitude on the hologram plane is the interference of the object beam and reference beam, represented as $O(x, y) + R(x, y)$. The total intensity pattern,

$$I(x, y) = |O(x, y)|^2 + |R(x, y)|^2 + 2 \operatorname{Re}\{O(x, y)R^*(x, y)\}, \quad (8-3)$$

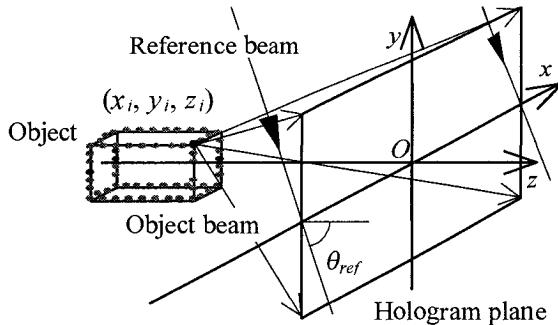


Figure 8-1. Calculation geometry for the hologram.

is a real physical light distribution comprised of three components and it can be the final hologram to be calculated.

The last thing to be considered is defining the reference beam distribution. In general, the reference beam $R(x, y)$ is represented as:

$$R(x, y) = a_{ref} \exp[i\varphi_{ref}(x, y)], \quad (8-4)$$

where a_{ref} and $\varphi_{ref}(x, y)$ are real-valued amplitude and relative phase distribution, respectively. If the reference beam is assumed as the plane wave incident from above as shown in Fig. 1, its distribution can be expressed as:

$$\varphi_{ref}(x, y) = ky \sin \theta_{ref}. \quad (8-5)$$

where θ_{ref} is an angle of the reference beam direction and z -axis.

2.2 Bipolar intensity

In computational holography, only the third term of Equation (8-3) should be calculated because the first and the second terms are unnecessary for image reconstruction (Waters, 1966; Lucente, 1993). Therefore, the holographic fringe pattern $H(x, y)$ can be obtained as:

$$H(x, y) = 2 \operatorname{Re}\{O(x, y)R^*(x, y)\}. \quad (8-6)$$

Since $H(x, y)$ can take both positive and negative value, it is named “bipolar intensity” (Lucente, 1993). It is necessary to add certain bias to make the physical hologram that takes positive values only.

3. CALCULATIONS OF HOLOGRAMS FOR WHITE LIGHT RECONSTRUCTION

To realize a practical digital holographic fringe printer for 3D display, it is important to calculate computer-generated holograms that can be reconstructed with a white light. The simplest way is the use of the image hologram (Rosen, 1966; Jendral et al. 1995). In this hologram, the real-image of the object is formed by lens near or on the holographic plate and the hologram is exposed with the reference beam. If the image is very close to the hologram plane, the spatial and the temporal coherence requirements is less critical. Therefore, white light source can be used for reconstruction with a limited depth object. CGH of the image hologram can be calculated just as same as mentioned in the section 2.

3.1 Computer-generated rainbow hologram

The “rainbow” or Benton hologram (Benton 1967) is well known as a white light reconstruction hologram. This hologram can be made similar manner as the image hologram with a horizontal slit that eliminates vertical parallax but gives very bright and sharp image with white light reconstruction. However, few papers have been published on computer-generated rainbow holograms (Leseberg and Bryngdahl, 1984; Leseberg, 1986), and those that have been are a very primitive approximation of the rainbow hologram’s nature. Their holograms are represented as the combination of an off-axis horizontal parallax only hologram and a simple vertically diffractive grating.

In this section, a simplified model is described to calculate the computer-generated rainbow hologram (Yoshikawa and Taniguchi, 1996, 1999) whose nature is almost the same as the optical rainbow hologram. Full color is also important for the practical display, and we demonstrate the preliminary results of the full color computer-generated rainbow hologram.

3.1.1 Optical model and simplified model

Since the rainbow hologram possesses horizontal parallax only (HPO), it is possible to represent the hologram with a vertically stacked array of one-dimensional holographic lines (Leseberg and Bryngdahl, 1984; Leseberg, 1986; St.-Hilaire et al., 1990; Lucente, 1993; Yoshikawa and Taniguchi, 1996, 1999). However, if we generate rainbow holograms with the rigorous optical model, the calculation procedure becomes similar to the full parallax holograms. The reason is that the wavefront from the object is diverging in both horizontal and vertical directions, although the vertical divergence

angle is limited by a slit, which is used to record the rainbow holograms. Therefore, complex amplitude of the object beam on the hologram becomes a function of both horizontal and vertical directions, so that the calculation of the hologram becomes two-dimensional as shown in equations (8-1) to (8-6). For simple and fast computation, Leseberg and Bryngdahl (1984) proposed combining the one-dimensional hologram line with a grating that has vertical spatial frequency. This grating diffracts the illumination beam in a vertical direction, as does the optical rainbow hologram reconstruction. The problem with their model is that the spatial frequencies in horizontal and vertical directions are independent so that an oblique fringe does not appear, unlike the optical hologram. This causes undesired diffraction, and reduces image brightness as a result. To solve these problems, we propose a simplified model but better approximation of the optical rainbow hologram. The additional restrictions to calculate the computer-generated rainbow hologram (CGRH) are:

1. The reference wavefront is constant along x-axis (horizontal) of the hologram.
2. The object wavefront is constant along y-axis (vertical) of the hologram.

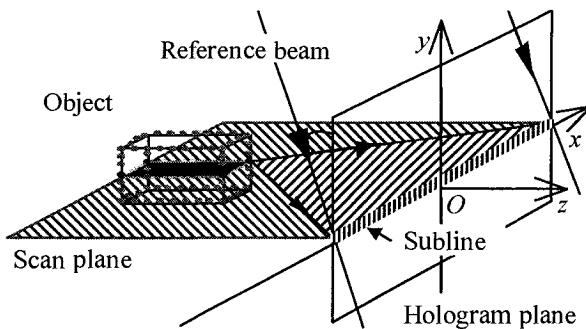


Figure 8-2. Calculation geometry for the subline.

The calculation geometry is shown in Fig. 8-2. The hologram is positioned at the $z = 0$ plane, and has horizontal and vertical axes of x and y , respectively. Restriction 1 is very moderate because it is often satisfied in the optical hologram. With the restriction 1, the reference wavefront on the hologram plane is represented as $R(y)$, which will be discussed in 3.1.3. Restriction 2 is equivalent to assuming that object points are vertically truncated line sources whose height is equal to the height of the holo-line, or the horizontal slit, which is used for rainbow holograms. Note that the line

source is assumed to radiate a cylindrical wave. Therefore, the reconstructed image also consists of vertical line sources and is observed to have an image blur. However, it should become small enough if we choose the proper parameters. With the restriction 2., the object wavefront on the hologram plane is represented as $O(x)$, which will be discussed in 3.1.2.

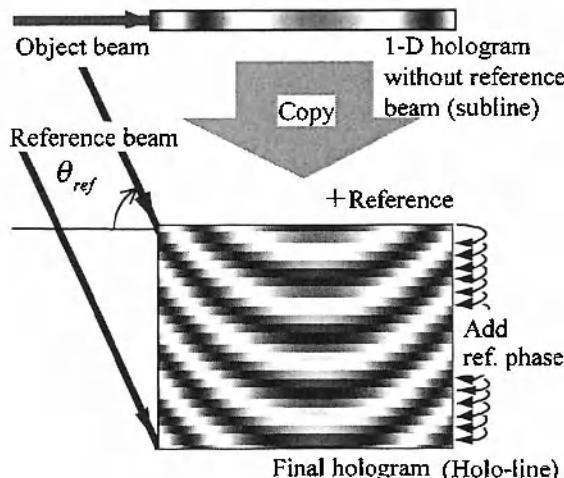


Figure 8-3. Two steps to calculate the "holo-line" of CGRH.

The procedure can be divided into two steps to calculate the single holographic line ("holo-line") of CGRH (Fig. 8-3). Note that this single holo-line consists of a vertically stacked array of one-dimensional sublines. The first step is calculating the complex amplitudes on the hologram plane caused by the object beam. The second step is adding an off-axis reference beam with a vertical incident angle. This procedure is employed to diffract the illumination beam in a vertical direction and can be performed to duplicate the subline vertically by adding a reference wavefront. The important feature of the present method is that we can process these two steps separately and store the first step results as intermediate data whose size is smaller than the final hologram. This allows high-speed computation and effective storage and/or transmission. The size and the speed factors against the rigorous model are equal to the number of sublines.

3.1.2 Calculating subline

The first step in calculating CGRH is calculation of the one-dimensional complex amplitude formed by the object beam. In this paper, we call this 1-D data a subline.

Calculation geometry for the subline is shown in Fig. 8-2. The object to be recorded is approximated as a collection of self-illuminated points. The locations of the points are specified as (x_i, y_i, z_i) . Note that the object can be put in front of the hologram or even overlapped with the hologram, and thus is known as an image hologram. Each point has real-valued amplitude a_i and relative phase ϕ_i . Since the rainbow hologram has horizontal parallax only, one subline should contain contributions only from points that lie on a single horizontal slice of the object. These slices or scan planes may be assumed to converge at the viewer's position to perform perspective correction (Leseberg, 1986). The complex amplitude $O(x)$ in the subline is determined from the superposition of the object wavefronts by

$$O(x) = \sum_{i=1}^N \frac{a_i}{r_i} \exp[i(kr_i + \phi_i)], \quad (8-7)$$

where N is the number of object points in the scan plane. The wave number is $k = 2\pi/\lambda$, where λ is the free-space wavelength of the light. The oblique distance $r_i(x)$ between the object point and the subline is defined as:

$$r_i = \sqrt{(x - x_i)^2 + z_i^2}. \quad (8-8)$$

In the conventional HPO CGH, the final holo-line is obtained by adding a reference beam to Eq. (8-7) with a horizontal angle of incident.

3.1.3 Adding reference beam with a vertical incident angle

The second step is adding off-axis reference beam $R(y)$ with a vertical incident angle. The total complex amplitude on the hologram plane is the interference of the object beam and reference beam, represented as $O(x) + R(y)$. The total intensity pattern,

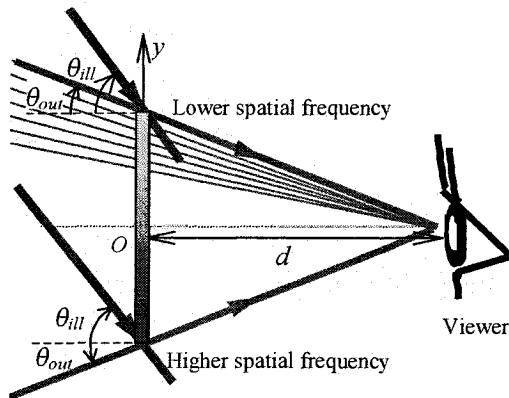


Figure 8-4. Calculation model for reference term with continuous subline spatial frequency change.

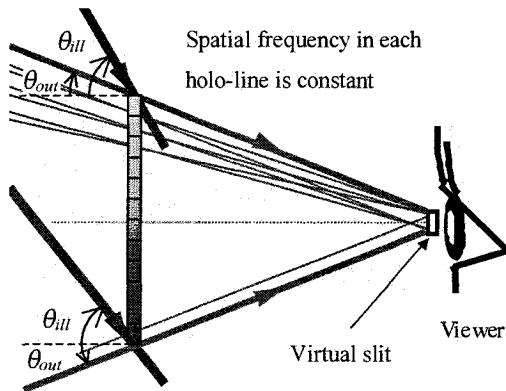


Figure 8-5. Calculation model for reference term with constant subline spatial frequency.

$$I(x, y) = |O(x)|^2 + |R(y)|^2 + 2 \operatorname{Re}\{O(x)R^*(y)\}, \quad (8-9)$$

is a real physical light distribution comprised of three components. In computational holography, only the third term should be calculated because the first and the second terms are unnecessary for image reconstruction (Lucente, 1993). Therefore, the final holo-line fringe pattern $H(x, y)$ is obtained as:

$$\begin{aligned} H(x, y) &= 2 \operatorname{Re}\{O(x)R^*(y)\} \\ &= 2 \operatorname{Re}\{O(x)\} \operatorname{Re}\{R(y)\} + 2 \operatorname{Im}\{O(x)\} \operatorname{Im}\{R(y)\}. \end{aligned} \quad (8-10)$$

There are many advantages using Eq. (8-10) to calculate the CGRH. The object term $O(x)$ is computed as rapidly as other HPO holograms. We can store and/or transmit $O(x)$ and $R(y)$ as intermediate data whose size is much smaller than the final hologram $H(x, y)$. Once $O(x)$ and $R(y)$ have been calculated, the computation of $H(x, y)$ can be done with simple multiplication and addition. This operation can also be implemented by hardware. Note that the reference $R(y)$ is independent of image contents, and can be stored in memory as a look-up table.

The last thing to be considered is defining the reference beam distribution. In general, the reference beam $R(y)$ is represented as:

$$R(y) = a_{ref} \exp[i\varphi_{ref}(y)], \quad (8-11)$$

where a_{ref} and $\varphi_{ref}(y)$ are real-valued amplitude and relative phase distribution, respectively. The relative phase distribution $\varphi_{ref}(y)$ is obtained from the relation between beam angles written as (Champagne, 1967):

$$\sin \theta_{obj} - \sin \theta_{ref} = \mu(\sin \theta_{out} - \sin \theta_{ill}), \quad (8-12)$$

where θ_{obj} , θ_{ref} , θ_{out} and θ_{ill} , are vertical angles of the object, the reference, the reconstructed (output) and the illumination beam, respectively. The angles of θ_{ill} and θ_{out} are also shown in Fig. 8-4 and 8-5. μ represents the ratio of recording and reconstruction wavelength, and it is set to one. The angle θ_{obj} is assumed to be zero. From equation (8-12) and those conditions, the phase distribution $\varphi_{ref}(y)$ can be expressed as:

$$\varphi_{ref}(y) = ky \sin \theta_{ref} = ky(\sin \theta_{ill} - \sin \theta_{out}). \quad (8-13)$$

To converge the reconstructed beam at the viewer's position, θ_{out} is defined as the function of y as:

$$\theta_{out} = \tan^{-1} \frac{y}{d}, \quad (8-14)$$

where d is distance between the hologram and the viewer. Since the CGRH has the structure of a vertically stacked array of the holo-lines, θ_{out} can also be assumed to have a constant value in each holo-line for simpler calculation as shown in Fig. 8-5. In the experimental results, the latter model is used.

3.1.4 Full Color Computation

Since the proposed method is quite similar to the optical rainbow hologram, it is also possible to synthesize full color CGRH in the same manner used in optical holography. We have made full color CGRHs by simply superimposing three primary holograms calculated with three primary wavelengths. In the experimental results, the wavelength combination 610-540-470 nm is used as the primary wavelength (Bjelkhagen and Vukicevic, 1994).

3.1.5 Results

Figures 8-6 to 8-8 show the white light reconstructed images from the rainbow holograms taken by a CCD camera. Because of a difference between viewing and camera parameters, these images are not faithful reproductions of the actual views. Moreover, the actual image has high dimensionality including focus and motion parallax cues. Since the holograms are rainbow type, color change occurs when the viewer moves up and down.

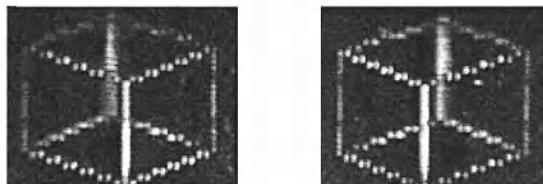


Figure 8-6. Stereoscopic pair of white light reconstructed images from different viewing positions. Left: View from left side. Right: View from right side. (Hologram size: 1,280 x 1,024 pixels; Pixel pitch: 4 μm)

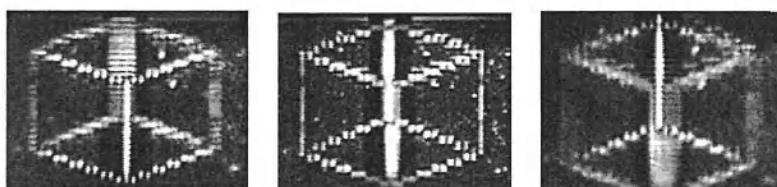


Figure 8-7. White light reconstructed images taken by CCD camera with different focal planes. Left: Focused in the front. Center: Focused in the middle. Right: Focused in the rear. (Hologram size: 1,280 x 1,024 pixels; Pixel pitch: 4 μm)

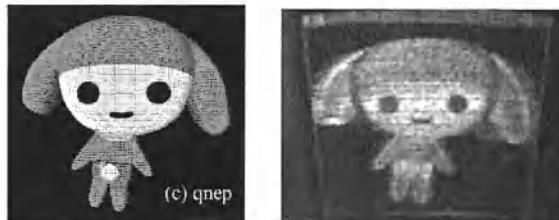


Figure 8-8. Monochrome reproduction of reconstructed image of full color rainbow CGH. Left: Perspective image of the object data. Right: Reconstructed image. (Hologram size: 4,096 x 4,096 pixels; Pixel pitch: 1.6 μm)

Figures 8-6 and 8-7 illustrate important 3D cues in the reconstructed images. Photographic reduction (see 4.1 for details) is used to materialize the holograms. The pixel numbers are 1,280 x 1,024 and the size of the hologram is about 5 mm x 4 mm. Therefore, pixel size of the hologram becomes roughly 4 μm square. The stereoscopic pair of the reconstructed image is shown in Fig. 8-6. Of course, since the CGRH is the hologram with horizontal parallax only, it displays continuous perspective changes when the viewer moves left and right. Figure 8-7 shows the focusing effect in the reconstructed image. These images are taken by the CCD camera with a different focus adjustment and show that the image has certain depth. Since the lens iris of the camera was set to the maximum diameter to emphasize the focusing effect, observation conditions are different from human vision.

A result of the full color reconstruction is shown in Fig. 8-8 as a monochromatic image. The hologram is recorded by the adapted CD-R writer (Sakamoto et al., 2004) described in section 4. The perspective image of the object is also shown.

3.2 Computer-generated rainbow holographic stereogram

We have applied holographic stereogram approach to calculate the object beam component for the CGRH and made full color computer-generated rainbow holographic stereogram (CGRHS) from an image array of 2D images (Yoshikawa and Sugawara, 1998).

3.2.1 Theory

The procedure to calculate the hologram is as same as CGRH described in 3.1. Only difference is calculation of the object beam $O(x)$, the complex amplitudes on the hologram plane caused by the object.

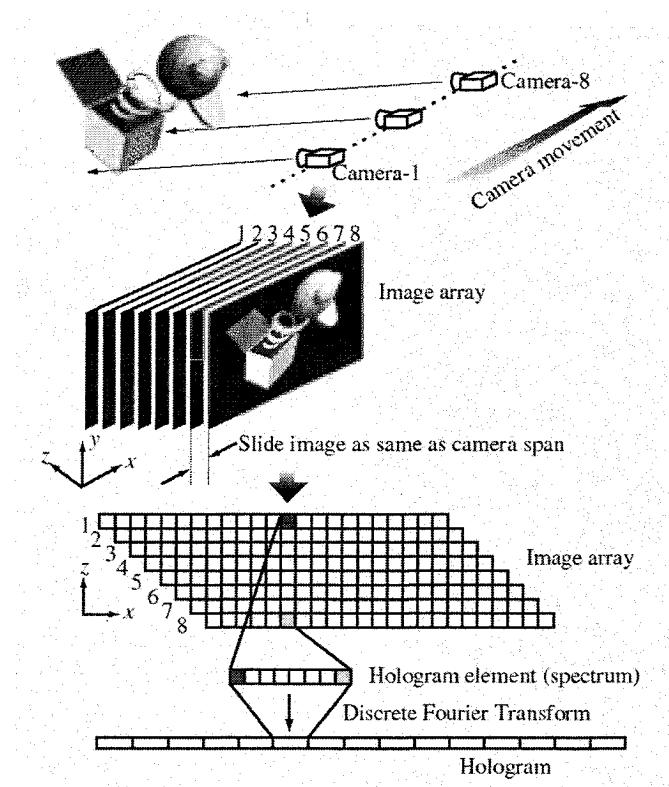


Figure 8-9. Procedure to transform an image array to generate image type holographic stereogram.

The first part to calculate CGRHS is almost same as that of conventional holographic stereogram (Yatagai, 1976; Yoshikawa and kameyama, 1995). However, since the reference beam is taken into account at the second step of the calculation, our result of the first step is obtained from only object beam. The procedure to obtain the subline, or 1-D hologram without reference term, is shown in Fig. 8-9. The subline is divided into elemental blocks to form the holographic stereogram (Yoshikawa and kameyama, 1995). Since it is the stereogram, an array of 2D images is required. This image array can be shot by moving camera, or generated as computer graphics from object data. The image array is stacked with sliding as same as a camera movement. Then the stack is sliced vertically to form the one-step image holographic stereogram (Benton, 1987). Since the sliced image can be used as spectrum data of the elemental block, the complex amplitude $O(x)$, caused by the object, is determined from the Fourier Transform of the sliced image.

The rest of the calculation is adding the off-axis reference beam with a vertical incident angle. This procedure is exactly same as mentioned in the section 3.1.

3.2.2 Results

We use photographic reduction to materialize computer-generated rainbow holographic stereograms. Fig. 8-10 shows the stereoscopic pair of the white light reconstructed images from the full color rainbow holographic stereogram, whose pixel numbers are $5,120 \times 4,096$. The size of the exposed hologram on the film is about 15 mm x 12 mm. Therefore, pixel size of the hologram becomes roughly $2.9 \mu\text{m}$ square. The recorded object is the cube colored with three different colors. Since the CGRHS is the hologram with horizontal parallax only, it displays perspective changes when viewer moves left and right. On the other hand, color change occurs when the viewer moves up and down.



Figure 8-10. Stereoscopic pair of white light reconstructed image from different viewing position. The object is the cube colored with three different colors. Left: View from left side. Right: View from right side. (Hologram size: $5,120 \times 4,096$ pixels; Pixel pitch: $2.9 \mu\text{m}$)

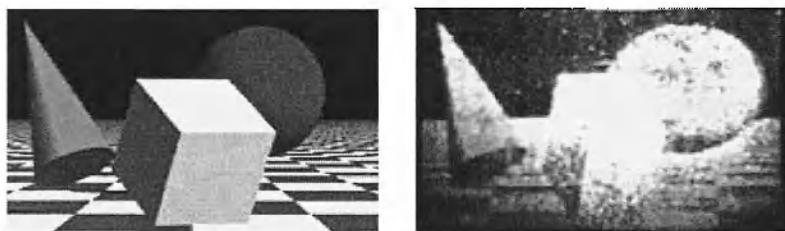


Figure 8-11. A white light reconstructed image of the full color holographic stereogram. The object consists of three primitive shapes (cone, cube and sphere) floating above a checkerboard. Left: One of the computer graphics. Right: Reconstructed image. (Hologram size: $10,240 \times 6,144$ pixels; Pixel pitch: $3.4 \mu\text{m}$)

Fig. 8-11 illustrates the white light reconstructed image from the full

color CGRHS. Pixel number of the hologram is 10,240 x 6,144 and the size is 35 mm x 21 mm. When observed about one meter apart from the hologram, one can recognize binocular parallax. The object consists of three shapes, cone, cube and sphere, floating above a checkerboard.

3.3 Hybrid hologram: Optical hologram made from CGH

The hybrid hologram is made from a computer-generated hologram, which is used as a master hologram and optically transferred to a second hologram (Wenzel et al., 1994). This hologram is originally proposed for comfortable observation with larger illumination angle than that of original CGH. The hybrid hologram is also useful to increase the viewing angle of the original CGH (Yoshikawa and Yamagishi, 1999). It is equivalent to reduce the pixel pitch of the hologram.

3.3.1 Theory

If the CGH is calculated with collimated reference beam, its viewing angle is same as its diffraction angle. Therefore, the viewing angle depends on the pixel pitch of the CGH. In contrast, the lens-less Fourier hologram (Stroke, 1965), recorded with a point reference source, makes the viewing angle independent from the pixel pitch.

Geometry to determine the viewing angle of the lens-less Fourier hologram is shown in Fig. 8.12. Note that the hologram is vertically off-axis and the reference point source is located over or under the object area. The viewing angle 2θ is proportional to hologram pixels P as expressed below (approximation is valid when the angle is small):

$$2\theta = 2 \tan^{-1} \left(\frac{Pd}{2r} \right) \cong \frac{P\lambda}{4D}, \quad (8-15)$$

where, D : Radius of an object area, P : A number of pixels in horizontal, d : Pixel pitch, r : Distance between hologram and center of object. Equation (8-15) also shows that reducing image size can also increase the viewing angle.

In the image reconstruction, the lens-less Fourier hologram requires a special illumination, a point light source that must be located near the hologram. Therefore, we made a secondary hologram that can be reconstructed with collimated white-light. The optical setup to make the secondary hologram is illustrated in Fig. 8-13. The holographic plate is located at the center of the reconstructed image from the master CGH to make image-plane hologram. Collimated reference beam is provided from

the same side of the object beam for transmission hologram or opposite side for reflection hologram. Rainbow hologram can be also produced when the master CGH is calculated with the proper manner (Yoshikawa and Kagotani, 2002).

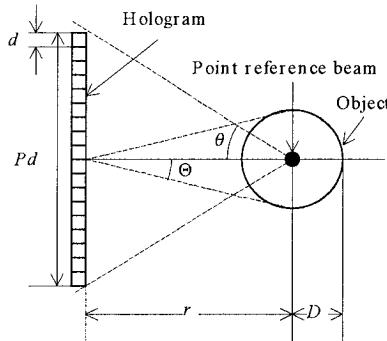


Figure 8-12. Top view of optical setup for recording hologram with point reference source.

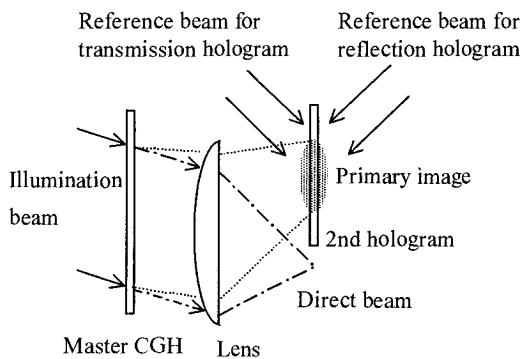


Figure 8-13. Side view of optical setup for recording an optical hologram of the reconstructed image from the CGH.

3.3.2 Results

A reconstructed image from the rainbow type hybrid hologram is shown in Fig. 8-14. The recorded object is a UFO consists of 1,526 point sources and its size on the second hologram is 3.0 mm x 4.0 mm. Pixel number of the master CGH is 3,200 x 2,200 and the pixel pitch is about 5.5 μm . The

horizontal viewing angle is 21 degrees. Since the master CGH is calculated as the rainbow hologram, it provides horizontal parallax only. However, discarding vertical parallax makes computational speed 20 times faster than the full parallax one. In addition, the reconstructed image becomes much sharper and brighter as same as optical rainbow holograms.

Figure 8-15 shows a reconstructed image from the reflection type multi-color hybrid hologram. The recorded object is two characters consists of 8,445 point sources and its size on the second hologram is 13.0 mm x 15.0 mm. The left character wears a red costume and the right character wears a green costume. Two master holograms (for red and green) are printed by the fringe printer described in section 4. Pixel number of the master CGH is 32,000 x 24,000 and the pixel pitch is about 1.3 μm . Each master hologram is transferred separately to the second hologram with red and green laser and the second holograms are stacked to provide a multi-color image. The horizontal viewing angle is 18 degrees. Since the master CGHs are calculated as the full parallax hologram, the second hologram also provides full parallax.

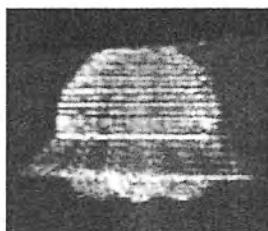


Figure 8-14. Reconstructed image from the hybrid rainbow hologram.

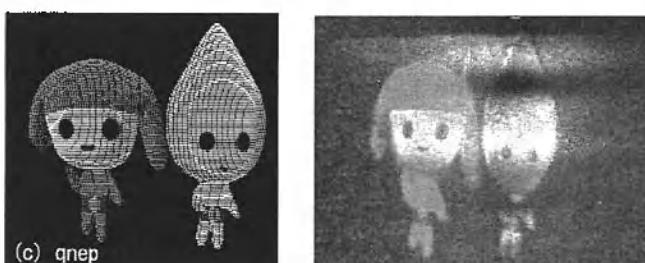


Figure 8-15. Reconstructed image from the hybrid multi-color reflection hologram.
Left: Perspective image of the object data. Right: Reconstructed image.

4. OUTPUT DEVICES OF COMPUTER-GENERATED HOLOGRAMS

There are many ways to obtain final hologram from computed data. This section describes some major method used for white light 3D CGH.

4.1 Photographic reduction

Photographic reduction is very simple and widely used from early CGH research (Waters, 1966). The computed hologram is printed on a paper or displayed on a video display. Then the image of the holographic fringe is taken by a photographic camera with high-resolution film. In combination with high-resolution lens and precise focus adjustment, pixel resolution of $2.9 \mu\text{m}$ was achieved (Yoshikawa and Sugawara, 1998). Many holograms in the section 3 were made with this technique.

4.2 Image setter

The “image setter” is an ultra high-resolution laser printer, whose resolution exceeds 4,000 dpi ($6.35 \mu\text{m}$). Since it is also possible to expose a transparent photographic film, the image setter is quite suitable for the output device of CGH. Although, the diffraction angle is limited, even a conventional laser printer can be used for output (Lee and Casasent, 1987).

4.3 Spatial light modulator

A high-resolution spatial light modulator (SLM) can be used as an output device of CGH. There are many publications using a liquid crystal panel as an SLM for the hologram reconstruction (Mok et al., 1986; Amako and Sonehara, 1990; Hashimoto et al., 1991). Accousto-optic device (St.-Hilaire et al. 1990) and digital micromirrors (Huebschman et al., 2005) can be also used for holographic displays. However, most of them use a laser for reconstruction, and a few use white light (Takano and Sato, 2002; Yamguchi and Yoshikawa, 2005).

A reconstructed image is shown in Fig. 8-16 (Yamguchi and Yoshikawa, 2005). The image type CGH is displayed on the reflective liquid crystal panel. The illumination light source is a white LED.



Figure 8-16. Reconstructed image from the image type hologram displayed on the liquid crystal panel. Left: Perspective image of the object data. Right: Reconstructed image. (Hologram size: 1,408 x 1,058 pixels; Pixel pitch: 10.4 μm)

4.4 Electron beam

The electron beam writer provides an excellent quality for 3D CGH recording (Hamano and Yoshikawa, 1998). Hamano and kitamura (2002) reported very bright full-color rainbow type CGH, whose size is 24 mm x 10 mm and 40 degrees of viewing angle. Since the equipment and running costs of the electron beam writer are very expensive, it is suitable for mass-production, but not for small lot or personal use.

4.5 Fringe printer

There have been reported many holographic printers. These holographic printers can be classified into two categories. The first category is “holographic stereogram printer” and the second is “holographic fringe printer.” The holographic stereogram printers (Yamaguchi et al., 1992; Klug et al., 1993) can be used only for holographic stereograms, which are made from an array of 2D images.

The holographic fringe printers (Ichioka et al., 1967; Yatagai et al., 1989; Noehre et al., 1994; Cable et al., 2003) print the fringe pattern itself and can be used for any kind of holograms. Although many fringe printers have been reported, most of them have very small pixel numbers such as 64 x 64 and not suitable for 3D display.

Sakamoto et al. (2004) adapted a CD-R writer for the fringe printer, and produced binary hologram up to 8,192 x 4,096 pixels within 20 minutes. This writer was used to record the computer-generated rainbow hologram shown Fig. 8-8. Their system is very compact, low cost and speedy. However, some problems remain such as; timing error between the tracks reduces image quality; maximum printing size is limited by the firmware.

Yamanaka and Matsushima (2004) developed a fringe printer consists of a laser diode and X-Y translation stage with stepper motors. Printed hologram shown in their paper has 8,192 x 4,096 pixels with 1.5 μm x 3 μm dot size. Since the holographic fringe is printed point-by-point, printing

speed is not so high. Yoshikawa and Tachinami (2005) developed a prototype holographic fringe printer consist of a laser, a spatial light modulator (SLM) and X-Y translation stage with stepper motors. Since the holographic fringe is printed area-by-area, total printing time can be made shorter than that of point-by-point printing. In their experimental result, CGH with 32,000 x 24,000 pixels and 2.8 μm of the pixel resolution was recorded.

5. CONCLUSION

This chapter describes the white light reconstruction 3-D computer-generated holograms.

The simplified model to calculate the computer-generated rainbow hologram has been described. The object to be recorded is approximated as a collection of self-illuminated and vertically truncated line sources. This model provided high speed computation and effective holographic data storage and/or transmission. The experimental results demonstrated that the described method is practical for the display of full color three-dimensional holographic images that can be reconstructed with the white light.

The stereographic approach to calculate the full color computer-generated rainbow hologram is described. Since the hologram is calculated from an array of 2D images, the image array can be shot by moving camera or generated as computer graphics. The experimental results demonstrated that the method is practical to display full color holographic stereogram that can be reconstructed with the white light.

The hybrid hologram that can enlarge the viewing angle of the computer-generated hologram is described. In the method, increasing pixel number of the holograms makes the viewing angle much wider.

Many ways to obtain final hologram from CGH are introduced. The electron beam writer gives best quality and suitable for mass production. Inexpensive and high quality output device is desired for small lot and personal use.

REFERENCES

- Amako, J. and Sonehara, T., 1990, CGH using TFT active matrix liquid crystal spatial light modulator, *Japan J. Appl. Phys.* **29**: L1533.
- Benton, S. A., 1969, Hologram reconstructions with extended incoherent sources, *J. Opt. Soc. Am.* **59**: 1545A.
- Benton, S. A., 1987, "Alcova" Holograms for computer-aided design, *Proc. Soc. Photo-Opt. Instrum. Eng.* **761**: 53-61.

- Bjelkhagen, H. I. and Vukicevic, D., 1994, Lippmann color holography in a single-layer silver-halide emulsion, *Proc. Soc. Photo-Opt. Instrum. Eng.* **2333**: 34-48.
- Brown, B. R. and Lohmann, A. W., 1969, Computer-Generated Binary Hologram, *IBM J. Res. Develop.* **13**: 160-168.
- Cable, A., Mash, P., and Wilkinson, T., 2003, Production of computer-generated holograms on recordable compact disk media using a compact disk writer, *Opt. Eng.* **42**(9): 2514-2529.
- Champagne, E. B., 1967, Nonparaxial imaging, magnification, and aberration properties in holography, *J. Opt. Soc. Am.* **57**(1): 51-55.
- Hamano, T. and Kitamura, M., 2000, Computer-generated holograms for reconstructing multi 3-D images by space-division recording method, *Proc. Soc. Photo-Opt. Instrum. Eng.* **3956**: 23-32.
- Hamano, T. and Yoshikawa, H., 1998, Image-type CGH by means of e-beam printing, *Proc. Soc. Photo-Opt. Instrum. Eng.* **3293**: 2-14.
- Hashimoto, N., Morokawa, S. and Kitamura, K., 1991, Real-time holography using the high-resolution LCVT-SLM", *Proc. Soc. Photo-Opt. Instrum. Eng.* **1461**: 291-302.
- Huebschman, M. L., Munjuluri, B., Hunt, J. and Garner, H. R., 2005, Holographic video display using digital micromirrors, *Proc. Soc. Photo-Opt. Instrum. Eng.* **5742**: 1-14.
- Ichioka, Y., Izumi, M. and Suzuki, T., 1967, Halftone plotter and its applications to digital optical information processing, *Appl. Opt.* **8**(12): 2461-2471.
- Jendral, A., Brauer, R. and Bryngdahl, O., 1995, Computer-generated image hologram for 3D display, *Optical Review* **2**(3): 187-188.
- Klug, M. A., Halle, M. W., Lucente, M. and Plesniak, W. J., 1993, A compact prototype one-step Ultragram printer, *Proc. Soc. Photo-Opt. Instrum. Eng.* **1914**: 15-24.
- Lee, A. J. and Casasent, D. P., 1987, Computer-generated hologram recording using a laser printer, *Appl. Opt.* **26**(1): 136-138.
- Leseberg, D. and Bryngdahl, O., 1984, Computer-generated rainbow holograms, *Appl. Opt.* **23**(14): 2441-2446.
- Leseberg, D., 1986, Computer-generated holograms: display using one-dimensional transforms, *J. Opt. Soc. Am. A* **3**(11): 1846-1850.
- Lesem, L. B., Hirsch, P. M. and Jordan Jr., J. A., 1969, The Kinoform: A New Wavefront Reconstruction Device, *IBM J. Res. Develop.* **13**: 150-155.
- Lucente, M., 1993, Interactive computation of holograms using look-up table, *J. Electron. Imaging* **2**(1): 28-34.
- Mok, F., Diep, J., Liu, H. and Psaltis, D., 1986, Real-time computer generated hologram by means of liquid-crystal television spatial light modulator, *Opt. Lett.* **11**(11): 748-750.
- Nochte, S., Dietrich, C., Fischer, G., Grosskopf, P., Lange, R., Manner, R. and Zhang, E., A High Quality Lithograph with Integrated Environment for Computer Generated holograms, 1994, *Topical meeting of the International Commission for optics*: 223.
- Rosen, L., 1966, Focused-image holography with extended sources, *Appl. Phys. Lett.* **9**: 337-339.
- Sakamoto, Y., Morishima, M and Usui, A., 2004, Computer-generated holograms on a CD-R disk, *Proc. Soc. Photo-Opt. Instrum. Eng.* **5290**: 42-49.
- St.-Hilaire, P., Benton, S. A., Lucente, M., Jepsen, M. L., J. Kollin, H. Yoshikawa and J. Underkoffler, 1990, Electronic display system for computational holography, *Proc. Soc. Photo-Opt. Instrum. Eng.* **1212**: 174-182.
- Stroke, G. W., 1965, Lensless Fourier-Transform Method for Optical Holography, *Appl. Phys. Lett.* **6**: 201-203.
- Takano, K. and Sato, K., 2002, Color electro-holographic display using a single white light source and a focal adjustment method, *Opt. Eng.* **41**(10): 2427-2433.

- Waters, J. P., 1966, Holographic image synthesis utilizing theoretical method, *Appl. Phys. Lett.* **9**: 405-407.
- Waters, J. P., 1968, Three-dimensional Fourier-Transform method for Synthesizing Binary Holograms, *J. Opt. Soc. Am.* **55**: 1284-1288.
- Wentzel, S., Jendral, A., Brauer, R. and Bryngdahl, O., 1994, Increase of the space-bandwidth product of computer-generated near-field holograms by optical means, *Opt. Lett.* **19**: 1352-1354.
- Yamanaka, S. and Matsushima, K., 2004, A New High-Resolution Printer for Digitally Synthetic Holograms (in Japanese), *J. of the Institute of Image Information and Television Engineers* **58**(11): 1665-1668.
- Yamaguchi, M., Ohya, N. and Honda, T., 1992, Holographic three-dimensional printer: new method, *Appl. Opt.* **31**(2): 217-222.
- Yamaguchi, T. and Yoshikawa, H., 2005, Interactive Holographic Television with Full-Parallax -Fast calculation by Image Hologram- (in Japanese), *HODIC Circular* **25**: 14-17.
- Yatagai, T., 1976, Stereoscopic approach to 3-D display using computer-generated holograms, *Appl. Opt.* **15**(11): 2722-2729.
- Yatagai, T., Camacho-Basilio, J.G.C. and Onda, H., 1989, Recording of computer-generated holograms on an optical disk master, *Appl. Opt.* **28**(6): 1042-1044.
- Yoshikawa, H., 2001, Fast Computation of Fresnel Holograms Employing Difference, *Optical Review* **8**(5): 331-335.
- Yoshikawa, H. and Kagotani, A., 2002, Full-color computer-generated rainbow hologram with enlarged viewing angle, *Optical Review* **9**(6): 251-254.
- Yoshikawa, H. and Kameyama, H., 1995, Integral holography, *Proc. Soc. Photo-Opt. Instrum. Eng.* **2406**: pp.226-233.
- Yoshikawa, H. and Sugawara, K., 1998, Computer generated rainbow holographic stereogram, *Topical meeting of the International Commission for Optics*: 380-390.
- Yoshikawa, H. and Tachinami, M., 2005, Development of direct fringe printer for computer-generated holograms, *Proc. Soc. Photo-Opt. Instrum. Eng.* **5742**: 259-266.
- Yoshikawa, H. and Taniguchi, H., 1996, Computer generated rainbow hologram, *Proc. Soc. Photo-Opt. Instrum. Eng.* **2885**: 26-33.
- Yoshikawa, H. and Taniguchi, H., 1999, Computer generated rainbow holograms, *Optical Review* **6**(2): 118-123.
- Yoshikawa, H. and Yamagishi, M., 1999, Enlarging viewing angle of computer-generated holograms, *Proc. Soc. Photo-Opt. Instrum. Eng.* **3749**: 611-612.

Chapter 9

WIDE-ANGLE COMPUTER-GENERATED HOLOGRAMS FOR 3-D DISPLAY

Toyohiko Yatagai and Yusuke Sando

Institute of Applied Physics

University of Tsukuba

Tsukuba, Ibaraki, Japan

yatagai@bk.tsukuba.ac.jp

Abstract A new method for synthesizing a full-color computer-generated hologram (CGH) of real-existing objects has been proposed. In this method, the synthesizing process of CGHs and the adjustments of magnifications for each wavelength are considered based on parabolic sampling of three-dimensional Fourier spectra. Our method requires only one-dimensional azimuth scanning of objects, does not require any approximations in the synthesizing process, and can perform efficient magnification adjustments required for color reconstruction. Experimental results have been presented to verify the principle and validity of this method. Then, we propose a fast calculation method for diffraction to nonplanar surfaces using the fast-Fourier transform (FFT) algorithm. In this method, the diffracted wavefront on a cylindrical surface is expressed as a convolution between the point response function and the spatial distribution of objects wherein the convolution is calculated using FFT. The principle of the fast calculation and the simulation results are presented.

Keywords: Computer holography, Fourier optics, 3-D display, Computer-generated hologram, Digital holography

1. INTRODUCTION

A computer-generated hologram (CGH) is one of the most important techniques for three-dimensional (3-D) imaging since it can yield ideal 3-D visual effects even for virtual 3-D objects. Some encoding methods for synthesizing CGHs were proposed[1–3]. In many methods, an object was usually considered to consist of many cross sections. The contributions from the cross sections at the hologram plane are calculated and then summed up these con-

tributions so as to obtain the light amplitude propagating from the object to the hologram plane. Waters[4] has proposed the use of overlapping Fresnel zone plates for CGH. Yatagai[5] has discussed the stereoscopic approach in CGHs to obtain large view fields. Ren *et al* have proposed a binary CGH with four-square apertures which can reconstruct four different images in the four different directions[6]. Yang *et al.* studied encoding of complex information into one phase-type CGH with area division and phase retardation [7]. The CGHs for 3-D display and a method to enlarge a viewing-angle have been discussed[8]. In this method, high-order diffracted lights by pixel structures of a display device are used for reconstructing objects. Hamano *et al* [9] have discussed CGHs for color-reconstruction where white light reconstruction is possible at the cost of vertical parallax.

However, it is extremely difficult to synthesize a CGH for real-existing 3-D objects because the 3-D object's information is necessarily required. Recording of projection images is one of the most promising techniques for outdoor recording because it can be performed using incoherent white light and it is not sensitive to external vibration. Multiple projection images from different view points are required to obtain the 3-D object's information. Yatagai [5] utilized projection process for synthesizing CGHs. However, an auto-stereoscopic display and not a CGH was fabricated at the final step.

In recent years, Abookasis *et al.* [10] and Sando *et al.* [11] proposed some methods for synthesizing a CGH from multiple projection images. These methods have potential for outdoor recording of 3-D objects and it is very easy to extend their methods to full-color reconstruction by using a color CCD camera [12]; however, in principle, they require two-dimensional(2-D) mechanical scanning of 3-D objects in their methods. This involved enormous time for recording and precise and complicated control of a CCD camera.

Although the proposed method can reduce computation time for synthesizing wide-angle holograms, enormous amount of time and memory are still required for calculating CGHs. Efficient algorithms for calculating diffraction is necessary for reducing the calculation time. The fast-Fourier transform (FFT) algorithm enables this reduction. Based on FFT, various types of algorithms for calculating diffraction have been developed [13]. Yoshikawa *et al.* [14] have proposed a fast calculation method for large size holograms by interpolation .

Most algorithms using FFT are effective only under the condition that both the input and observation surfaces are finite planes that are parallel to each other. Some authors proposed fast calculation methods using FFT that can be applied to the case where the input plane is not parallel to the observation plane [15, 16]. These methods are very useful for the calculation of reconstructed images observed from different points of view [17, 18]. However, since the observation surfaces, in any of the methods, are assumed to be planes, very high-resolution display devices are necessary to enlarge the viewing an-

gles. Moreover, although such devices are available, the reconstructed images cannot be observed from the opposite side of a hologram.

A remarkable technique developed to achieve 360° field of view is 360° holography [19]. If a computer-generated 360° hologram is synthesized in a computer, a numerical simulation of the diffraction on the nonplanar observation surfaces is required. Rosen [20] synthesized a CGH on a spherical observation surface, but this method does not yield a 360° field of view because the origin of the object does not correspond to the center of the observation sphere, and considerable computing time is required since the FFT algorithm cannot be applied in this method. To our knowledge, there exist no reports on the methods that enable the fast calculation of diffraction on spherical or cylindrical observation surfaces.

In this chapter, at first, we have proposed an efficient method for synthesizing a CGH from projection images. The synthesizing process of CGHs is considered geometrically in the 3-D Fourier space. This reduces redundant 2-D scanning of objects [10] into essential one-dimensional (1-D) azimuth scanning for the recording process. In addition, an efficient adjustment method of magnifications is also proposed. In order to verify the effectiveness of this method, experimental results have been demonstrated with actually recorded projection images.

Next, we have proposed such a fast calculation method based on the convolution theorem using the FFT algorithm. In this method, both the object and observation surfaces are cylindrical and concentric about the cylindrical axis. Therefore, this method yields a 360° field of view.

2. HOLOGRAM SYNTHESIS BY 3-D FOURIER SPECTRUM

The method proposed here mainly consists of three processes. In the first process, projection images are recorded (in this report, synthesized using a computer for ease) with a color CCD scanning two-dimensionally and parts of the 3-D Fourier spectrum of objects are calculated in accordance with the principle of CT[21]. The Fourier components required to generate a CGH are extracted from the 3-D Fourier spectrum in the second process. Finally, a Fresnel hologram is synthesized from it directly.

2.1 Relation between object waves and 3-D Fourier spectrum

A virtual optical system for recording 3-D objects is shown in Fig. 9-1. We assume that 3-D object surfaces reflect waves emerging from external light sources isotropically. In general, the spatial reflectivity distribution of 3-D objects is expressed as complex values. However, we have to treat the spatial

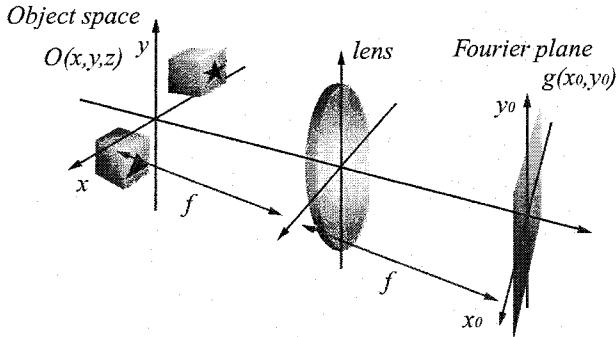


Figure 9-1. A virtual optical system for 3-D objects.

phase distribution of the reflectivity as spatially unvaried because a CCD can not detect the phase information without interferometer. The 3-D spatial distribution with spatially unvaried phase is represented by $O(x, y, z)$, which corresponds to the square root of the intensity. The complex wavefronts reflected on the objects are observed in the Fourier plane in Fig. 9-1. The distribution $g(x_0, y_0)$ in the Fourier plane is expressed as follows [11]:

$$g(x_0, y_0) = \iiint O(x, y, z) \exp \left\{ -\frac{i2\pi}{\lambda} \left[\frac{x_0 x + y_0 y}{f} - \frac{(x_0^2 + y_0^2)z}{2f^2} \right] \right\} dx dy dz, \quad (9-1)$$

where λ and f are the virtual wavelength of incident light and the focal length, respectively, of the lens introduced in Fig. 9-1.

Our principle is based on the relation between $g(x_0, y_0)$ and the 3-D Fourier spectrum of the 3-D distribution $O(x, y, z)$. The relation is revealed after substituting $u = x_0/\lambda f$ and $v = y_0/\lambda f$ in Eq. (9-1).

$$\begin{aligned} g(u, v) &= \iiint O(x, y, z) \exp \left\{ -i2\pi \left[ux + vy - \frac{\lambda}{2}(u^2 + v^2)z \right] \right\} dx dy dz \\ &= \left\{ \iiint O(x, y, z) \exp[-i2\pi(ux + vy + wz)] dx dy dz \right\} \Big|_{w=-\lambda(u^2+v^2)/2} \\ &= \mathcal{F}[O(x, y, z)]|_{w=-\lambda(u^2+v^2)/2}, \end{aligned} \quad (9-2)$$

where $\mathcal{F}[\cdot]$ denotes a 3-D Fourier-transform operator. Subscripts in Eq. (9-2) represent a paraboloid of revolution in 3-D Fourier space (u, v, w) . Consequently, we can determine that the wavefront distribution at the Fourier plane in Fig. 9-1 is completely identical to components on the paraboloid of revolution in the 3-D Fourier space of $O(x, y, z)$ without any approximations, unlike other similar methods [10–12].

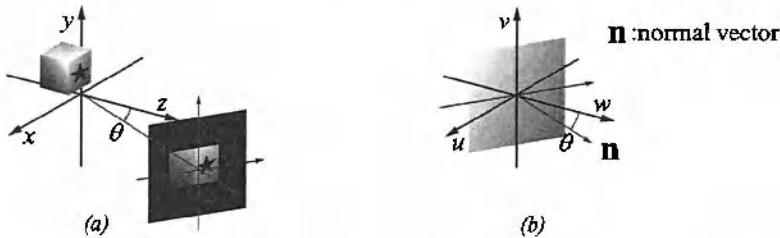


Figure 9-2. Schematics of the principle of 3-D CST. (a) Orthogonal projection in the real space and (b) a sectional plane in the 3-D Fourier space obtained from a projection image.

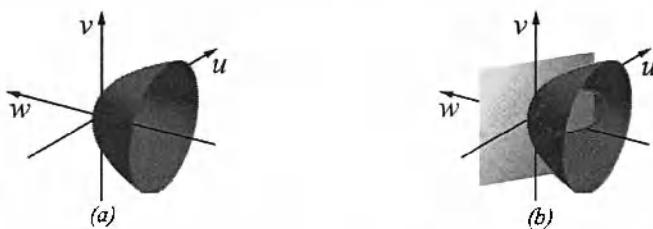


Figure 9-3. Paraboloid of revolution. (a) Components identical to objects waves and (b) intersections between the paraboloid of revolution and a sectional Fourier plane.

2.2 Extraction method for paraboloid of revolution

Equation (9-2) implies that indirect acquisition of the distribution $g(u, v)$ becomes possible by direct access to the 3-D Fourier space of $O(x, y, z)$. To achieve this, the principle of the 3-D central slice theorem (CST) is essential [21]. The principle of 3-D CST ensures that partial components of the 3-D Fourier spectrum of a 3-D object are obtained from an orthogonal projection image of the object. In this principle, at first, the 3-D object is projected onto a plane whose normal vector is inclined by θ to the z-axis on the z-x plane. Subsequently the projection image is 2-D Fourier-transformed. The 2-D Fourier spectrum then corresponds with a sectional Fourier plane whose normal vector is inclined by θ to the w-axis on the w-u plane in the 3-D Fourier space of the object, as shown in Fig. 9-2. Therefore, it is possible to obtain partial components on the paraboloid of revolution represented by Eq. (9-2) from one projection image using this principle, as shown in Fig. 9-3.

Figure 9-3(a) shows the components represented by Eq. (9-2). Only intersections between the sectional Fourier plane in Fig. 9-2(b) and the paraboloid of revolution in Fig. 9-3(a) can be extracted from the sectional Fourier plane. The intersections are calculated by solving the following simultaneous equa-

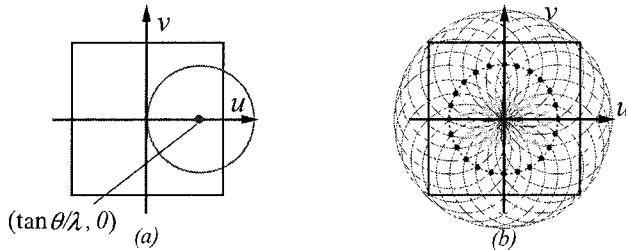


Figure 9-4. Extractive area on the u - v plane from (a) one projection image and (b) a series of projection images.

tions:

$$w \cos \theta + u \sin \theta = 0 \quad (9-3)$$

$$w = -\frac{\lambda}{2}(u^2 + v^2). \quad (9-4)$$

Equations (9-3) and (9-4) represent the planar equation shown in Fig. 9-2(b) and the equation of the paraboloid of revolution shown in Fig. 9-3(a), respectively. This simultaneous equations give the following solution:

$$\left(u - \frac{\tan \theta}{\lambda} \right)^2 + v^2 = \left(\frac{\tan \theta}{\lambda} \right)^2, \quad w = -u \tan \theta. \quad (9-5)$$

This solution shows that the intersections between the sectional Fourier plane and the paraboloid of revolution form an ellipse on the sectional Fourier plane, which is extracted components from one projection, as expressed with a red line in Fig. 9-3(b). Moreover, from this solution it is found that the projection of the ellipse onto the u - v plane in the 3-D Fourier space becomes a circle with a radius $\tan \theta/\lambda$, shown in Fig. 9-4(a). The position of the center and the radius depend on the direction of projection. On the other hand, to obtain all the components on the paraboloid of revolution is equal to filling the u - v plane with red circles. In order to fill the two-dimensional (2-D) u - v plane, 2-D scanning of 3-D objects is not necessarily needed. 1-D scanning is enough because 1-D components on the 2-D u - v plane are extracted from only one projection. Although there are some scanning methods to accomplish this, the diagram shown in Fig. 9-5 indicates the best scanning method in terms of the feasibility of a recording optical system. In this diagram, tangential θ and azimuth ϕ determine the radius of the extractive circle and the azimuth position of its center, respectively, ($\phi = 0$ in Fig. 9-4(a)). 3-D objects are imaged onto a CCD plane by an imaging lens. This projection is not exactly an orthogonal one; however, it can be approximated as an orthogonal projection in the case that the distance from the origin to the CCD camera is considerably longer than the depth of the

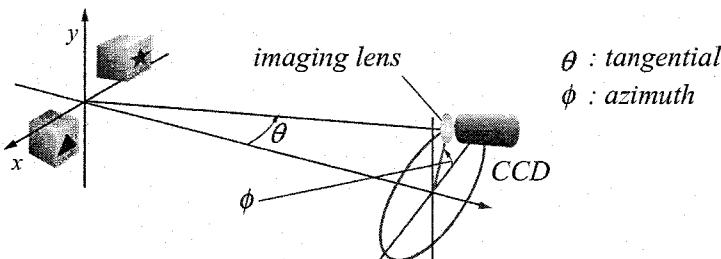


Figure 9-5. A recording optical system.

3-D objects. The CCD camera records the intensity of such projection images by revolving around the z-axis. Since the locus of the CCD camera and the Fourier components shown in Fig. 9-3(a) are in rotational symmetry to the z- and w- axes, respectively, the extractive components from a series of projection images recorded by this system can fill the u-v plane with red circles, as shown in Fig. 9-4(b). Hence, it is found that this scanning method can provide all the components on the paraboloid of revolution and requires only 1-D azimuth scanning of objects, unlike other similar methods [10–12].

2.3 Adjustments of magnifications for each wavelength

To extend this method to color reconstruction is rather simple provided that a color CCD camera is used. Sando *et al.* have already proposed such a full-color CGH [12]. In full-color reconstruction, three light sources with wavelengths of λ_R , λ_G , and λ_B corresponding to red, green, and blue are essential. Moreover, adjustments of magnifications for each wavelength are required. These were performed by Sando *et al.* by changing the movement area of the CCD camera for each wavelength. However, this adjustment method is not applicable to the method proposed here because the locus of a CCD camera is fixed for every wavelength. In this method, the adjustment of the magnification of the z-axis, whose diagram is shown in Fig. 9-6, is performed by changing a radius of an extractive circle for each wavelength. U and V denote maximum spatial frequencies that depend not on wavelengths but on the size and pixel number of a projection image. Therefore, all the three color components are obtainable by a single one azimuth scanning using a color CCD camera. The magnifications of the x- and y- axes are fixed to one [12].

2.4 Reconstruction of color 3-D images

In order to verify the principle described above, we have demonstrated an experiment. A total of 90 projection images were recorded with a recording optical system proposed in Fig. 9-5. Some typical examples are shown in Fig.

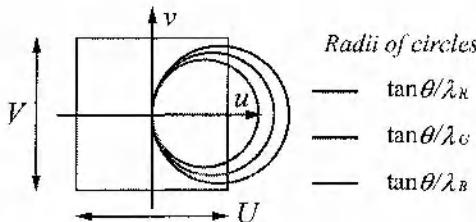


Figure 9-6. Adjustments of magnifications.

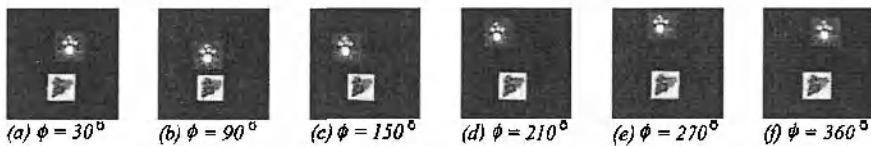


Figure 9-7. Color projection images at $\theta = 17^\circ$.

9-7. These images are divided into color components and each component is binarized. The noise component is also removed from them. The size and pixel numbers of each image are $1 \times 1\text{cm}$ and $256 \times 256\text{pixels}$, respectively. A grape and a mushroom on square planes are located at $z \approx 0\text{mm}$ and $z \approx 4.9\text{mm}$ in the object space, respectively. Three wavelengths, 632.8nm, 514.5nm, and 488.5nm, corresponding to red, green, and blue, respectively, are used for full-color reconstruction. The angle between the optical axis of a CCD camera and the z-axis is $\theta \approx 17^\circ$. Under these conditions, the magnification of the z-axis is approximately 53 [12]. Therefore, the grape and the mushroom should be reconstructed at $z \approx 0\text{cm}$ and $z \approx 26\text{cm}$, respectively. According to the principle proposed here, the distributions at the Fourier plane in Fig. 9-1 are synthesized from the above projection images. Thus, the reconstructed images can be easily obtained by calculating back propagation from the distribution at the Fourier plane onto arbitrary sectional planes in the objects space. These procedures are performed for each color component and the three color components are then superimposed at the final step. Such reconstructed images are shown in Fig. 9-8. As can be observed from Fig. 9-8, each of the two different objects can be clearly reconstructed at each corresponding position. The adjustments of magnifications for each wavelength are also successful. Thus, it is verified that this method can reconstruct 3-D full-color objects from substantially smaller number of projection images as compared to the previous method [12].

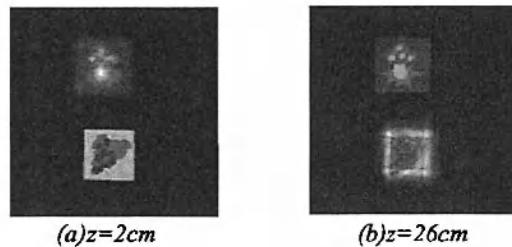


Figure 9-8. Numerical reconstructed images.

3. HOLOGRAM SYNTHESIS IN POLAR COORDINATES

3.1 Relation between the object and observation surfaces

To achieve fast calculation, we employ the FFT algorithm to calculate the convolution. Thus, the diffracted wavefront on the observation surface must be expressed in the form of a convolution with the distribution on the object surface. In case of a computer-generated 360° hologram, the shape of the observation surface is usually cylindrical. Therefore, the shape of the object surface must be also cylindrical to form a convolution between them, and both the surfaces must be concentric. The schematic of the geometrical relation between them is shown in Fig. 9-9, where r and R denote the radii of the

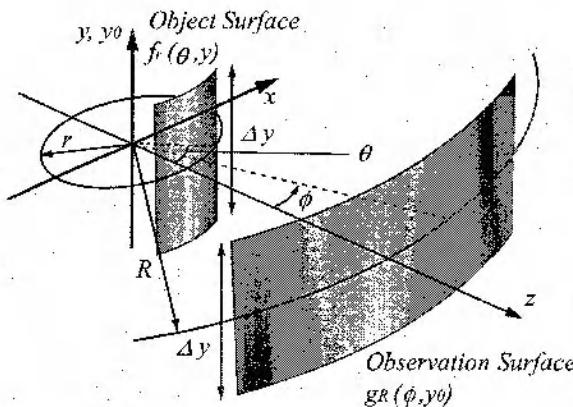


Figure 9-9. Schematic of the geometrical relation between the object and observation surfaces.

object and observation surfaces, respectively. The distributions on the object and observation surfaces are represented by $f_r(\theta, y)$ and $g_R(\phi, y_0)$, respectively, as shown in Fig. 9-9. Based on the assumption that the object is the aggregate

of many point light sources, the distribution $g_R(\phi, y_0)$ is the summation of the spherical wavefronts that emerge from the point light sources, and it is given by

$$g_R(\phi, y_0) = C \iint \frac{f_r(\theta, y) \exp(ikL)}{L} d\theta dy, \quad (9-6)$$

with

$$L = [R^2 + r^2 - 2Rr \cos(\theta - \phi) + (y - y_0)^2]^{1/2}, \quad (9-7)$$

where k and C denote the wavenumber of the incident light and a constant, respectively, and L represents the distance between two points. Here, the point response function (PRF) is defined as

$$h(\theta, y) = \frac{\exp[ik(R^2 + r^2 - 2Rr \cos \theta + y^2)^{1/2}]}{(R^2 + r^2 - 2Rr \cos \theta + y^2)^{1/2}}. \quad (9-8)$$

Then, Eq. (9-6) is expressed in the form of a two-dimensional convolution as follows:

$$\begin{aligned} g_R(\phi, y_0) &= C \iint f_r(\theta, y) h(\theta - \phi, y - y_0) d\theta dy \\ &= C f_r * h(\phi, y_0), \end{aligned} \quad (9-9)$$

where $*$ denotes the convolution integral. Therefore, fast acquisition of the diffracted wavefront on the cylindrical observation surface becomes possible by calculating Eq. (9-9) with an FFT based on the convolution theorem.

However, in case of diffraction for a 3-D object, the 3-D object is defined in cylindrical coordinates. Therefore, the radius of the object surface r must be treated not as a constant but a variable, and the PRF expressed in Eq. (9-8) becomes dependent on r . As a result, the diffracted wavefront on the observation surface is given as the superposition of all the wavefronts diffracted from every cylindrical object surface.

3.2 Spectral band widths

It is impossible to analytically Fourier transform $h(\theta, y)$ in Eq. (9-8). Therefore, it is necessary to numerically Fourier transform $h(\theta, y)$. In order to achieve this, the sampling theorem should be satisfied in both the azimuthal, θ , and the vertical, y , directions during the discretizing process. Therefore, the spectral band width of each direction is discussed here.

3.3 The azimuthal direction

The local spatial frequency $v_\theta(\theta, y)$ in the azimuthal direction at the local point (θ, y) is calculated as follows:

$$v_\theta(\theta, y) = \frac{1}{2\pi} \frac{\partial h(\theta, y)}{\partial \theta}, \quad (9-10)$$

where it is assumed that the variation in $(R^2 + r^2 - 2Rr \cos \theta + y^2)^{-1/2}$ is much less than that in $\exp[ik(R^2 + r^2 - 2Rr \cos \theta + y^2)]$. After simple calculations, the maximum of $v_\theta(\theta, y)$ and the spectral band width W_θ are given by

$$|v_\theta|_{max} = \frac{k\mathcal{L}}{2\pi}, \quad W_\theta = \frac{k\mathcal{L}}{\pi}, \quad (9-11)$$

where \mathcal{L} represents the smaller length of r and R .

3.4 The vertical direction

Based on the same assumption as in the case of the azimuthal direction, the local spatial frequency $v_y(\theta, y)$ in the vertical direction is calculated as follows:

$$v_y(\theta, y) = \frac{1}{2\pi} \frac{\partial h(\theta, y)}{\partial y}. \quad (9-12)$$

Similarly the maximum of $v_y(\theta, y)$ and the spectral band width W_y are given by

$$|v_y|_{max} = \frac{k\Delta y}{4\pi} \frac{1}{[(R-r)^2 + (\Delta y/2)^2]^{1/2}}, \quad W_y = \frac{k\Delta y}{2\pi} \frac{1}{[(R-r)^2 + (\Delta y/2)^2]^{1/2}}, \quad (9-13)$$

where Δy represents the height of the object surface, which is the same as that of the observation surface, as shown in Fig. 9-9.

3.5 Aliasing in the azimuthal direction

$f_r(\theta, y)$ and $h(\theta, y)$ are periodic functions with period 2π in the azimuthal direction. Thus, the analytical convolution between them is also a periodic function with the same period. On the other hand, in case of the numerical convolution between them based on the FFT, aliasing caused by the digitization is significant. However, fortunately, aliasing error can be avoided, but only in the azimuthal direction. Aliasing arises from the superposition of identical signals with different parallel shifts, where the amount of the parallel shift is an integer multiple of the azimuthal size of $f_r(\theta, y)$ and $h(\theta, y)$. Hence, it is possible to match the amount of the parallel shift with integer multiples of the period of f_r and h , namely 2π . Therefore, the aliasing error is completely avoided in the azimuthal direction.

3.6 Numerical experiments

3.6.1 The spectral band widths of point response function. Figure 9-10 shows the amplitude distribution of the Fourier spectrum of PRF calculated numerically from Eq. (9-8). The values of r and R are fixed to 1 cm and 10 cm, respectively and that of Δy is 10 cm. In the range of optical wavelengths, very large pixel sizes are required to satisfy the sampling theorem

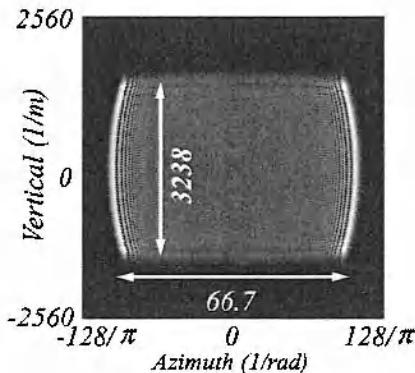


Figure 9-10. Amplitude distribution of the Fourier spectrum of point response function.

under these conditions. Hence, the wavelength is set to $300\mu\text{m}$, which is in the tera-hertz region. Then, the pixel number in both the azimuthal and vertical directions is 512. Thus, the angular sampling pitch of both θ and ϕ is given by $2\pi/512$. If a CGH were designed with a wavelength of 632.8 nm , the number of pixels required would be $N \cong 2.0 \times 10^5$. On the other hand, according to Eqs. (9-11) and (9-13), the spectral band widths in the azimuthal and vertical directions are estimated to be 66.7 rad^{-1} and 3238 m^{-1} , respectively. These numerical values correspond to the spectral widths shown in Fig. 9-10.

3.6.2 Cylindrical observation of Young's fringe. We have simulated the simplest example of our method, referred to as Young's experiment, where only two point light sources are distributed on the object surface. The two point sources are located at the same position in the vertical direction, while they are located at $\theta = -\pi/32$ and $\pi/32$, respectively, in the azimuthal direction. The values of r , R , Δy , and the wavelength are the same as those in section 3.6.1. The interference patterns on the cylindrical observation surface are shown in Fig. 9-11.

Here, the pitches of the fringes vary with the azimuth angles. The calculation time required in our method is 0.9 s. On the other hand, when the same simulation is demonstrated by the direct method used to directly calculate Eq. (9-6), the calculation time exceeds 9×10^4 s. Therefore, our method reduces the calculation time by 10^5 times that achieved by the direct method, thus enabling faster calculation of the diffraction on the cylindrical surface.

3.6.3 Reconstructed images from the cylindrical CGH. Other examples of our method are presented here. The object used for this simulation is shown in Fig. 9-12. The parameters used for this simulation are the same as

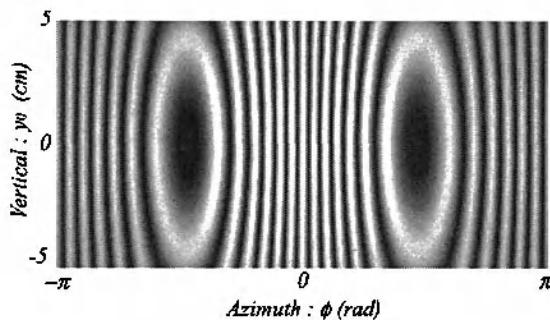


Figure 9-11. The interference patterns on the cylindrical surface.

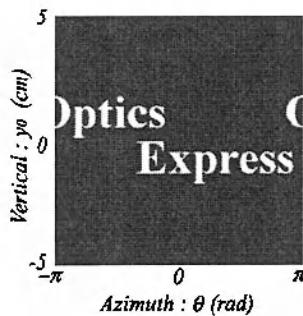


Figure 9-12. Object distribution on the cylindrical surface.

those used for the previous example. According to our method, a CGH consisting of complex values on the cylindrical surface is synthesized from Fig. 9-12. Then, the reconstructed images on the sectional planes are calculated from the CGH by the direct method. The schematic for the reconstruction is shown in Fig. 9-13. The position of the light source is virtually set to the origin to produce a spherical wave, which is uniformly illuminated on the cylindrical CGH. Figure 9-14 shows the reconstructed images in the sectional planes of (a) $z = 1$ cm and (b) -1 cm. Since the radius of the cylindrical object surface is $r = 1$ cm, the character located in the vicinity of $\theta = 0$ in Fig. 9-12, namely "x", is in focus in the sectional plane of $z = 1$ cm, as shown in Fig. 9-8(a); further, the character "O", which is located in the vicinity of $\theta = \pi$ in Fig. 9-12, is in focus in the sectional plane of $z = -1$ cm, as shown in Fig. 9-8(b). Therefore, these results verify our method.

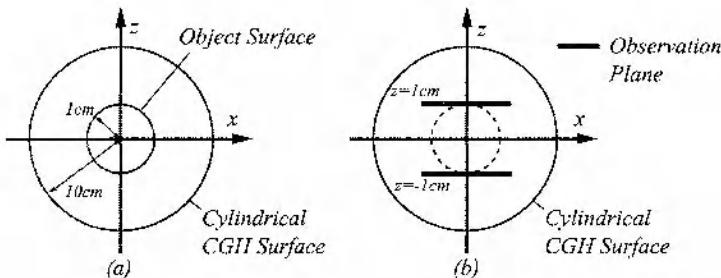


Figure 9-13. Top views of the geometrical relations (a) between the object surface and the CGH surface and (b) between the CGH surface and the observation plane.

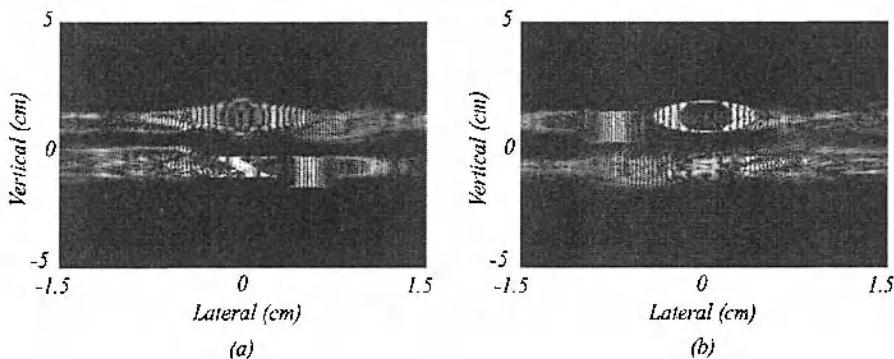


Figure 9-14. The reconstructed images in the sectional planes of (a) $z = 1\text{ cm}$ where "x" is in focus and (b) $z = -1\text{ cm}$ where "O" is in focus.

4. CONCLUSION

An efficient method for synthesizing full-color CGHs from multiple projection images has been proposed. In this method, holographic patterns are calculated using 3-D Fourier spectra of 3-D objects. The proposed method requires only 1-D mechanical scanning in an object space, unlike our previous method and does not require any approximations. Therefore, this method is superior to other similar methods in terms of efficiency of projection data, quality of reconstructed images, and practical feasibility of the recording system. Experimental demonstration has verified the principle and the validity of this method.

A method for the fast calculation of diffraction on cylindrical surfaces has been proposed in this paper. The algorithm of our method and the analysis of the spectral band widths are discussed. The simulation of Young's fringe is demonstrated as the simplest example of our method. The reconstructed im-

ages in the sectional planes from the cylindrical CGH are also presented. The calculation time required in our method is 10^5 times less than that of the direct method. Therefore, our method is extremely useful in creating computer-generated 360° holograms.

REFERENCES

- [1] A. W. Lohmann and D. P. Paris, "Binary Fraunhofer holograms, generated by computer," *Appl. Opt.*, **6**, 1739–1748 (1967).
- [2] B. R. Brown and A. W. Lohmann, "Computer-generated binary holograms," *IBM J. Res. Develop.*, **13**, 160–168 (1969).
- [3] L. B. Lesem, P. M. Hirsch and J. A. Jordan, Jr., "Computer synthesis of holograms for 3-D display", *Commun. ACM.*, **11**, 661–674 (1968).
- [4] J. P. Waters, "Holographic image synthesis utilizing theoretical methods", *Appl. Phys. Lett.*, **9**, 405–407 (1966).
- [5] T. Yatagai, "Stereoscopic approach to 3-D display using computer-generated holograms," *Applied Optics* **15**, 2722–2729 (1976).
- [6] C. H. Ren and L. Zhou and W. Q. Gao, "Four-channel self-focus computer-generated hologram", *Appl. Opt.*, **36**, 8844–8847 (1997).
- [7] M. Yang and J. Ding, "Area encoding for design of phase-only computer-generated holograms," *Opt. Commun.* **203**, 51–60 (2002).
- [8] T. Mishina, M. Okui, and F. Okano, "Viewing-zone enlargement method for sampled hologram that uses high-order diffraction," *Applied Optics* **41**, 1489–1499 (2002).
- [9] T. Hamano and M. Kitamura, "Computer-generated holograms for reconstructing multi 3-D images by space-division recording method", *Proc. SPIE*, **3956**, 23–32 (2000).
- [10] D. Aboekasis and J. Rosen, "Computer-generated holograms of three-dimensional objects synthesized from their multiple angular viewpoints." *J. Opt. Soc. Amer., A* **20**, 1537–1545 (2003).
- [11] Y. Sando, M. Itoh and T. Yatagai, "Holographic three-dimensional display synthesized from three-dimensional Fourier spectra of real existing objects," *Opt. Lett.* **28**, 2518–2520 (2003).
- [12] Y. Sando, M. Itoh, and T. Yatagai, "Color computer-generated holograms from projection images," *Opt. Express* **12**, 2487–2493 (2004).
- [13] D. Leseberg, "Sizable Fresnel-type hologram generated by computer," *J. Opt. Soc. Amer.*, **6**, 229–233 (1989).
- [14] N. Yoshikawa, M. Itoh, and T. Yatagai, "Interpolation of reconstructed image in Fourier transform computer-generated hologram," *Opt. Commun.*, **119**, 33–40 (1995).
- [15] D. Leseberg and C. Frère, "Computer-generated holograms of 3-D objects composed of tilted planar segments," *Appl. Opt.*, **27**, 3020–3024 (1988).
- [16] T. Tommasi and B. Bianco, "Computer-generated holograms of tilted planes by a spatial frequency approach," *J. Opt. Soc. Amer.*, **10**, 299–305 (1993).

- [17] Y. Takaki and H. Ohzu, "Hybrid Holographic Microscopy: Visualization of Three-Dimensional Object Information by use of Viewing Angles," *Appl. Opt.*, **39**, 5302–5308 (2000).
- [18] L. Yu, Y. An and L. Cai, "Numerical reconstruction of digital holograms with variable viewing angles," *Opt. Express*, **10**, 1250–1257 (2002).
- [19] O. D. D. Soares and J. C. A. Fernandes, "Cylindrical hologram of 360° field of view," *Appl. Opt.*, **21**, 3194–3196 (1982).
- [20] J. Rosen, "Computer-generated holograms of images reconstructed on curved surfaces," *Appl. Opt.*, **38**, 6136–6140 (1999).
- [21] M-Y. Chiu, H. H. Barrett and R. G. Simpson, "Three-dimensional reconstruction from planar projections," *J. Opt. Soc. Amer.* **70**, 755–762 (1980).

Chapter 10

HORIZONTAL-PARALLAX-ONLY OPTICAL SCANNING HOLOGRAPHY

Ting-Chung Poon

Optical Image Processing Laboratory, Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, Virginia 24061

Abstract: The principle of optical scanning holography (OSH) has been recently proposed by Poon et al. to acquire horizontal-parallax-only (HPO) holographic information electronically. Although many ideas of HPO-holography have been proposed and studied, to the best of our knowledge, this is the first proposed electronic holographic technique to acquire HPO-holographic information. In this Chapter, we first briefly review OSH and then discuss how HPO-holographic information can be acquired using OSH. We provide computer simulations to illustrate and clarify the idea. We also propose and discuss a possible optical reconstruction scheme for twin-image elimination using only two holograms as well as a digital scheme for HPO-holographic reconstruction.

Key words: Electronic/digital holography, optical scanning holography, horizontal-parallax-only holographic information, optical heterodyning

1. INTRODUCTION

The first television (TV) transmission of a hologram was demonstrated in 1966 [1]. A TV camera was used to record the interference between the Fresnel diffraction pattern of an object transparency and an off-axis plane wave. The holographic information was then transmitted over a closed-circuit TV and displayed on a 2-D monitor. The displayed 2-D record was then photographed to form a hologram, which was reconstructed subsequently by a coherent optical system. Since a hologram has a tremendous amount of information, information reduction techniques have

been investigated in order to alleviate the problems associated with the restricted field of view upon holographic reconstruction. In addition, if holographic information is to be refreshed at a TV rate, information reduction techniques can lower data transmission rate through the channel linking the holographic acquiring end and the holographic reconstruction/display end. Indeed live 3-D TV using hologram formation and transmission is a formidable problem. Nevertheless, much progress has been made [2-13] and novel devices have been invented [14-16]. Recently, Poon has proposed a real-time holographic TV system using optical scanning holography (OSH) to acquire holographic information electronically and employing spatial light modulators (SLMs) for eventual coherent 3-D display [17]. Using OSH does not suffer from the limited spatial resolution of the TV camera used for acquiring holographic information because in OSH, a photomultiplier is used (not a TV camera as in conventional electronic holographic systems) for collecting holographic information. Also, twin-image noise can be eliminated without the use of an off-axis plane wave for holographic recording, thereby lowering the data transfer rate for transmission and the spatial resolution requirement of spatial light modulators for display. Horizontal-parallax-only (HOP) holography is an excellent way to reduce holographic information and many holographic systems have been proposed and studied [18-19]. However, there has not been a single electronic holographic recording technique being investigated regarding the HPO-approach until the one proposed by Poon et al. in 2005 [20]. Electronic holography is an electronic processing technique that is used in the context of optical holography to bypass the use of films for recording. In this Chapter, we investigate HPO-electronic holography using optical scanning holography (OSH).

In Section 2, we first briefly review standard optical holography and then discuss optical scanning holography. In Section 3, we discuss how horizontal-parallax-only holographic information can be acquired by using OSH. In Section 4, we provide computer simulations of the idea and in the last Section, we make some concluding remarks.

2. OPTICAL SCANNING HOLOGRAPHY

In this Section, we briefly review traditional optical holography and then discuss optical scanning holography. Figure 1 shows the conventional optical holographic recording of a point source object, located z_0 away from the recording film. The point source object is the pinhole, which generates a

spherical wave as an object wave on the recording film. Together with the plane reference wave, the intensity recorded on the film is then given by

$$t(x, y, z_0) = |A + \frac{jB}{\lambda z_0} \exp[-\frac{j\pi}{\lambda z_0}(x^2 + y^2)]|^2, \quad (10-1)$$

where A and B are some constants, representing the amplitude of the plane reference wave and the spherical object wave, respectively. λ is the wavelength of the laser.

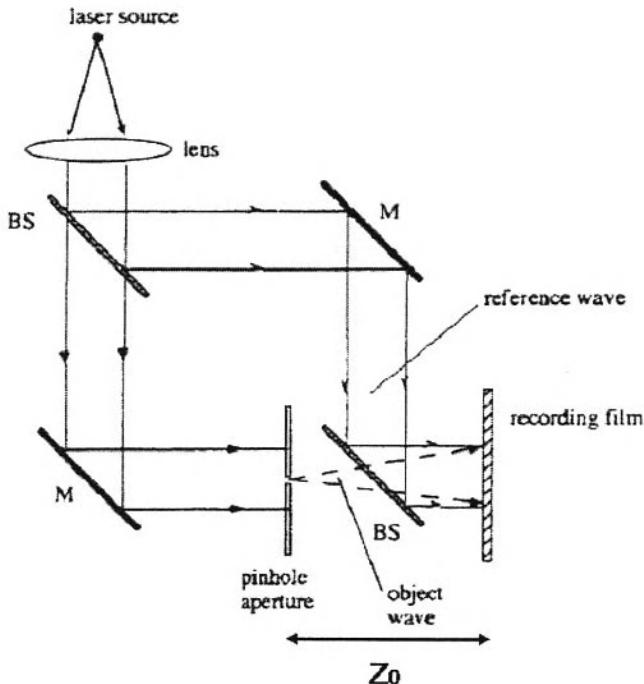


Figure 10-1 Holographic recording of a point source object (pinhole aperture)

Equation (10-1) can be expanded and written as

$$t(x, y, z_0) = a + b \sin(\frac{\pi}{\lambda z_0}(x^2 + y^2)), \quad (10-2)$$

where $a = A^2 + b(B/\lambda z_0)^2$, $b = 2AB/\lambda z_0$. Equation (10-2) is the well-known *Fresnel zone plate (FZP)*, which is the hologram of a point object. In short, a point object will give rise to a FZP in the hologram. Hence, for an arbitrary 3-D object, we consider it as a collection of point objects and hence the recorded hologram can be considered as a collection of Fresnel zone plates.

Optical scanning holography (OSH) is a novel 3-D imaging technique, invented by Poon, in that 3-D optical information of an object can be extracted by a single 2-D optical scan of the object [21]. The optical beam used to 2-D scan the object is called a *time-dependent Fresnel zone plate (TD-FZP)*, which is the superposition of a plane wave and a spherical wave of different temporal frequencies. The situation is shown in Fig. 10-2.

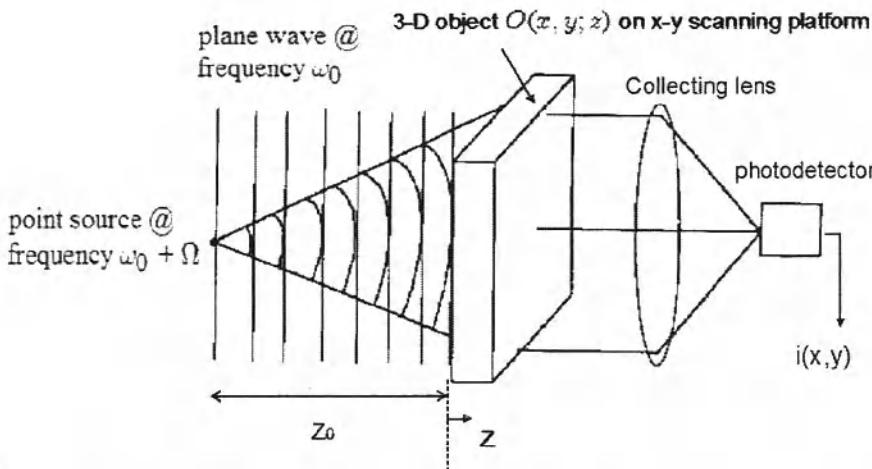


Figure 10-2: Principle of optical scanning holography: 3-D information of an object is extracted by 2-D scanning of a time-dependent Fresnel zone plate (TD-FZP) consisting a plane wave and a spherical wave of different temporal frequencies

The scanning beam intensity, as shown in Fig. (10-2), is given by

$$|A \exp(j\omega_0 t) + \frac{jB}{\lambda z_0} \exp[-\frac{j\pi}{\lambda z_0}(x^2 + y^2)] \exp[j(\omega_0 + \Omega)t]|^2, \quad (10-3)$$

where λ is the wavelength of the laser, ω_0 and $\omega_0 + \Omega$ are the temporal frequencies of the plane wave and the spherical wave, respectively. z_0 is the distance from the point source (which generates the spherical wave as one of

the scanning sources) to the object. Scanning of the plane wave and the spherical wave on the object can be done by using scanning mirrors or as shown in Fig. 10-2 where the object is placed on an $x - y$ scanning platform. The composite scanning beam, as given by Eq. (10-3), can be written as

$$I(x, y, z_0, t) = a + b \sin\left[\frac{\pi}{\lambda z_0}(x^2 + y^2) - \Omega t\right], \quad (10-4)$$

where again z_0 is the distance measured away from the point source of the scanning beam. a and b have been defined in Eq. (10-2). This composite scanning beam is called the time-dependent Fresnel zone plate (TD-FZP) [22]. Ω is called the *heterodyne frequency* and it is the result of a mixing between the plane wave and the spherical wave.

This TD-FZP is used to 2-D scan the object $O(x, y; z)$, where $O(x, y; z)$ represents an intensity distribution of a slice of the 3-D object in which we model the 3D object as a stack of transverse slices, and z is the depth or thickness parameter of the 3D object. The collecting lens then collects all the transmitted light and focuses the light on the photodetector (Note that if the object is diffusely reflecting, a photomultiplier can be used to pick up the scattered light). For the sake of explaining the concept, let us assume that a single point object $O(x, y; z) = \delta(x, y)$ is located z_0 away from the point source of the scanning beam. Now, the 2-D scanning of the TD-FZP on the point object will cause the photodetector to deliver a heterodyne current $i(x, y)$, as shown in Fig. 10-2:

$$i(x, y) \propto a + b \sin\left[\frac{\pi}{\lambda z_0}(x^2 + y^2) - \Omega t\right]. \quad (10-5)$$

The above result is mathematically explainable simply by knowing that the 2-D correlation of a function with a delta function gives the original function. Relating to our physical situation, the function in question is the TD-ZFP given by Eq. (10-4), and the delta function represents the point object. Two-dimensional scanning corresponds to 2-D correlation.

To explain further and refer to Fig. 10-2, we consider on the plane z_0 away from the point source of the scanning beam, we have a delta function correlating with the beam intensity $I(x, y, z_0; t)$ to give $i(x, y)$, which is given by

$$I(x, y, z_0, t) \otimes \delta(x, y) = \{a + b \sin\left[\frac{\pi}{\lambda z_0}(x^2 + y^2) - \Omega t\right]\} \otimes \delta(x, y)$$

$$= a + b \sin\left[\frac{\pi}{\lambda z_0}(x^2 + y^2) - \Omega t\right], \quad (10-6)$$

where \otimes denotes 2-D correlation, $x(t) = vt$ and $y(t) = vt$ in the above equation represent the instantaneous 2-D position of the scanning beam with respect to the object (which is $\delta(x, y)$ in our case), and are determined by the motion of the scanning mechanism with v the scanning speed of the scanning beam.

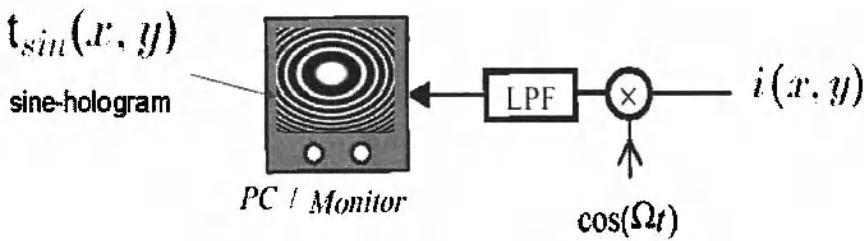


Figure 10-3 Demodulation scheme: $i(x, y)$ is first electronically multiplied and lowpass filtered to give a signal to write onto the PC/monitor

Now, the scanned current $i(x, y)$ from the photodetector can be demodulated according to Fig. 10-3, where it is first electronically multiplied by $\cos(\Omega t)$ and lowpass filtered before feeding into a PC for electronic storage or monitor for real time display of the scanned object. Mathematically, we see that using Eq. (10-6), we have

$$\begin{aligned} i(x, y) \times \cos(\Omega t) &= \left\{ a + b \sin\left[\frac{\pi}{\lambda z_0}(x^2 + y^2) - \Omega t\right] \right\} \times \cos(\Omega t) \\ &= a \cos(\Omega t) + b \left\{ \frac{1}{2} \sin\left[\frac{\pi}{\lambda z_0}(x^2 + y^2)\right] + \frac{1}{2} \sin\left[\frac{\pi}{\lambda z_0}(x^2 + y^2)\right] - 2\Omega t \right\} \end{aligned}$$

where we have used the trigonometric identity $\sin \alpha \cos \beta = (1/2)[\sin(\alpha + \beta) + \sin(\alpha - \beta)]$ to obtain the last part of the above equation. By performing lowpass filtering electronically to the above result as shown in Fig. 10-3, we obtain the second term of the above equation to give the holographic information of the point object as

$$b \sin\left[\frac{\pi}{\lambda z_0}(x^2 + y^2)\right], \quad (10-7)$$

where again $b \propto 1/\lambda z_0$. If we want the hologram to be displayed on a real-time monitor, we can add some bias c to the above result to give

$$t_{\sin}(x, y) = c + b \sin\left[\frac{\pi}{\lambda z_0}(x^2 + y^2)\right]. \quad (10-8)$$

The above result is basically identical to Eq. (10-2), which is the hologram of a point source object located z_0 away from the holographic recording film. However, note that in Eq. (10-8), we have a constant bias c in stead of the annoying space-varying bias a normally encountered in traditional optical holography (the term a will be a function of x and y if the object is more complicated than a point object being considered). In the context of optical scanning holography, the point source object is located z_0 away from the scanning beam (see Fig. 10-2). Eq. (10-8) is called the *sine-Fresnel zone plate (FZP) hologram* or simply *sine-hologram* of the point object, $t_{\sin}(x, y)$.

It is interesting to point out that in classical holography, FZP is generated as a hologram of the pin hole object in a photographic film, while in optical scanning holography, we reverse the order in that holographic information of a pin hole is being captured by synthesizing the TD-FZP to scan over it. Hence, in general, when a 3-D object $O(x, y; z)$ is scanned by a time-dependent Fresnel zone plate $I(x, y, z_0; t)$ [see Eq. (10-4)], each point on the object will give rise to a demodulated electrical signal corresponding to Eq. (10-7), specified by its location and depth, giving the sine-hologram of the object, $H_{\sin}(x, y)$. Mathematically, we have [23]

$$H_{\sin}(x, y) = \int \frac{\pi}{\lambda(z + z_0)} \sin\left[\frac{\pi}{\lambda(z + z_0)}(x^2 + y^2)\right] \otimes O(x, y; z) dz \quad (10-9)$$

where \otimes denotes 2-D correlation in x and y , and the integration is over the thickness or depth of the object. This 2-D storage of the information is then a sine-hologram $H_{\sin}(x, y)$ of the 3-D object. It is important to point out that in Fig. 10-3, we can form another electronic channel simultaneously to use $\sin(\Omega t)$ to multiply the current $i(x, y)$ instead of $\cos(\Omega t)$, we will then have a *cosine-FZP hologram*, $H_{\cos}(x, y)$, i.e., in Eq. (10-9), the sine is replaced by the cosine.

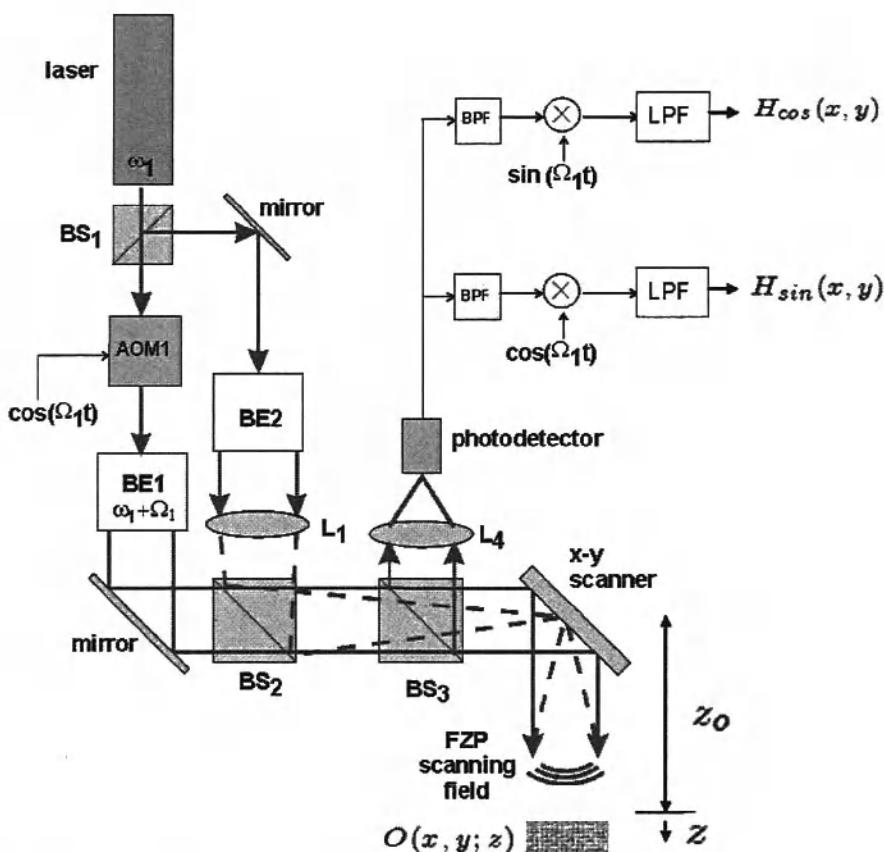


Figure 10-4 Optical scanning holographic recording - BS1, BS2, BS3: beamsplitter; AOM: acousto-optic modulator (to shift the frequency of the laser); BE1, BE2: beam expanders to collimate the light; L1, L4: lens; BPFs: bandpass filter (to capture the heterodyne frequency from the photodetector), LPFs: lowpass filter.

In essence, in order to perform optical scanning holography, one needs to implement Figs. 10-2 and 10-3 together. A practical version of the overall setup is shown in Fig. 10-4. Note that the FZP scanning field is the superposition of a spherical wave and a plane wave with different temporal frequencies. A collimated beam after beam expander BE_1 is a plane wave at frequency $\omega_1 + \Omega_1$, where ω_1 is the temporal frequency of the laser. Lens L_1 is used to focus the collimated beam to the $x-y$ scanner to form a spherical wave toward the 3-D object, $O(x, y; z)$. The FZP is used to scan over the 3-D object in two dimensions. The backscattered light from the 3-D object is received by BS_3 and lens L_4 to reach the photodetector. The sine-and cosine-holograms are to be displayed or stored in a PC for holographic

reconstruction (digitally or optically).

With the sine- and cosine-FZP holograms acquired simultaneously, we can construct a *complex hologram* as follows:

$$H^c(x, y) = H_{\cos}(x, y) \pm jH_{\sin}(x, y) \quad (10-10)$$

The choice of the plus sign or the minus sign in Eq. (10-10) allow us to choose either a real image reconstruction or a virtual image reconstruction. For optical holographic reconstruction, for example, we can illuminate the hologram $H(x, y)$, either sine-, cosine-, or complex hologram written on a *spatial light modulator* (SLM). Image reconstruction can be observed at different z away from the hologram according to Fresnel diffraction as follows:

$$\psi(x, y; z) = H(x, y) * h(x, y; z),$$

where $h(x, y; z) = \exp(-jk_0z) \frac{jk_0}{2\pi z} \exp[-j\frac{k_0}{2z}(x^2 + y^2)]$ is the free-space spatial impulse response in Fourier optics with k_0 being the wavenumber of the laser light [24], and $*$ denotes 2-D convolution. If we do not want to use a phase-type SLM for reconstruction of the complex hologram, but rather using the sine- and cosine-holograms, we propose the following optical system, as shown in Fig. 10-5, for reconstruction. The flat glass in one of the arms is to retard the phase of the light wave by 90 degrees, which corresponds to the “j” different, as seen in Eq. (10-10), between the sine- and cosine- hologram needed for reconstruction. We want to point out that holographic reconstruction can be obtained without the twin image noise if a complex hologram is used [25, 26], an important advantage over phase-shifting holography [27] where at least 3 holograms are acquired serially to use to eliminate the twin image.

Reconstruction can also simply be done by convolving digitally the hologram with the free space impulse response matched to the depth parameter z_0 , $h(x, y; z_0)$. In fact, this is what really has been known as *electronic holography* or *digital holography* - once the hologram has been captured electronically or digitally, reconstruction or manipulation of holographic information can be performed digitally.

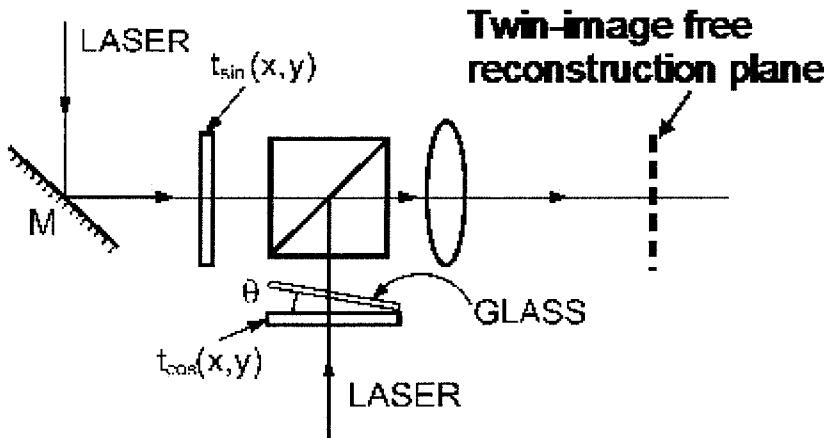


Figure 10-5 Proposed optical reconstruction without twin-image noise

3. ELECTRONIC HORIZONTAL-PARALLAX-ONLY HOLOGRAPHY

In 2005, Poon et al. suggested ways of generating electronic HPO-holographic information from real objects [20]. By recognizing that OSH is accomplished by scanning the 3-D object with a 2-D time-dependent Fresnel zone plate (TD-FZP), they visualize that horizontal-parallax-only (HPO) optical scanning holography can then be accomplished by 2-D scanning of the object by a 1-D TD-FZP to obtain holograms. Under this context, for HPO-sine holograms, Eq.(10-9) becomes

$$H_{hpo-s}(x, y) = \int \frac{\pi}{\lambda(z + z_0)} \sin\left[\frac{\pi}{\lambda(z + z_0)}(x^2 + y^2)\right] \exp\left[-\left(\frac{\pi}{\omega_x}\right)^2 - \left(\frac{\pi}{\omega_y}\right)^2\right] \otimes O(x, y; z) dz \quad (10-11)$$

where we have modeled the 1-D scanning zone plate by masking the 2-D zone plate with a narrow Gaussian function. With $\omega_x \gg \omega_y$, there will be no variation of the zone plate along the y -direction. The 1-D time-dependent FZP, for example, can be obtained by first replacing the x - y scanner with a mirror [see Fig. 10-4]. Away from the mirror we then place a slit and have the slit imaged onto the object through an x - y scanner. Indeed, by performing 2-D scanning of the 1-D TD-FZP onto the 3-D object, we can record two HPO-holograms: the 1-D HPO-sine hologram [see Eq. (10-11)] and the HPO-cosine hologram if two channels of electronic multiplexing are

used. In what follows, we show some computer simulations in standard OSH as well as in HPO-OSH.

For 2-D objects and according to Eq. (10-9), we can consider the sine-FZP hologram as the correlation between the sine-FZP and the object $O(x, y; z_0)$, located z_0 away from the focused spot on the mirror as shown in Fig. 10-4:

$$H_{\sin}(x, y) = \sin\left[\frac{\pi}{\lambda z_0}(x^2 + y^2)\right] \otimes O(x, y; z_0) \quad (10-12a)$$

and similarly for the cosine-FZP hologram, we have

$$H_{\cos}(x, y) = \cos\left[\frac{\pi}{\lambda z_0}(x^2 + y^2)\right] \otimes O(x, y; z_0). \quad (10-12b)$$

Figure 10-6a) shows the 2-D TD-FZP scanning beam [see Eq. (10-4)] at a fixed time, say at $t = t_0 = 0$. Note that the strange fringes along the four edges of the image are due to aliasing. Figure 10-6b) and c) show the sine- and cosine-FZP holograms of a 4-dot object, respectively. Figure 10-6d) shows the reconstruction when using a complex hologram described by Eq. (10-10). Note that there is no twin-image noise in the reconstruction.

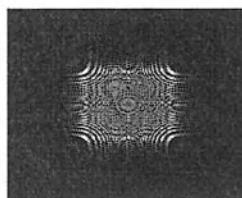


Figure 10-6a) 2-D TD-FZP scanning beam at a fixed time, see Eq. (10-4)

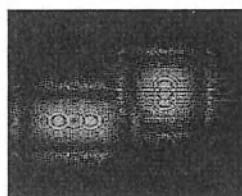


Figure 10-6b) Sine-FZP hologram of a 4-dot object

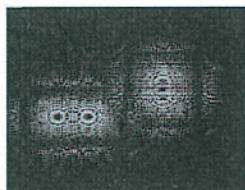


Figure 10-6c) Cosine-FZP hologram of a 4-dot object



Figure 10-6d) Reconstruction of the 4-dot object using a complex hologram

In contrast, for HPO-holograms, the 2-D scanning beam will be masked by a Gaussian function, as indicated in Eq. (10-11). Figure 10-7a) shows the 1-D scanning beam. The equation below shows the HPO-sine hologram for the 2-D object $O(x, y; z_0)$.

$$H_{hpo-s}(x, y) = \sin\left[\frac{\pi}{\lambda z_0}(x^2 + y^2)\right] \exp\left[-\left(\frac{\pi}{\omega_x}\right)^2 - \left(\frac{\pi}{\omega_y}\right)^2\right] \otimes O(x, y; z_0) \quad (10-13)$$

In Fig. 10-7b) and c), we show the HPO-sine and HPO-cosine hologram of the 4-dot object [see Fig. 10-6d)], respectively.



Figure 10-7a) 1-D TD-FZP scanning beam at a fixed time

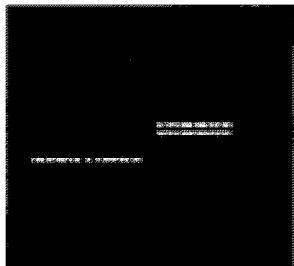


Figure 10-7b) HPO-sine hologram of the 4-dot object

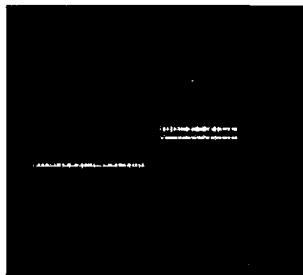


Figure 10-7c) HPO-cosine hologram of the 4-dot object

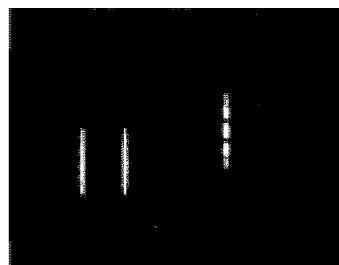


Figure 10-7d) Simulation result of optical reconstruction of complex-HPO hologram

We can construct a complex HPO-hologram according to Eq.(10-10) using the two holograms:

$$H_{hpo}^c(x, y) = \exp\left[-\frac{j\pi}{\lambda z_0}(x^2 + y^2)\right] \exp\left[-\left(\frac{\pi}{\omega_x}\right)^2 - \left(\frac{\pi}{\omega_y}\right)^2\right] \otimes O(x, y; z_0), \quad (10-14)$$

where we have used complex subtraction to obtain the above equation. In Fig. 10-7d), we show its optical reconstruction by computer simulations. Again, this is simply done by convolving digitally the complex HPO-hologram with the free space impulse response $h(x, y; z_0)$. The result is interesting in that we see along the x-direction the “dots” are properly holographically focused. However, along the y-direction, the light spreads, which is due to the slit-type holograms for each point of the object. To compensate this vertical spreading, Poon et al. has proposed the optical system shown in Fig. 10-8 to reconstruct the HPO-hologram [20]. A cylindrical lens of focal length of $z_0/4$ would compensate the y-spreading. Indeed, reconstruction simulation of Fig. 10-8 shows basically the original object given by Fig. 10-6d) [28].

cylindrical lens with focal length = $z_0/4$

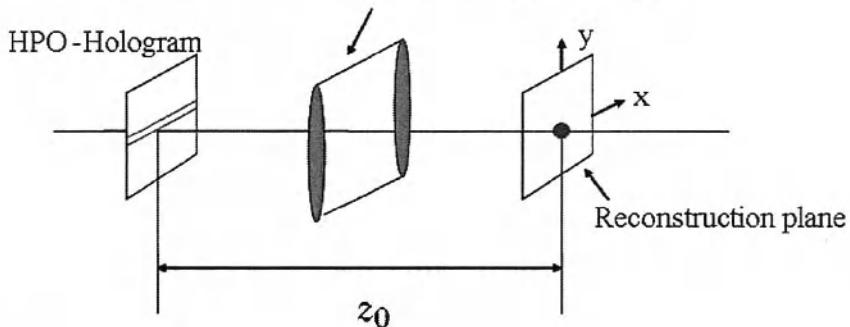


Figure 10-8 Proposed optical reconstruction system for HPO-holograms[adopted from [20]]

As another example involving HPO-holography, we show the 1-D scanning beam and the original 2-D object, respectively in Fig. 10-9a) and b).



Figure 10-9a) 1-D TD-FZP scanning beam at a fixed time



Figure 10-9b) Original 2-D object

In Fig. 10-9c) and d), we show the HPO-sine and cosine holograms generated using the 1-D scanning beam shown in Fig. 10-9a). Fig. 10-9e) shows the reconstruction of the HPO-complex hologram (see Eq. 10-14)) by convolving it with $h(x, y; z_0)$. Again, we see the spreading along the y-direction. One can use the proposed optical system in Fig. 10-8 to reconstruct without spreading along the y-direction. However, once the holograms are stored in a digital computer, we can reconstruct them using digital reconstruction. For digital reconstruction of HPO-holograms, we propose the use of a *chirp-Gaussian function* given below:

$$h^c_{hpo}(x, y) = \exp\left[-\frac{j\pi}{\lambda z_0}(x^2 + y^2)\right] \exp\left[-\left(\frac{\pi}{\omega_x}\right)^2 - \left(\frac{\pi}{\omega_y}\right)^2\right] \quad 10-15$$

to extract the object back from the hologram, as we one can show that

$$H^c_{hpo}(x, y) * h^c_{hpo}(x, y) \approx O(x, y, z_0). \quad (10-16)$$

Figure 10-9f) shows the reconstruction using Eq. (10-15).



Figure 10-9c) HPO-sine hologram

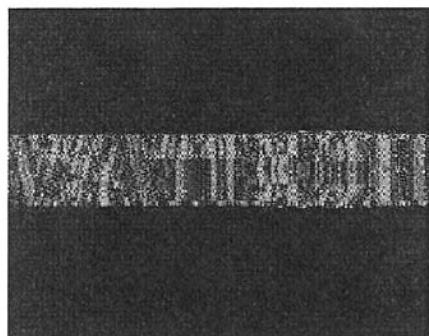


Figure 10-9d) JIPO-cosine hologram

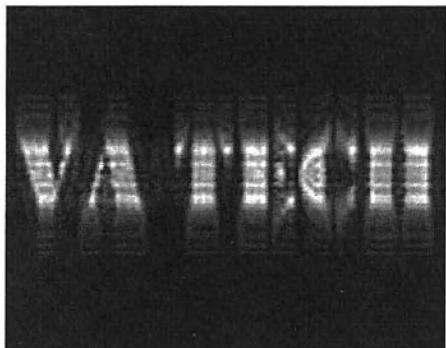


Figure 10-9e) Reconstruction of HPO-complex hologram



Figure 10-9f) Digital reconstruction using the proposed chirp-Gaussian function given by Eq. (10-15)

4. CONCLUDING REMARKS

We have briefly reviewed standard optical holography and optical scanning holography (OSH). In standard holography, holographic process can be considered as a collection of Fresnel zone plates, while in standard OSH we use a 2-D time-dependent Fresnel zone plate (TD-FZP) to scan over the object in two dimensions to obtain the holographic information of a 3-D object. In this Chapter, We have also proposed an optical system [see Fig. 10-5] to reconstruct a complex hologram to eliminate twin-image noise without the use of at least three holograms in phase-shifting holography.

We have also discussed how HPO-electronic holography can be accomplished in OSH by simply using a 1-D TD-FZP to scan over the object, as first proposed by Poon et al. While optical experiments have not been constructed, we performed some simulations to illustrate the idea. Along the line of digital reconstruction, we have proposed a chirp-Gaussian function to reconstruct HPO-holograms and simulations results have confirmed the usefulness of the proposed function. We submit that optical scanning holography is a powerful electronic holography technique for 3-D imaging and hope that this Chapter will stimulate further research in novel electronic holography and its various applications.

REFERENCES

1. L. H. Enloe, J.A. Murphy and C. B. Rubinstein, "Hologram Transmission via Television," Bell Syst. Techn. 25, 335-339 (1966).
2. S.B. Gurevich et al., "Holographic Image Transmission by Television," Sov. Phys. 13, 378-383 (1968).
3. C. B. Burckhardt and E.T. Doherty, "Formation of Carrier Frequency Holograms with an On-axis Reference Beam," Applied Optics 7, 1191-1192 (1968).
4. L. H. Lin," A Method of Hologram Information Reduction by Spatial Frequency Sampling," Applied Optics 7, 545-548 (1968).
5. K. A. Haines and D. B. Brumm," Holographic Data Reduction," Applied Optics, 6, 1185-1189 (1968).
6. A. Macovski, "Considerations of Television Holography," Opt. Acta 18, 31-39 (1971).
7. N. Hashimoto, S. Morokawa and K. Kitamura, "Real-time Holography Using the High-Resolution LCTV-SLM," Proc. SPIE, Vol. 1461, pp. 291-302, 1991)
8. S. A. Benton, "Experiments in Holographic Video," Proc. SPIE IS-08, 247-267 (1991).
9. M. Yamaguchi, H. Sugiura, T. Honda, and N. Ohyama, J. Opt. Soc. Am. A. 9, 1200 (1992)
10. T.-C. Poon, B. W. Schilling, M. H.Wu, K.Shinoda and Y. Suzuki, "Real-Time Two-Dimensional Holographic Imaging Using an Electron Beam Addressed Spatial Light Modulator," Optics Letters, Vol. 18, 63-65 (1993).
11. K. Sato, K. Higuchi, and H. Katsuma, "Holographic Television by Liquid Crystal Device," Proc. SPIE 1667, 91-31 (1993).
12. T.-C. Poon, K. Doh, B. Schilling, K. Shinoda, Y. Suzuki, and M. Wu, " Holographic Three-Dimensional Display using an Electron-Beam-addressed Spatial-Light-Modulator," Optical Review, 567-571 (1997).
13. P. St. Hilaire, " Holographic Video: the Ultimate Visual Interface?" Optics & Photonics News, p. 35, August (1997).
14. K. Shinoda, and Y. Suzuki, M. Wu and T.-C. Poon," Optical Heterodyne Scanning Type Holography Device," U.S. Patent No. 5064257, Nov. 1991.
15. H. B. Brown, S.C. Noble and B.V. Markevitch, "Three-dimensional Television System using Holographic Technique," U.S. Patent No. 4,376,950, March, 1983.
16. R. L. Kirk," Electronically Generated Holography," International Patent No. WO 84/00070, Jan., 1984.
17. T.-C. Poon, "Three-Dimensional Television Using Optical Scanning Holography," Journal of Information Display, 3, 12-16 (2002). *Invited.*

18. S.A. Benton," Hologram Reconstruction with Extended Incoherent Sources," J.Opt. Soc. Am., 59, 1545A (1969).
19. A. Beauregard and R. A. Lessard," Rainbow holography of 3-D Stationary Objects with no Slit," Applied Optics, 23, 3095-3729 (1999).
20. T.-C. Poon, A. Akin, G. Indebetouw, T. Kim," Horizontal-Parallax-Only Electronic Holography," Optics Express, 6 pages (2005).
21. T.-C. Poon," Scanning holography and Two-Dimensional Image Processing By Acousto-Optic Two-Pupil Synthesis," J. Opt. Soc. Am., 2, pp. 521-527 (1985).
22. T.-C. Poon, M. Wu, K. Shinoda, and Y. Suzuki, "Optical Scanning Holography," Proceedings of the IEEE , Vol. 84, pp. 753-764, (1996), *Invited review paper*.
23. T.-C. Poon and T. Kim, "Optical Image Recognition of Three-Dimensional Objects," Applied Optics, 38, pp. 370-3811999.
24. T.-C. Poon and P.P. Banerjee, *Contemporary Optical Image Processing With MATLAB*, Elsevier Science, Oxford, U.K. (2001).
25. T.-C. Poon, T. Kim, G. Indebetouw, M. H. Wu, K. Shinoda, and Y. Suzuki, "Twin-Image Elimination Experiments for Three-Dimensional Images in Optical Scanning Holography," Optics Letters, 25, pp.215-217 (2000).
26. T.-C. Poon," Three-Dimensional Image Processing and Optical Scanning Holography," in *Advances in Imaging and Electron Physics*, Vol. 126, 329-350, Academic Press, 2003 (ISBN-0-12-014768-8).
27. I. Yamaguchi and T. Zhang, "Phase-Shifting Digital Holography," Optics Letters, 22, pp.1268-1270 (1997).
28. T. Akin, MS Thesis, in progress, Dept. of Electrical and Computer Engineering, Virginia Tech, USA.

Chapter 11

THREE-DIMENSIONAL PROJECTION DISPLAY SYSTEM

Eun-Soo Kim

Dept. of Electronic Engineering, Kwangwoon University, Seoul Korea

Abstract: In this chapter, a current state of the art of 3D projection displays will be addressed. This begins with a review on conventional and three-dimensional projection display systems. And, some discussions on stereoscopic projection displays such as the polarization-based, the LC shutter-based and the ZScreen-based will be followed. Then, time-sequential multiview projection display and projection integral-imaging display will be discussed. Finally real 3D projection displays such as the Fresnel lens-based and the hologram-based will be reviewed.

Key words: 3D Display, Projection

1. INTRODUCTION

1.1 Overview of projection displays

In general, an electronic display is defined as a device or system that converts electronic signal information representing video, graphic or text to a viewable image of this information. Displays can be largely classified into direct-view or projection. Direct-view displays produce their images on the surface being viewed. On the other hand, the images from projection displays are formed on auxiliary surfaces, which are physically separated from the image-generating component. The direct-view technologies include cathode-ray tubes(CRTs) in televisions and computer monitors as well as flat-panel displays, such as liquid crystal displays(LCDs), plasma display

panels(PDPs), and organic light emitting diode displays(OLEDs) in various information terminals. All these displays have their own capabilities of high-resolution and satisfactory luminance. But, it is difficult and expensive to make a large-scale direct-view display enough to accommodate several viewers simultaneously.

Human visual system is known to have an angular resolution of approximately 1 minute of arc. In case an viewer would watch the HDTV(high-definition television) at a distance of 2 meters from the television as shown in Figure 11-1, the TV size must be as large as $\sim 70''$ in diagonal to fully resolve the TV video images with a resolution of high definition, which is shown in 1920 X 1080, $\sim 0.6\text{mm}$, full-color pixels [1].

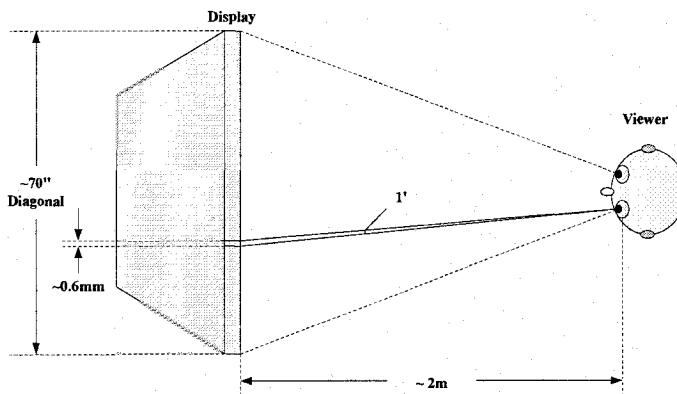


Figure 11-1. Display geometry for viewing direct-view HDTV

It is a surely expensive and challenging problem to implement a direct-view display of this size at present even though several prototypes of LCDs and PDPs of this size were being demonstrated very recently[2, 3].

Alternatives to the large-scale direct-view displays, Many types of projection displays have been devised[4]. Projection displays make use of an optical imaging system to magnify a small picture created either by conventional direct-view technologies, such as CRTs, or by modulating the light from an illumination system with a device called a panel. Projection displays produce larger images from electronic signals than is normally achieved by direct-view technologies. Figure 11-2 shows where projection displays are located in the display market with regard to resolution and screen size[1, 4].

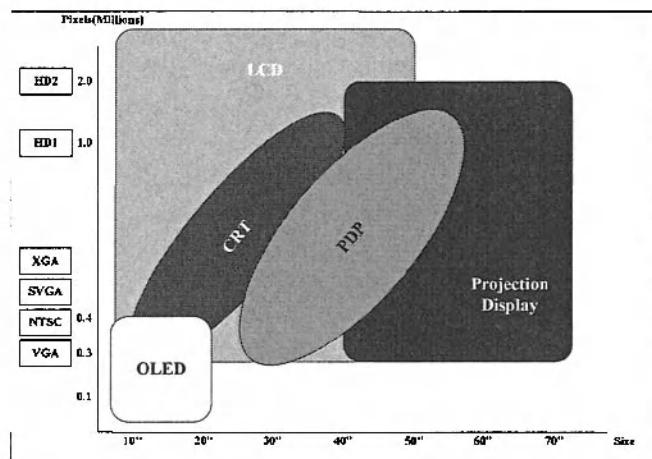


Figure 11-2. Applications of projection displays as a function of screen size and resolution

Projection displays are commonly categorized as front-projection and rear-projection types depending on the optical path used as shown in Figure 11-3. In case of front-projection types, the image source and the viewer are on the same side of the projection screen and the image is viewed by reflection from the screen surface, whereas in rear-projection types the image source and the viewer are on the opposite side of the projection screen and the image is projected to the rear of the screen and observed through the translucent screen material. So far, projection display systems can offer an economical solution to large and high-resolution information displays.

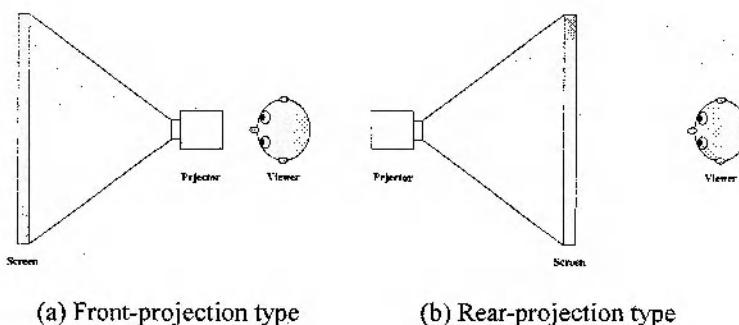


Figure 11-3. Two types of projection display system

There are two key technologies providing impetus for the development of advanced projection displays. The first technology is the personal computer, which can now provide portable information sources of high-quality electronic video signals, so that projector makers have developed transportable projectors for displaying these signals over the last 10 years. A second technology to make deep influence on the present development of projection displays is the digital HDTV. To implement direct-view displays for the purpose of HDTV is very difficult and expensive, because the present size range of direct-view displays cannot effectively display all the resolution contained in a HDTV signals as mentioned above. Accordingly various types of projection display systems will provide cost-effective displays for this need.

1.2 Projection display systems

The core part of all projectors is the light engine. The light engine is the subsystem within the projector, which converts incoming electrical signals into the intensity-modulated two-dimensional optical images. The key components required for the design of a light engine include image forming devices, lamps, optical components, and attributes of these components strongly influence the overall system performance. There are many kinds of projection display systems on market, the CRT-based, the LCD-based, the LCoS(liquid crystal on silicon)-based and the DMD(digital micromirror device)-based. These conventional projection display systems have been briefly reviewed here.

1.2.1 CRT-based projection display

So far the CRT-based is the most common projection display systems at the commercial market. In this type of projectors, separate image sources are typically used for generation of the three primary colors; red, green and blue, so that the light engine of these CRT projectors is composed either of the three-CRT/ three-lens or three-CRT/ one-lens configurations as shown in Figure 11-4[1].

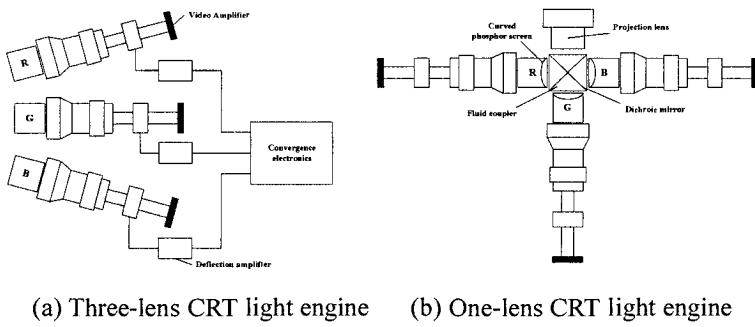


Figure 11-4. Two configurations of the CRT light engine

Three monochrome tubes, each optimized for luminance and beam width of a specific primary color, are imaged onto the screen and combined to form a full color image. In case of three-lens CRT projectors, convergence of CRT projection systems is a most challenging problem. Because red and blue channels are in an off-axis arrangement in the three-lens CRT projection system as shown in Figure 11-4(a), a trapezoidal image known as keystone distortion can be generated from these off-axis tubes[5]. Even though single-lens CRT projectors shown in Figure 11-4(b) are free from this keystone distortion occurred in the three-lens system, they also suffer from some other sources degrading convergence of CRT projectors such as the long back focal length and the relatively high f-number, etc. [6]. Majority of conventional projection display systems are based on the three-lens architecture and these are made for both front- and rear-projection installations.

1.2.2 Transmissive three-panel projection display

Most of front and rear transmissive projectors, which are commercially available now, incorporates high-temperature poly silicon(HTPS) panels in a standard three-panel, X-cube configuration[1]. Figure 11-5 shows a standard three-panel transmissive LC projection display system, which was first introduced commercially by Seiko-Epson in 1997[7]. In this projector, the incident white light is split into three primary colors by dichroic filters and each primary color passes through an LCD panel sandwiched between two sheet polarizers. Then, the output light is spatially modulated by the voltage applied to the pixels and the modulated three color beams are combined by a dichroic X-cube before the projection lens.

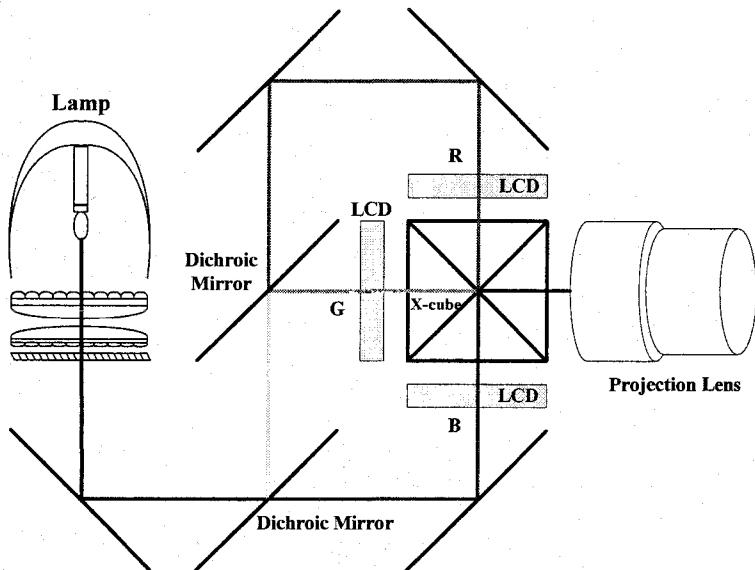


Figure 11-5. System architecture of a standard 3-panel transmissive projection display

1.2.3 Reflective three-panel projection display

Commercial success of the standard transmissive three-panel projection system has driven down the cost of individual components. Accordingly a birth of the first three-panel LCoS projection architecture might be a derivative of the transmissive architecture, with each modulating subsystem(polarizer/panel/analyzer) being replaced by a PBS (polarization beam splitter) and LCoS panel. This architecture has been called ‘3xPBS/X-cube’[8]. Figure 11-6 shows the reflective three-panel LCoS projection display based on three cube MacNeille PBSs surrounding an X-cube, which was first developed by IBM and Nikon[9]. As an alternative, an off-axis LCoS system was also developed[10, 11], in which the incident and reflective beams do not counter-propagate. Sheet polarizers can be used instead of PBSs to pre-polarize the incident beam and separately analyze the reflected light with off-axis designs. But, the panel convergence and projection lens design have been known to be difficult in this off-axis LCoS system.

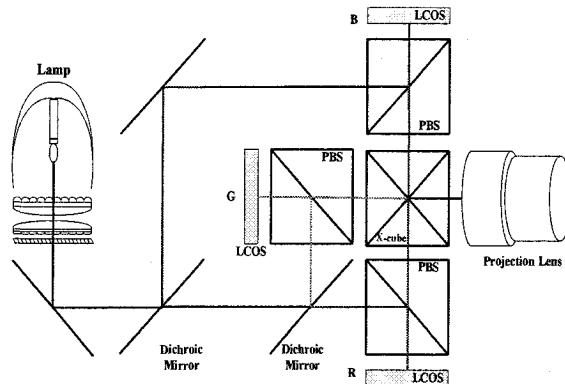


Figure 11-6. Architecture of the 3xPBS/X-cube reflective 3-panel LCoS projection display

More recently, three-panel LCoS architectures employing RSFs(retarder stack filters) with MacNeille PBSs, which combine the polarizing/analyzing of color, have been introduced[12]. This technology is capable of very compact color management systems. The ColorQuadTM[12] shown in Figure 11-7 is one of LCoS projection architectures based on this approach.

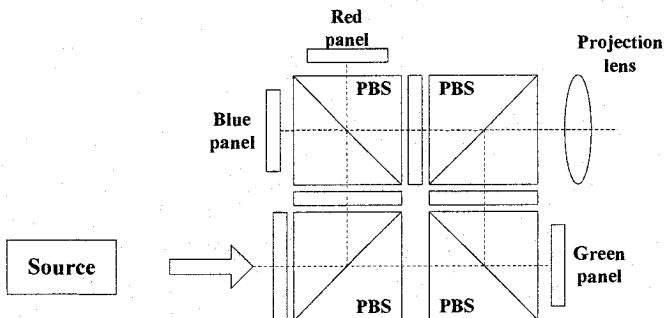


Figure 11-7. LCOS projection system based on ColorQuadTM

1.2.4 DMD-based projection display

There is an optical semiconductor known as the DMD at the heart of every DLP projection system. This DMD chip, which was invented by Dr. Larry Hornbeck of Texas Instruments[13], might be regarded as the most sophisticated light switch in the world. It consists of a rectangular array of up to 1.3 million hinge-mounted microscopic mirrors. Each of these micromirrors is measured to be less than 10 μm in size and corresponds to

one pixel in a projected image. When a DMD chip is coordinated with a digital video signal, a light source and a projection lens, its mirrors can reflect an all-digital image onto a screen. The DMD and the sophisticated electronics that surround it are what we call DLP(digital light processing) technology. Figure 11-8 shows a photomicrograph of a DMD mirror array[14].

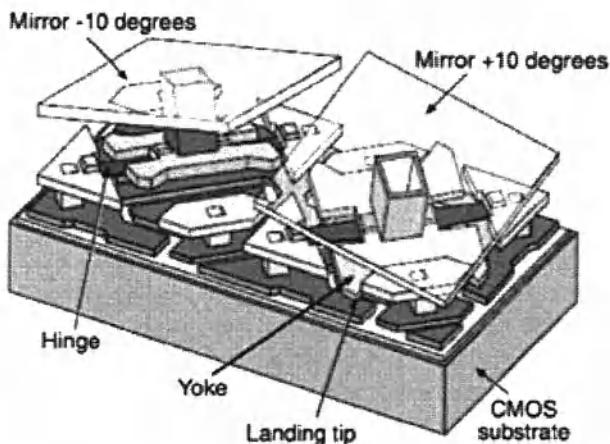


Figure 11-8. Digital micromirror device(DMD)

A DMD panel's micromirrors are mounted on tiny hinges that enable them to tilt either toward the light source in a DLP projection system or away from it, creating a light or dark pixel on the projection surface. The bit-streamed image code entering the semiconductor directs each mirror to switch on and off up to several thousand times per second. When a mirror is switched on more frequently than off, it reflects a light gray pixel; a mirror that's switched off more frequently reflects a darker gray pixel. In this way, the mirrors in a DLP projection system can reflect pixels in up to 1,024 shades of gray to convert the video or graphic signal entering the DMD into a highly detailed grayscale image.

1.3 Three-dimensional projection display

Human beings see the different viewpoint images of an object through two eyes, the right and left. Then, the human brain recognizes three-dimensional (3D) stereopsis of the object by synthesizing them with the binocular disparity of a stereo input image pair[15]. Most of the conventional 3D projection display systems have been implemented by

imitating this human visual system (HVS), so that these systems normally need a pair of camera for capturing the left and right images of an object, and a pair of projectors for projecting the captured left and right images to the screen, and some special optical devices for separately inputting the projected left and right images to the corresponding eyes, contrary to the two-dimensional projection displays, in which only one camera, one projector is needed for capturing and projecting images of an object.

There are more than 50 kinds of 3D display systems developed up to now[15]. They can be largely categorized as stereoscopic and autostereoscopic 3D display systems depending on image separation methods as shown in Figure 11-9. In the stereoscopic 3D display system, the viewer is required to wear special glasses such as polarized or shutter glasses for separated reception of the left and right images on the eyes. But autostereoscopic 3D display systems can present 3D image to the viewer without a need for any special glasses, in which optical elements such as lenticular sheets or parallax barriers are attached on the display panels and they make the left and right images displayed on the panel to be collected on the corresponding eyes without any interference between them. In short, in the stereoscopic display systems the viewer takes on the polarized or shutter glasses, whereas the display panel is forced to take on special optical elements such as a lenticular sheet or a parallax barrier in the autostereoscopic display systems.

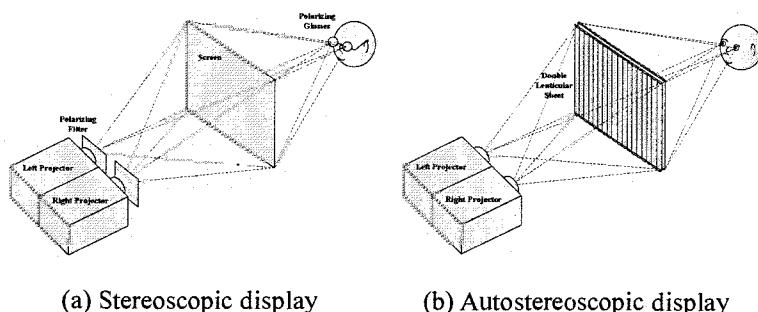


Figure 11-9. Types of conventional 3D projection display system

Meanwhile a holographic display, which is totally different from the stereoscopic and autostereoscopic approaches, has been regarded as one of the most attractive approaches for creating the most authentic illusion of observing volumetric objects. It is because the holographic technology can supply high-quality images and accurate depth cues viewed by human eyes without any special observation devices [16].

2. STEREOSCOPIC PROJECTION DISPLAYS

With stereoscopic projection displays two different perspective views are generated simultaneously and then these left and right views of a scene are projected onto the screen and then distributed to the left and right eyes of the viewer independently. Often special viewing devices are required to direct the appropriate view to the correct eye and block the incorrect view to the opposite eye. This technology is called *stereoscopic display*. Several methods for stereoscopic projection display have been reviewed here.

2.1 Polarized stereoscopic projection display

2.1.1 Overview

There are many kinds of projector systems commercially available such as CRT, LCD and DLP projectors mentioned above[4]. All these projector systems are now being used for polarized stereoscopic projection display. Figure 11-10 shows a general geometry of polarized stereoscopic projection display systems, which consists of two video projectors, two external polarizers, screen and a pair of polarized glasses. Basically, polarized stereoscopic projection involves the polarization of the two (left and right) projected images in orthogonal directions[4, 17]. The general standard for linear polarized projection is $+45^\circ/-45^\circ$ from the vertical axis. But, vertical (90°) and horizontal (0°) polarization for the two views can be also used. For the case of circular polarization, the two views are clockwise and counter-clockwise polarized.

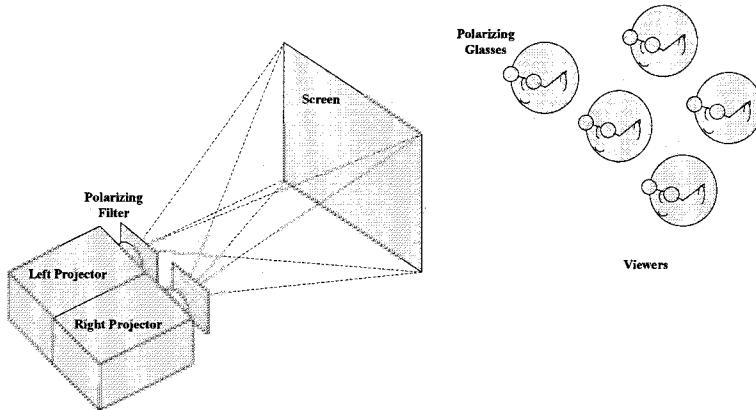


Figure 11-10. General geometry of polarized stereoscopic projection display

Commercial video projectors have some different operational characteristics. Especially each projector has its unique polarization state of the out beam. Accordingly in order for these video projectors to be used with polarized stereoscopic projection, lots of attention must be given to the possible interaction between the polarization orientation of the projected output and the desired polarization direction for stereoscopic projection.

2.1.2 Configurations of projectors for polarized stereoscopic projection

In general, video projectors mostly used for polarized stereoscopic projection can be classified into three types depending on the polarization state of their outputs[17] The majority of commercial LCD projectors has the linear polarized output with two colors in one direction and the other color in an orthogonal direction, which are classified as the *Type-1* LCD projector here. Some other LCD projectors also have the linear polarized outputs with all colors in the same directions, which are classified as the *Type-2* LCD projector here. On the contrary to the LCD projectors, the CRT, DMD and DLP projectors have the unpolarized outputs, which are classified as the *Type-3* projector here.

Figure 11-11 shows the output polarization states of three types of projectors. *Type-1* LCD projectors have the linear polarized outputs with two colors of red and blue in the vertical direction and the other color of green in the horizontal direction. *Type-2* LCD projectors have the linear polarized outputs with all colors in the same directions and *Type-3* projectors have the unpolarized outputs.

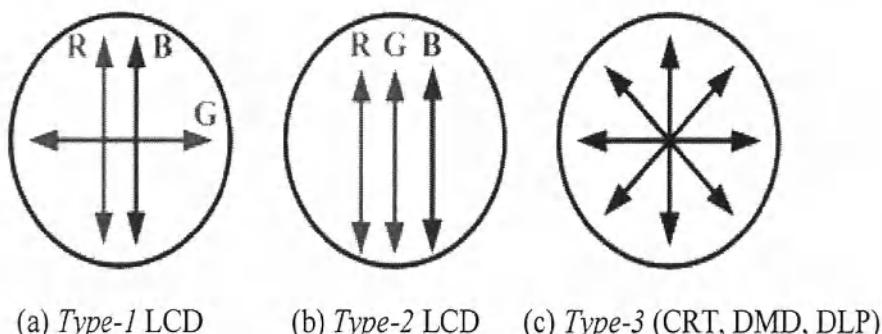


Figure 11-11. Illustration of output polarization of three types of projector

For these three types of commercial projectors to be used for polarized stereoscopic projection, various configurations of optical polarizers must be employed to obtain correct polarization output and color balance with minimal light loss and color distortion.

Figure 11-12 shows an overall optical layout of the conventional polarized stereoscopic projection system, where RR, RB, RG and LR, LB, LG represent the red, blue and green color components for the right and left projectors, respectively[18-20]. As shown in Figure 11-12, this system consists of two projectors, two external linear polarizers, Fresnel screen and a pair of glasses with linear polarizers.

Basically, polarized stereoscopic projection involves the polarization of the two (left and right) projected views in the orthogonal direction. Accordingly, two external linear polarizers must be mounted at either $+45^\circ$ or -45° in front of the right and left projectors for achieving correct orthogonal polarization and color balance as shown in Figure 11-12. These left and right beams polarized in the orthogonal direction by two external linear polarizers are simultaneously projected onto the Fresnel screen. And then, observer can view the stereoscopic 3D image by putting on a pair of glasses polarized in the same directions with those of the external linear polarizers ($+45^\circ$ for the right eye, -45° for the left eye).

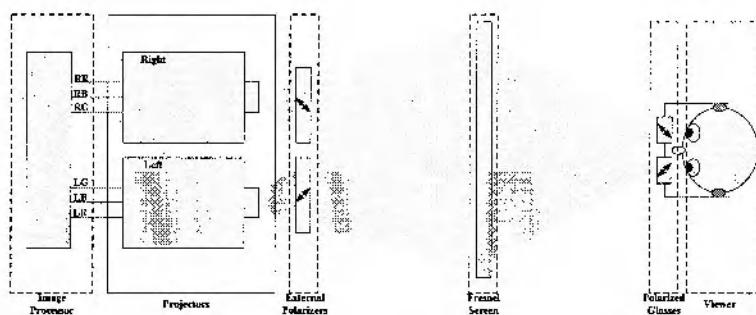


Figure 11-12. Overall optical layout of the conventional polarized stereoscopic projection system

For the case of *Type-1* LCD projectors having the polarization states of Figure 11-13(a), which are mostly consisted of three LCD panels, two external linear polarizers must be mounted at either $+45^\circ$ or -45° in front of the right and left projector for achieving the correct orthogonal polarization and color balance as shown in Figure 11-13(b). Here, two circular polarizers (one is clockwise, the other is counterclockwise) with their axis at $\pm 45^\circ$ could be also used instead of two external linear polarizers.

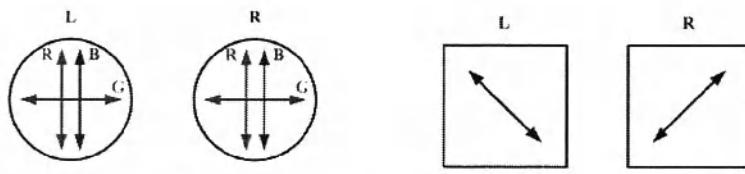


Figure 11-13. Method of linearly polarized stereoscopic projection with *Type-1* LCD projectors

For the case of *Type-2* LCD projectors having the polarization states of Figure 11-14(a), which are consisted of single or three LCD panels, the vertically polarized output of the projectors are rotated into the $+45^\circ$ and -45° orientations using half-wave retarder and then, external linear polarizers are placed in front of the retarders at the desired directions of $+45^\circ$ and -45° as shown in Figure 11-14(b). Here, also two circular polarizers (one is clockwise, the other is counterclockwise) with their axis at $\pm 45^\circ$ can be also used instead of two external linear polarizers.

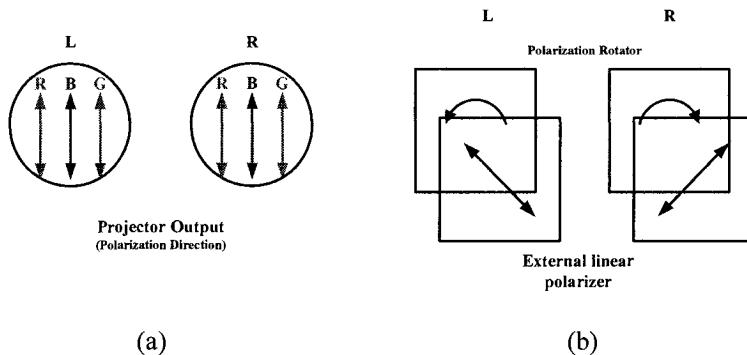


Figure 11-14. Method of linearly polarized stereoscopic projection with *Type-2* LCD projectors (a) Polarized outputs of *Type-2* LCD (b) Half-wave retarders and external linear polarizers

On the contrary to the *Type-1*, *Type-2* LCD projectors the output of the *Type-3* projector of Figure 11-15(a) is unpolarized, so that two external linear polarizers are simply needed for polarized stereoscopic projection. As shown in Figure 11-15(b), the external linear polarizers at the output of the projectors are oriented at $\pm 45^\circ$ from the vertical axis.

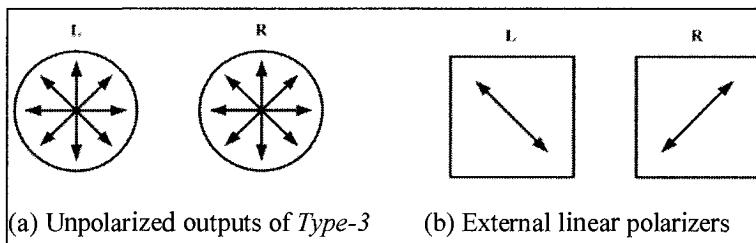


Figure 11-15. Method of linearly polarized stereoscopic projection with *Type-3* projectors

Configurations of polarized stereoscopic projection display by use of three types of commercial projectors in combination with external polarizers explained above are summarized in Figure 11-16. That is, the *Type-1* LCD projector-based, *Type-2* LCD projector-based, *Type-3* projector-based polarized stereoscopic projection display systems are shown in Figure 11-16(a), (b) and (c), respectively.

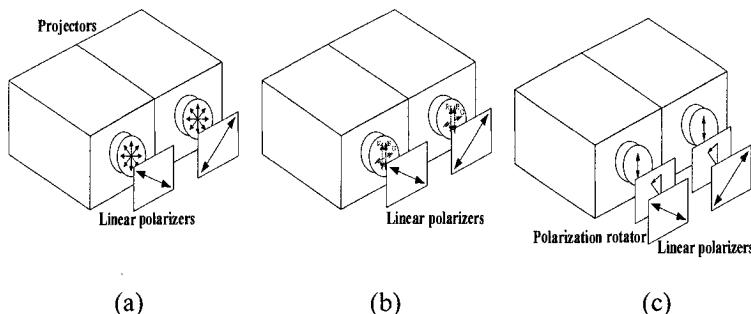


Figure 11-16. Three kinds of configurations of commercial projectors for polarized stereoscopic projection (a) Type-1 LCD projector-based (b) Type-2 LCD projector-based (c) Type-3 projector-based

2.2 LCD polarized stereoscopic projection display with improved light efficiency

2.2.1 Overview

Generally, it is well known that in the conventional projector configurations for polarized stereoscopic projection mentioned above, more than 50% of light energy might be inevitably lost in the process of polarizing the two projected views in the orthogonal directions. In particular light loss of these conventional polarized stereoscopic projection systems highly depends on the configuration of two projectors for polarized stereoscopic projection. A. Woods discussed various configurations of the commercial projectors for polarized stereoscopic projection and evaluated their performances in terms of light loss[17]. His experiments reveal that about 68%, 43%, 75% of light energy is measured to be lost for the cases of *Type-1* LCD, *Type-2* LCD and *Type-3* projectors, respectively in the polarization process.

In case of *Type-3* projectors, two polarizers must be placed in front of each projector for the two views to be polarized in the orthogonal directions, because the outputs of *Type-3* projector are unpolarized, so that there can be no chance to get rid of the additional light loss caused by these polarizers in this system. On the other hand, in cases of *Type-1* and *Type-2* LCD projectors, their outputs are already polarized, so that a new type of LCD polarized stereoscopic projection display without using these additional polarizers can be devised by effectively taking advantage of the inherent polarization property of the conventional LCD projectors. Recently, S. C

Kim and E. S Kim[18] developed a new configuration of LCD polarized stereoscopic projection display with improved light efficiency, in which two external polarizers have been excluded and as a result, light loss could be dramatically reduced.

2.2.2 Type-1 LCD-based polarized stereoscopic projection display

Figure 11-17 shows a new configuration of the *Type-1* LCD polarized stereoscopic projection display proposed by S. C Kim and E. S Kim[18, 21]. By comparing with Figure 11-12, this system also consists of two LCD projectors, Fresnel screen, and a pair of glasses with linear polarizers, but there are some differences. That is, two external polarizers employed in the conventional system are taken away, the right LCD projector is physically rotated by 90 degrees with respect to the left one and the green components of two projectors are mutually exchanged in this system.

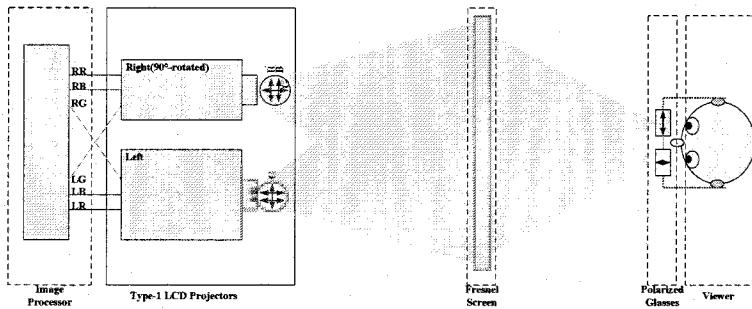
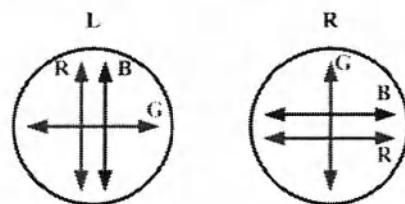


Figure 11-17. Overall optical layout of a new *Type-1* LCD polarized stereoscopic projection system

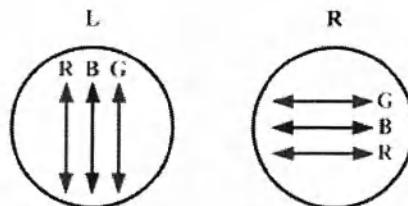
Figure 11-18(a) shows the left projector and it has output polarization of the normal *Type-1* LCD projector, in which red and blue components of the projector are vertically polarized, whereas green component is horizontally polarized[25]. If this right projector is physically rotated by 90° with respect to the left one then, the red and blue components, which are originally polarized at the vertical directions, are converted into horizontal polarization, whereas the green component, which is originally polarized at the horizontal direction, is transformed into the vertical polarization as shown in Figure 11-18(b).



(a) Left projector (b) 90°-rotated right projector

Figure 11-18. Output polarization of the left and 90°-rotated right projectors for the case of Type-I LCD

Thus, red and blue components of the left projector and green component of the 90°-rotated right projector are vertically polarized. At the same time, red and blue components of the 90°-rotated right projector and green component of the left projector are horizontally polarized. Accordingly, by simultaneously exchanging the green color components between the left and 90°-rotated right projectors through the signal processing technique, two full linearly polarized color sets, which should meet the requirements of color balance and orthogonal polarization, can be obtained without a need of external polarizers as shown in Figure 11-19. That is, the left color set is vertically polarized, whereas the right one is horizontally polarized.



(a) Left projector (b) 90°-rotated right projector

Figure 11-19. Transformed output polarization of the left and 90°-rotated right projectors for the case of Type-I LCD

Figure 11-20 shows a flowchart of stereo image processing for the new Type-I LCD projector-based polarized stereoscopic projection system. Firstly, the left and right video images captured by stereo camera are separated into three-color components of red, blue and green, respectively. Then, a new left image for the left projector, which is called a transformed left image here, is generated by mixing the red and blue components of the left image with the green component of the right image. At the same time, a

new right image for the 90°-rotated right projector is generated by mixing the red and blue components of the right image with the green component of the left image. But, because the right projector has been initially rotated by 90° with respect to the left one in the proposed scheme, the right image projected by the right projector is also rotated by 90° with respect to the left one. Therefore, the new right image must be adjusted in its orientation and aspect ratio to match with those of the left one through some image processing techniques before it is loaded onto the 90°-rotated right projector, which is now called a transformed right image here. These newly transformed left and right images are sent to the corresponding left and 90°-rotated right projectors and the size of the projected stereo image pair is matched on the screen by using a 3D reform function of the projector. Then, by using a pair of glasses with linear polarizers oriented at the horizontal(right eye) and vertical(left eye) directions, the newly transformed stereoscopic images can be finally viewed.

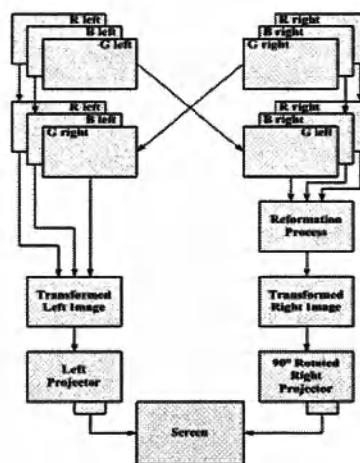


Figure 11-20. Flowchart of stereo image processing for the proposed *Type-1* LCD projector-based stereoscopic display

It is certain that output beam intensities of the new *Type-1* LCD-based polarized stereoscopic projection system are increased two times than those of the conventional system, which means 50% of light energy might be inevitably lost in the conventional system because the external polarizers are employed. Moreover the external polarizers mentioned above are assumed to be ideal for theoretical analysis, but in case the commercial polarizers are used more than 50% of light loss might be expected. According to the experimental results of A. Woods[17], 68% of light loss is measured in the

practical *Type-1 LCD*-based polarized stereoscopic projection system. From some experimental results with NEC MT 1060R projectors belonging to the *Type-1 LCD* projector[21, 22], S. C. Kim and E. S Kim[18, 21] show that light efficiency of the new system could be maximized and the stereoscopic video image projected from the new system can be made to be 213%, 75% and 300% brighter than those projected from the conventional *Type-1 LCD* projector-based, *Type-2 LCD* projector-based and *Type-3* projector-based systems, respectively. Figure 11-21 shows experimental results for stereoscopic video images of ‘Korean traditional wedding’ projected on the screen from the conventional and new *Type-1 LCD* polarized stereoscopic projection systems, respectively. These figures visually conform that output light beam projected from the new system is much brighter than that of the conventional system.



(a) Conventional method



(b) New method

Figure 11-21. Experimental results of the *Type-1 LCD*-based polarized stereoscopic projection system

2.2.3 Type-2 LCD-based polarized stereoscopic projection display

A new configuration of *Type-2 LCD* polarized stereoscopic projection system is shown in Figure 11-22[23, 24]. By comparing with Figure 11-12, this system also consists of two LCD projectors, Fresnel screen, and a pair of glasses with linear polarizers, but two external polarizers are eliminated and the right LCD projector is physically rotated by 90 degrees.

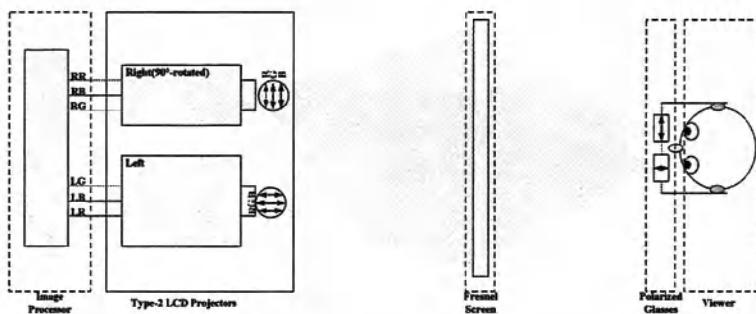
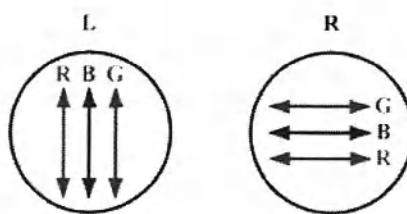


Figure 11-22. Overall optical layout of a new *Type-2* LCD polarized stereoscopic projection system

Figure 11-23(a) shows the left projector and it has output polarization of the normal Type-2 LCD projector, in which all color components of LCD projector are vertically polarized. If the right projector is physically rotated by 90° with respect to the left one then, all color components of the projected right image, which are originally polarized at the vertical direction, are converted into the horizontal polarization as shown in Figure 11-23(b). Thus, all color components of the left and the 90°-rotated right projectors are vertically and horizontally polarized, respectively. Accordingly, on the contrary to the case of *Type-1* LCD projectors, a process of color separation and exchange between two stereo images are not needed any more in case of *Type-2* LCD projectors, because two full linearly polarized color sets, which should meet the requirements for correct color balance and orthogonal polarization are already obtained just by rotating the right projector by 90° with respect to the left one as shown in Figure 11-23. Here the left color set has vertical polarization, whereas the right one has horizontal polarization.



(a) Left projector (b) 90°-rotated right projector

Figure 11-23. Output polarization of the left and 90°-rotated right projectors for the case of *Type-2* LCD

Figure 11-24 shows a flowchart of stereo image processing for the new *Type-2* LCD projector-based polarized stereoscopic projection system. In

case of *Type-2* LCD projector, because the polarization directions of all color components are equal, a process of color separation and exchange is not required. But, because the right projector has been initially rotated by 90° with respect to the left one, the right image projected from it is also rotated by 90° with respect to the left one, so that the new right image must be reformed in its orientation and aspect ratio to match with those of the left one through some image processing techniques before it is input to the 90° -rotated right projector just like the case of the *Type-1* LCD projector.

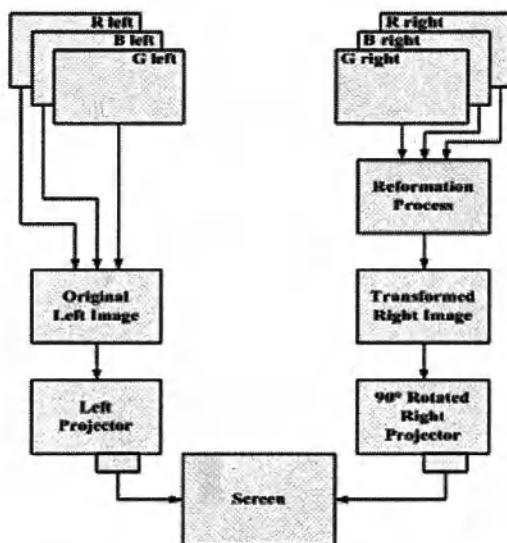


Figure 11-24. Flowchart of stereo image processing for the case of *Type-2* LCD projector

The original left image and the newly transformed right image are sent to the left and 90° -rotated right projectors, respectively and the size of the projected stereo image pair is also matched on the screen by using a 3D reform function of the projector. Then, by taking on same glasses used for the case of *Type-1* LCD projector, we can observe the newly transformed stereoscopic images on the screen with a good quality.

It is certain that output beam intensities of the new *Type-2* LCD-based polarized stereoscopic projection system must be improved by comparing with those of the conventional system. In case of the conventional *Type-2* LCD-based polarized stereoscopic projection system, half-wave retarders and polarizers must be employed, so that these optical devices inevitably cause lots of light loss and image distortion. According to the experimental results of A. Woods[17], 43% of light loss is measured in the conventional

Type-2 LCD-based polarized stereoscopic projection system. Moreover, it is also reported that light losses of the JVC's Type-2 LCD projectors of Model M1500 and G150CL caused by the half-wave retarders and polarizers are tested to be 49% and 39%, respectively[26].

2.3 Time-sequential stereoscopic projection display

2.3.1 LC shutter-based active eyewear system

As an another approach for stereoscopic projection displays, a time-sequential stereoscopic projection with a pair of LC shutters has been suggested, in which the two left and right views are alternatively projected on the screen and these two views are separated by using a pair of LC shutters synchronized with alternating video signals as shown in Figure 11-25[27].

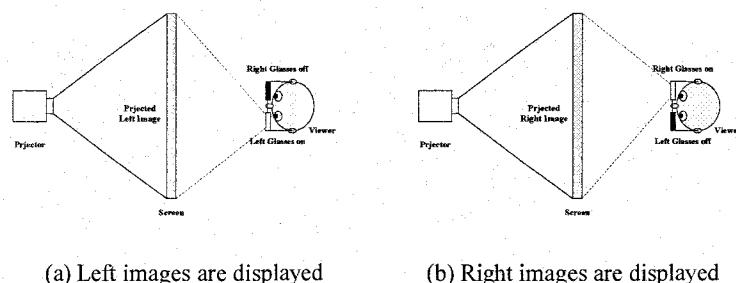
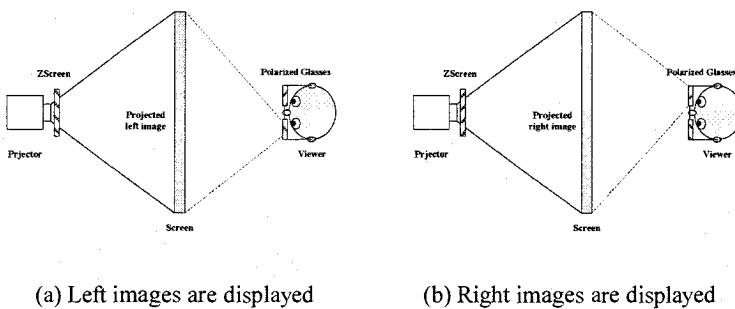


Figure 11-25. A schematic of LC shutter-based stereoscopic projection display

A pair of LC shutters is mostly employed as an active eyewear system. In this approach wireless battery-powered eyewear with LC shutters, which are running in synchrony with the video field rate, is employed. Synchronization information between the alternating video signals and a pair of LC shutters can be communicated by means of an infrared emitter. The emitter looks at the computer's video signal and seeing the vertical blanking synchronization pulse broadcasts coded IR pulses to signify when the left eye and the right eye images are being displayed[27]. The eyewear incorporates an IR detection diode which sees the emitter's signal and tells the eyewear shutters when to make on and off. The shutter is a sandwich made of two linear sheet polarizers (whose axes are orthogonal) on either side of the liquid crystal cell. The cell itself is made up of a thin film of LC material contained between two parallel sheets of glass.

2.3.2 ZScreen-based passive eyewear system

An alternative to the active eyewear system is the ZScreen, which is a special kind of LC polarization modulator[28]. It is placed in front of the projection lens just like a sheet-polarizing filter, and this changes the characteristic of polarized light and switches between left and right-handed circularly polarized light at the video field rate. Viewers are forced to wear circular polarizing analyzing eyewear in this stereoscopic projection system as shown in Figure 11-26. Although most of theaters putting stereo films on the screen uses linearly polarized light for image selection, circular polarized light has the advantage of allowing a great deal of head tipping before the stereoscopic effect is lost.



(a) Left images are displayed

(b) Right images are displayed

Figure 11-26. ZScreen-based stereoscopic projection display system

The ZScreen uses two liquid crystal cells (called pi-cells) in optical series with their optical axes orthogonal. The ZScreen uses phase shifting of linearly polarized light to achieve its electro-optical effect. A linear polarizer, at the surface of one of the cells (closest to the projection lens), has its axis oriented to bisect the pi-cells' orthogonal axes. The pi-cells are driven to quarter-wave retardation but with their electric potential out of phase. A low-voltage bias must be applied to the cells in order to tune their birefringence so that the vector sum of the cells' phase shift achieves a quarter-wave retardation.

The projector's light is first linearly polarized which is next subjected to phase shifts based on the voltages applied to the cells. In combination the cells function as a variable retarder so that left and right-handed circularly polarized light is output in synchrony with the video field rate. tipping is possible. The combination of ZScreen and polarizing eyewear forms a shutter, although one could make a case for classifying it as a polarization selection method.

3. TIME-SEQUENTIAL MULTIVIEW PROJECTION DISPLAYS

Depending on the development stages of the 3D display technologies, various degrees of resolution and spatial perception are now offered[29]. In stereoscopic projection displays discussed above, they require viewers to wear special glasses such as a pair of polarized or shuttered glasses to feel the depth, which has been regarded as a main shortcoming in practical applications. Therefore, many kinds of autostereoscopic projection displays have been developed, because autostereoscopic displays could present 3D image to viewers without a need for any special glasses. Here a time-sequential projection multiview display system developed in Cambridge University has been discussed.

3.1 Concept of time-sequential multiview projection display

The time-sequential multiview projection display, which has been developed in the rainbow group in Cambridge university, consists of a high-speed liquid crystal display, a Fresnel lens and a series of abutting bar shaped light sources as shown in Figure 11-27(a)[30]. The light sources are positioned just beyond the focal plane of the Fresnel lens so that an image of the light bars is projected into the user's view space(they called this image of the light bars the 'eye box'). Each light bar is illuminated in turn and, in synchronization with this, successive laterally adjacent views of an object are displayed on the high-speed liquid crystal display. The effect of the lens is that each view is visible in a different window in front of the display. Provided the multiviews are repeatedly illuminated sufficiently rapidly, the user can perceive a three-dimensional image with both stereo and horizontal movement parallax, as long as both of the eyes are within the eye box. Of course the best position from which to view autostereoscopic images is at the eye box, but a good 3D effect can be obtained over a large range of distances, from 50 cm to several meters[31].

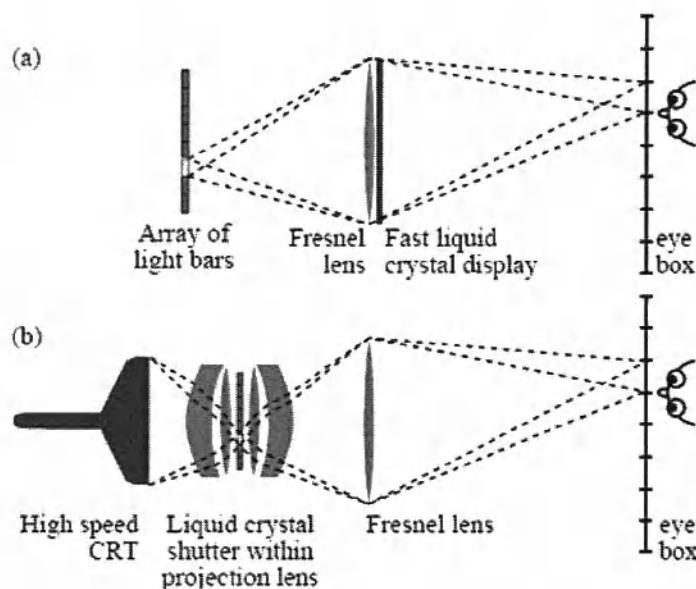


Figure 11-27. Time-sequential display; (a) Theoretical implementation, (b) Practical implementation

For the practical implementation of this time-sequential multiview projection display, a very fast liquid crystal display panel should be required. That is, in case 8-view images of an object might be displayed at a 60Hz refresh rate, it requires a liquid crystal display with a field rate of 480 Hz (8×60 Hz). Moreover, for the case of 32-view images, it would require a field rate of almost 2 kHz (32×60 Hz). So far these high field rate cannot be feasible with present nematic liquid crystals, but it may be attainable with ferroelectric liquid crystal displays(FLCD) if image data can be transferred to the ferroelectric liquid crystal arrays sufficiently quickly[32].

3.2 Practical implementation

A practicable monochrome 16-view version of a Cambridge display has been developed by use of a high speed CRT, an 'image transfer' lens and a ferroelectric LC shutter element in 1992 as shown in Figure 11-28 [33-35]. The implemented system is capable of 16-view at 320×240 resolution or 8-view at 640×480 (both interlaced) on a 10-inch diagonal screen. This requires horizontal and vertical scan rates of 150 kHz and between 400 and 1000 Hz, to give individual view direction refresh rates of 50–60 Hz, and an eye box of about 250 mm width at a viewing distance of 1 meter.

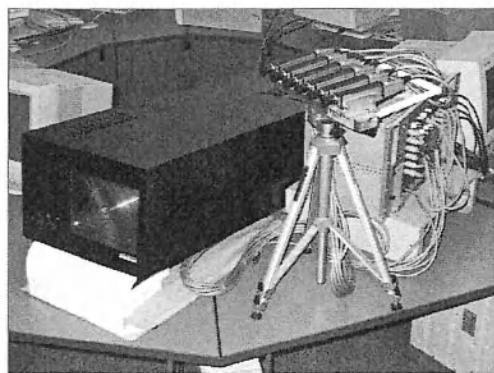


Figure 11-28. A practicable monochrome 16-view version of a Cambridge display

Figure 11-29 shows an optical schematic of the 50 inch Cambridge full-color projection display system[36]. All Cambridge projection displays built before 1995 were monochrome. Color was achieved in late 1995 using a color sequential solution, since shadow-mask based color CRTs are not capable of the required luminance.

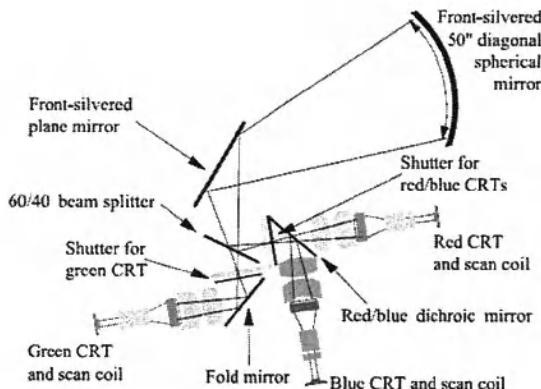


Figure 11-29. Optical schematic of the Cambridge full-color projection display system

In the prior system, a color sequential shutter is used, which gave no color convergence problems, but the system required a factor of three increase in the CRT frame rate to produce the three color fields needed in each cycle. As a result this prototype used four CRTs abutted together horizontally had shown three faint seams in the field of view. And in the posterior system, three-CRTs are used to represent the color image as shown in Figure 11-29. The different color fields in the chosen RGB system must

be combined at some point in the optical train to be co-axial. Green CRT phosphors are much more efficient than red and blue phosphors, so a neutral 60/40 dielectric beam-splitting surface is used to fold in the green, wasting 60% of that CRT light. And a single dichroic surface is used, which efficiently transmits red and reflects blue light, to combine these two colors with very little loss. This is because, with no green light at this point in the optical train, we can design the dielectric-coating stack to have a broad transition region in the green part of the spectrum.

4. PROJECTION INTEGRAL IMAGING DISPLAYS

Integral imaging (II) or real-time integral photography has been studied for optical display and visualization of true three-dimensional images in space with incoherent light[37]. In II, three-dimensional images are formed by crossing the rays coming from two-dimensional elemental images using a lenslet array. Although II can provide observers with true 3D images with full parallax and continuous viewing points as in holography, it has some drawbacks also. For example, the viewing angle, depth-of-focus and resolution of 3D images are limited because lenslet arrays are employed. In addition, 3D images produced in direct-pickup II are pseudoscopic(depth-reversed) images. To overcome those limitations, projection-type of II systems with either a lenslet array or a micro-convex-mirror array has been developed[38].

4.1 Projection types of integral imaging system

There are two kinds of projection types of II system, the lenslet array-based and the micro-convex-mirror array-based. A projection II using a lenlet array can be implemented as shown in Figure 11-30(a). The projector casts entire elemental images onto the corresponding lenslet array. Of course the diverging angle θ of the projection beam should be close to 0. In this projection II scheme using a lenslet array, however, the Pseudoscopic/Orthoscopic(P/O) conversion is required and the viewing angle is narrow as in the conventional scheme.

Figure 11-30(b) shows a 3D projection II scheme using a micro-convex-mirror array. In this case, orthoscopic virtual images are automatically reconstructed, when the elemental images obtained from direct camera pickup. This is because each micro-convex mirror does not rotate the corresponding elemental image around its own center optic axis in the 3D image reconstruction process. Therefore, the display of raw elemental images using a micro-convex mirror is exactly equivalent to the display of

P/O-converted elemental images using either a lenslet array or a micro-concave-mirror array. Each convex mirror element could have an $f/\#$ smaller than 1. For example, if $f/\# = 0.5$, the viewing angle ψ becomes 90 degrees, which is acceptable for many practical applications.

Note that optical barriers are not required in projection schemes, if 1) the focal length of micro-convex mirrors is shorter than the depth-of-focus of the relay optics, and 2) the diverging angle θ of the projection beam is small. These conditions are easily satisfied in projection II systems. Then, each elemental image is projected onto its own micro-convex mirror, i.e., elemental images are not displayed through their neighboring micro-convex mirrors. Therefore, as observers' viewing direction deviates from the optical axis that is normal to the display lenslet array, they do not experience flipping of the reconstructed 3D image.

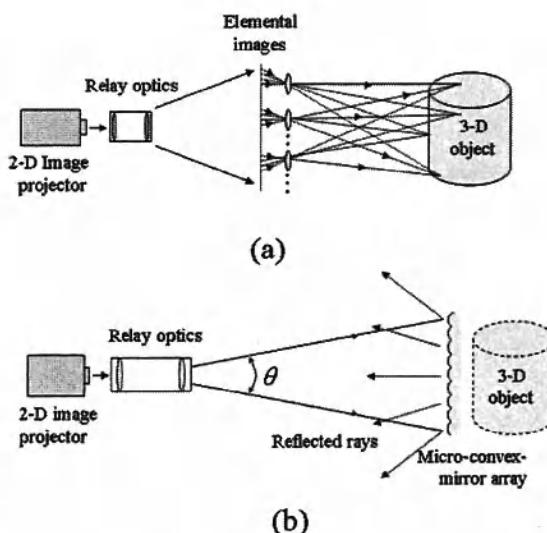


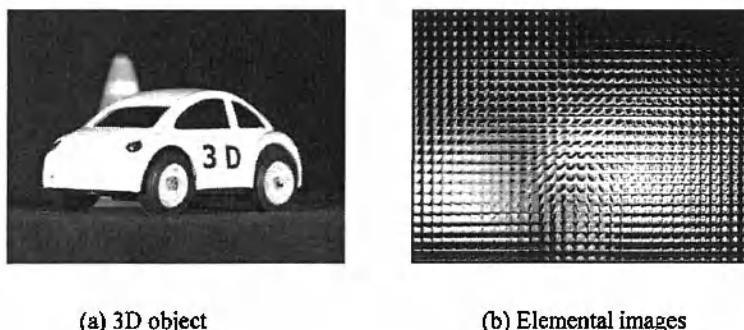
Figure 11-30. Projection II (a) with a lenslet array, (b) with a micro-convex-mirror array

4.2 Experiments on projection integral imaging

A feasibility of projection II using a micro-convex-mirror array could be demonstrated through a simple optical setup, which is composed of a micro-convex-mirror array, imaging lens and projector[38]. Figure 11-31(a) shows the 3D object to be imaged in the projection II experiments. A pickup lenslet array with 53×53 plano-convex lenslets and a color LCD projector with three panels were employed in the experiments. For the micro-convex-mirror

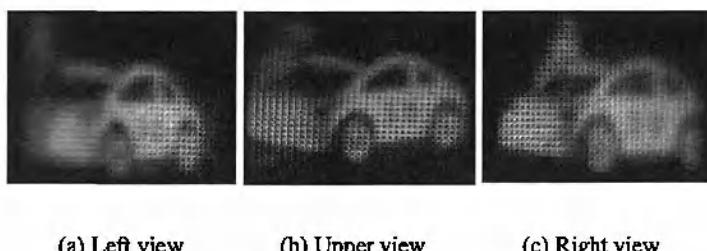
array, the pickup lenslet array itself was used. This is possible, because some light energy could be reflected for normally incident light at the surface of the lenslet array. Therefore, to produce the micro-convex-mirror array, the lenslet array is positioned so that the convex surfaces of lenslets face the 2D image projector. Then, the reflected light from the surface of the lenslet array can be observed. If the lenslet array is flipped so that the opposite side faces the 2D image projector, the effect of a micro-concave-mirror array can be obtained.

The raw elemental images obtained from direct camera pickup are shown in Figure 11-31(b) and they are projected onto the micro-convex-mirror array for reconstruction of a 3D orthoscopic virtual image. The left, upper, and right views of the reconstructed 3D image are finally shown in Figure 11-32, in which viewing directions for the three images were deviated from the optical axis by ~30 degrees, respectively. In case the lenslet array is used as a micro-concave-mirror array by facing its plano side to the viewer, the reconstructed image became a pseudoscopic real image, as we expected.



(a) 3D object

(b) Elemental images

Figure 11-31. 3D object and its elemental images

(a) Left view

(b) Upper view

(c) Right view

Figure 11-32. Reconstructed orthoscopic virtual 3D images

5. FRESNEL LENS-BASED PROJECTION DISPLAYS

5.1 Concept of image floating

Image floating is a simple 3D display method in which the feel of depth is emphasized using a floating lens. Figure 11-33(a) shows a schematic concept of the floating display[39]. The floating image, however, can be located in front of or behind the floating lens, and the position of the floating image is determined by the lens equation related with the focal length of the floating lens and the position of the object. A system in which the distance between the object and the floating lens is always a greater distance than the focal length is amenable for use in the floating system. Consequently, the floating system, the floating image of which is located in front of the floating lens, is mostly employed in the practical applications. As shown in Figure 11-33(a), the floating image is rotated by 180° because of the imaging phenomenon of the floating lens.

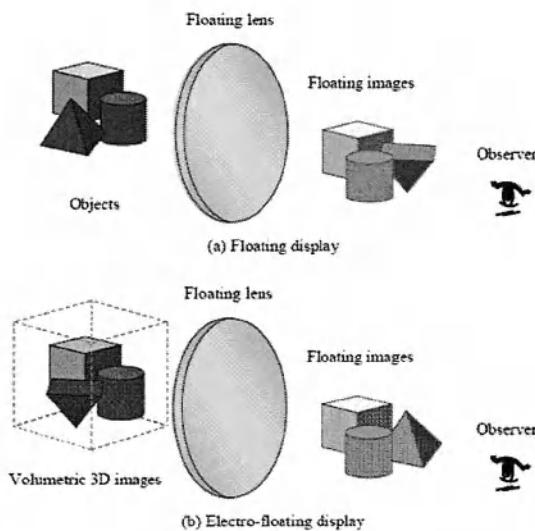


Figure 11-33. Concept of image floating

The floating display system cannot make a 3D image from a plane image but can only change the position of the image, which means that a 3D image cannot be generated by a floating display system in which an ordinary 2D display device replaces a real object. Therefore, the electro-floating display

system which does not use a real object requires a 3D display component which is capable of providing the volumetric 3D images in order to represent 3D moving pictures. Figure 11-33(b) shows the concept of the electro-floating display. In this case, the inversion of the floating image can be easily corrected through rotating the object image, as shown in Figure 11-33(b). The floating lens shown in Figure 11-32 can also be changed into a concave mirror because the concave mirror is similar to the convex lens from the viewpoint of optics. The floating image can be observed through the floating lens. As a result, the viewing area of the floating system is restricted by the aperture of the floating lens. This restriction becomes severe when the distance between the floating lens and the floating image, referred to as the floating distance, is longer, which is determined by the lens equation with the focal length of the floating lens and the distance between the floating lens and the object, referred to as the object distance.

5.2 Doublet Fresnel lens(DFL) system

In the floating display systems, which can provide a floating 3D image with real depth in the air, Fresnel lens is mostly employed as a floating lens[40-43]. Fresnel lens is regarded as a very useful and attractive optical device in case a large-size lens might be required in the optical system, because it can be made to be very thin and flat contrary to the conventional optical concave and convex lenses [44, 45]. In particular, a doublet Fresnel lens(DFL) can be effectively used for projecting a 2D image into the front focal length of it as a form of the floating image with some depth. Operating in conjunction with the display system such as a CRT or the like, the DFL is found to achieve a substantial image improvement in terms of image contrast, signal-to-noise ratio and field of view by comparing with those of the single Fresnel lens. A pair of Fresnel lenses acts as a single optical element, in which each of Fresnel lenses exhibits the same focal length. However, acting as a pair of Fresnel lens, the focal length will change to one-half that of the individual Fresnel lenses. Figure 11-34 shows a schematic diagram for the DFL system.

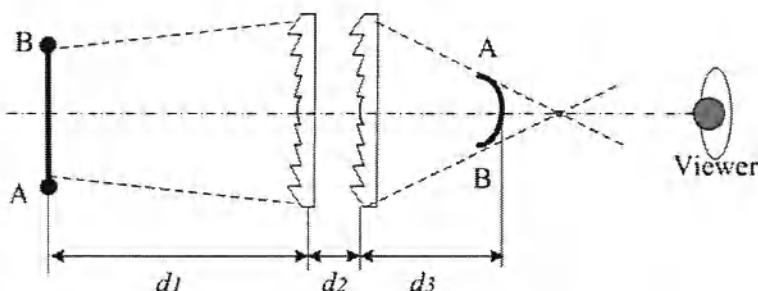


Figure 11-34. Schematic diagram of a DFL system

Figure 11-34 illustrates the operation principle of a DFL system. Generally, a single Fresnel lens system shows considerable chromatic aberrations, so that it causes some distortions in the floated images. To reduce this aberration of a single Fresnel lens, a DFL system might be preferred, in which the chromatic aberration and the focal point caused by refraction can be reduced. The DFL system shown in Figure 11-34 is composed of two Fresnel lenses having a positive focal length. The first lens acts as a collimator to direct light rays from points on a source image to the second lens and by using the second lens the light rays transmitted by the first lens are collected and focused in the front of the second lens. Thus, the DFL serves to direct and cause light rays from the source image to converge at locations in front of the second lens, so that a real image appears in front of the second lens.

In Figure 11-34, a source image is appeared as a line having two end points represented by A and B, and its projection along a common optical path is represented by the dotted line. A curved bold line A-B' in conjunction with projection lines can represent the substantially transmitted focal plane. As the floated image size is related to the spacing between two Fresnel lenses, d_3 , its size can be changed in proportion to the distance of the spacing d_2 . The distance of the spacing d_2 should be also selected as large enough to avoid the moiré fringe effects at the projected image.

The back focal length of the lens, d_1 , is related to the distance of floating image in front of the second lens, d_3 . That is, variation of the distance d_1 leads the changes of displayed image size and the floating distance d_3 in front of the screen. An increase of distance d_1 results in the decrease of the image size at the distance d_3 , whereas a decrease of the distance d_1 leads to the magnification of virtual images in front of screen.

At the distance d_3 , floating image creates a dynamic movement when an observer shifts its perspective. For all cases, the size of the aperture should

be at least as great as the corresponding dimension of the source image at screen.

When a flat image derived from the LCD or PDP display system is projected to such a curved focal plane, a form of object parallax is evoked wherein the viewer observes portions of the originally flat image at either side of the curved focal plane. In the projection system employing a schematic of Figure 11-34, this curved output focal plane may be generated in conjunction with relatively wide fields of view to permit observer or eye station motion about the projected focal plane and coincident image such that human cognition will tend to synthesize and interpret a three-dimensional effect.

5.3 DFL-based projection display system

Recently a large-scale DFL-based 3D projection display system called 'Holocube' has been implemented with a LCD projector, curved screen and DFL system as shown in Figure 11-35[40-43]. In this system, 2D video image is projected from the LCD projector in a rear projection mode and reflected from the curved screen and then, this image is floated into the space through the DFL system.

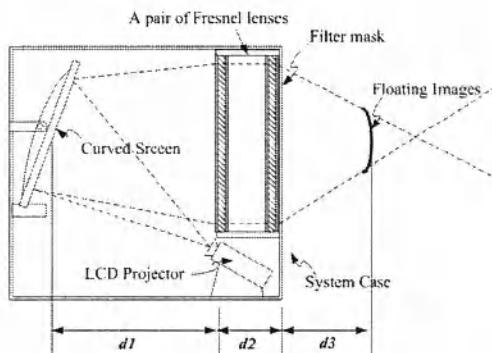


Figure 11-35. Schematic diagram of a DFL-based 3D projection display system

This system can be divided into two functional units; the display and rear-projection unit. The display unit consists of a DFL system and a dark film. The dark film could be used for improving the floated image contrast and observation condition. A pair of large planar Fresnel lens is placed behind this dark film. A commercial LCD projector is used for projecting the 2D video image onto the curved screen. The curved screen reflects the input

image back to the center of the DFL. Meanwhile the DFL was manufactured in all one form and its diagonal size was made to be 100 inch.

Figure 11-36 shows the original image on the video projector and its floated 3D image onto the space through the implemented projection display system.



(a) Original image

(b) Projected 3D image

Figure 11-36. Original image and its projected 3D image

A typical image size at the screen is 30 inches high and 40 inches wide. The output focal plane of the DFL is positioned about 35~47 inches in front of the exit aperture of the Fresnel lenses. The active size of the Fresnel lenses is 70 and 94 inches in height and width, respectively. The size of the 3D image is found to be approximately 35~47 inches and the viewing angle is also found to be approximately 60 degrees.

6. HOLOGRAPHIC PROJECTION DISPLAY

6.1 Introduction

Recently Harold R. Garner have constructed a system that projects true dynamic 3D holographic images from computer-generated holograms utilizing the lowest orders of diffracted light from a laser illuminated DMD[46]. They have demonstrated the utility of the DMD as a 3D image holographic medium by producing virtual and real 3D images at finite distances, an essential condition for image reconstruction with depth. Their aim is to create a real-time, multi-color projection system for all digital holograms.

6.2 Optical system

An optical system for reconstruction of dynamic holographic images by use of a DMD has been constructed as shown in Figure 11-37(a). It is composed of a 15 mW HeNe Laser, spatial filter, collimating lens with a focal length of 10 cm, DMD, converging lens with a focal length of 40 cm and an image reconstructor for real image viewing.

The real image reconstructor could be a frosted glass plate, fiber optic magnifier, or CCD/digital camera for visualization of a planar cross-section of a 3D image; or it could be a translucent block such as a thick Agarose gel to create a suspension of micro-scatter bodies to simultaneously view the whole 3D real image. The “original image” depicted in the upper left in Figure 11-37(a) is a bitmap of a 2D irregular perspective object. Its computed interferogram is represented on the computer monitor. The picture at the “image reconstructor” is a CCD camera photo of the actual image reconstructed on a frosted glass plate. This is an illustration of the DMD’s capability to reconstruct 2D irregular perspective objects as well as full 3D holographic scenes.

The 3D holographic virtual image can be observed by looking directly into the DMD as shown in Figure 11-37(b). The convergent lens and image reconstructor are removed from the optical system and the laser intensity is substantially reduced with neutral density filters for viewing directly by eye. The DMD functions as a reflective holographic medium in either projection mode.

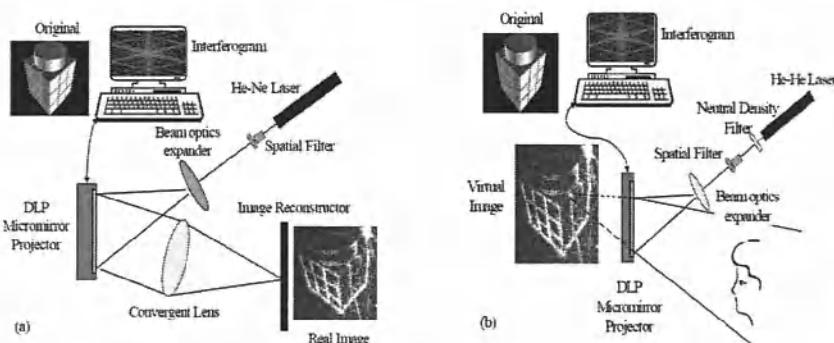


Figure 11-37. Schematic of DMD hologram projection system for (a) real and (b) virtual images.

6.3 System demonstration

For system demonstration, a simple scene containing two objects has been considered, which is shown in Figure 11-38. That is, a bitmap-image of a jet is placed at the rear plane ($z = 27.5$ cm), and a bitmap-image of a helicopter is placed at the front plane ($z = 30$ cm). For calculation of each interferogram, every pixel in the image at each plane was treated as a point source of light with intensity equal to its gray scale value.

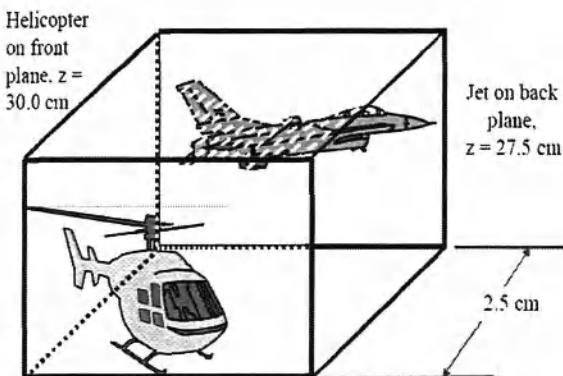


Figure 11-38. Bitmap-image layout to demonstrate 3D image projection with a DMD hologram.

The calculated interferograms can be scaled in software to adjust the size of the displayed scene and to associate them with physical spatial coordinates. Here in calculations of the hologram, a laser wavelength of 633 nm and a converging lens with a focal length of 25 cm and the bit-maps located as shown in Figure 11-38 were used.

In case the computer-generated hologram is transcribed to a HeNe laser illuminated DMD, a real and a virtual 3D image of the jet and helicopter can be seen as shown in Figure 11-38. For this demonstration, they located a 10cm collimating lens 10 cm in front spatial filter, the DMD 28 cm from the collimating lens, and the 40cm converging lens 11.5 cm from the DMD. The angle between the incident illumination axis and the image axis is about 20 degrees due to the cant angle of the mirrors [47, 48]. A frosted glass reconstructor was translated in front of the converging lens to image different slices of the real images produced at the two different focal ($z = 27.5$ cm and $z = 30.0$ cm from the 40 cm lens) positions, thus validating that the whole computer-generated scene was 3-dimensional. By taking photographs at each focal position, the original helicopter and jet was

obtained as shown in Figure 11-39(a) and 11-39(b), respectively. The different diffraction orders result in multiple images separated by the appropriate diffraction angles relative to the brightest diffraction order. Likewise for each diffraction order, there is a corresponding inverted image. The normal and inverted images occur at equal distances, plus or minus, from the focal distance of the lens. These other diffraction order images around the edges of the photos have been cropped for presentation.

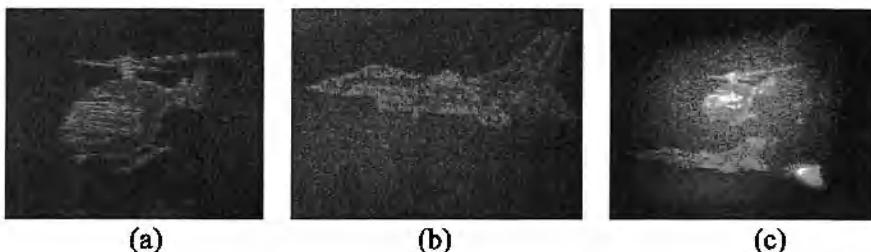


Figure 11-39. Reconstructed real and virtual holographic images (a) reconstructed real 3D image at 30 cm from the converging lens, (b) reconstructed real 3D image at 27.5 cm from converging lens, and (c) the virtual image looking into the DMD

The more traditional way to visualize the reconstructed 3D scene is to directly view the virtual image that appears in the DMD as shown in Figure 11-38(c). The helicopter could be appeared in front of the fighter just as in the real image. The camera was positioned in front of the DMD to capture the virtual image and focused for a distance of 3 ft and placed 22 cm in front of the DMD for this picture. The depth of field of the camera was maximized to simultaneously capture both objects in the virtual image volume.

REFERENCES

- [1] Jianmin Chen, Michael Robinson, Gary Sharp, *Polarization Engineering for LCD Projection*, John Wiley & Sons, Ltd, Chichester, 2005.
- [2] Sang-Soo Kim, "The World's Largest (82-in.) TFT-LCD," *Digest of Technical Papers*, Society for Information Display International Symposium, vol. 36, Issue 1, pp. 1842-1847, 2005.
- [3] <http://www.samsungsdi.com/>
- [4] Edward H. Stupp, Matthew S. Brennesholtz, *Projection Displays*, John Wiley & Sons, Ltd, Chichester, 1999.
- [5] R. Uokenbrock and W. Rowe, "Self-convergent, 3-CRT projection TV system," *Digest of Technical Papers*, Society for Information Display International Symposium, pp. 108-109, 1982

- [6] J. G. George, "Minimum adjustment analog convergence system for curved faceplate projection tubes," *IEEE Trans. Consumer Electro.*, vol. 44, Issue 3, pp. 536-539, 1995
- [7] Y. Itoh, J.-I. Nakamura, K. Yoneno, H. Kamakura, and N. Okamoto, "Ultra-high-efficiency LC projector using a polarized light illumination system," *Digest of Technical Papers*, Society for Information Display International Symposium, pp. 993-996, 1997.
- [8] R. L. Melcher, M. Ohhata, and K. Enami, "High-information-content projection display based on reflective LC on silicon light valves," *Digest of Technical Papers*, Society for Information Display International Symposium, pp. 25-28, 1998.
- [9] IBM, Special session for high-resolution displays, *IBM J. Res. Dev.*, vol. 42, no. 3/4, 1998.
- [10] M. F. Bone, M. Francos, P. Menard, M. E. Stefanov, and Y. Ji, "Novel optical system design for reflective CMOS technology," *Proc. of 5th Annual Flat Panel Display Strategic and Technical Symposium*, p. 81, 1998.
- [11] M. F. Bone, "Front projection optical system design for reflective LCOS technology," *Proc. of Microdisplay 2000*, 2000.
- [12] M. Robinson, J. Korah, G. Sharp and J. Birge. "High contrast color splitting architecture using color polarization filters," *Digest of Technical Papers*, Society for Information Display International Symposium, vol. 31, pp. 92-95, 2000.
- [13] L. J. Hornbeck, "128x128 deformable mirror device," *IEEE Trans. Electron Dev.*, ED-30, pp. 539-545, 1983.
- [14] <http://mems.colorado.edu/c1.res.ppt/ppt/g.tutorial/slide050.htm>
- [15] N. Holliman, *3D Display Systems, Handbook of Optoelectronics*, Institute of Physics Press, ISBN 0 7503 0646 7, November 2003.
- [16] O. Matoba, T. J. Naughton, Y. Frauel, N. Bertaux, and B. Javidi, "Real-time three-dimensional object reconstruction by use of a phase-encoded digital hologram," *Appl. Opt.* vol. 41, Issue 29, pp. 6187-6192, 2002.
- [17] Andrew J. Woods, "Optimal usage of LCD projectors for polarized stereoscopic projection," *Proc. SPIE*, vol. 4297, pp. 5-7, 2001.
- [18] Seung-Cheol Kim, Eun-Soo Kim, "A new liquid crystal display-based polarized stereoscopic projection method with improved light efficiency", *Optics Communications*, vol. 249, pp. 51-63, 2005.
- [19] Eun-Soo Kim, Seung-Cheol Kim, *Polarized stereoscopic display device and method*, PCT Patent, No. PCT/KR2004/001542, June 25, 2004.
- [20] Eun-Soo Kim, Seung-Cheol Kim, *Polarized stereoscopic display device without loss*, PCT Patent, No. PCT/KR2005/002283, July 15, 2005.
- [21] Seung-Cheol Kim, Dong-Kyu Kim, Dae-Heum Kim, Eun-Soo Kim, "A New Configuration of LCD Polarized Stereoscopic Projection System without Light Loss," *Proc. of SPIE*, vol. 5664A, pp. 439-449, San Jose, California USA, January, 2005.
- [22] Seung-Cheol Kim, Jae-Woong Moon, Dong-Hwi Lee, Eun-Soo Kim, "A New LCD-based Polarized Stereoscopic Projection Method with Improved Light Efficiency", *IMID 2005*, pp. 1607-1610, Seoul, Korea, July 2005.
- [23] Seung-Cheol Kim, Jae-Woong Moon, Dong-Whi Lee, Eun-Soo Kim, "A New Configuration of LCD Projectors for Polarized Stereoscopic Projection with Improved Light Efficiency", *Proc. of SPIE* vol. 5908, pp. 348-359, San Diego, California USA, August, 2005.
- [24] Seung-Cheol Kim, Dong-Whi Lee, Eun-Soo Kim, "Implementation of a New LCD Polarized Stereoscopic Projection System with Improved Light Efficiency", *Proc. of SPIE*, vol. 5962, Jena, Germany, September, 2005.
- [25] Victor A. Elkhov, Yuri N. Ovechkis, "Light loss reduction of LCD polarized stereoscopic projection," *Proc. of SPIE*, vol. 5006, pp. 45-48, 2003.
- [26] http://www.jvcdig.com/3d_kits.htm

- [27] S. Pastoor and M.D. Wöpking, "3-D displays: a review of current technologies," *Displays*, vol. 17, pp. 100–110, 1997.
- [28] Vivian Walworth, "Three-dimensional projection with circular polarizers," *Proc. of SPIE*, vol. 462, Optics in Entertainment II, 1984.
- [29] Neil A. Dodgson, "Autostereoscopic 3D displays", *Computer*, Vol. 38, Issue. 8, pp. 31-36, 2005.
- [30] Adrian R. L. Travis, "Autostereoscopic 3-D display," *Applied Optics*, vol. 29, no. 29, pp. 4341–4342, 1990.
- [31] Neil A. Dodgson, "Analysis of the viewing zone of the Cambridge autostereoscopic display," *Applied Optics*, vol. 35, no. 10, pp. 1705–1710, 1996.
- [32] Adrian R. L. Travis and S. R. Lang, "The design and evaluation of a CRT-based autostereoscopic 3-D display," *Proc. of SID*, vol. 32, pp. 279–283, 1991.
- [33] S. R. Lang, A. R. L. Travis, O. M. Castle and J. R. Moore, "A 2nd generation autostereoscopic 3-D display," *Proc. of 7th Eurographics Workshop on Graphics Hardware*, pp. 5–6, 1992
- [34] J. R. Moore, A. R. L. Travis, S. R. Lang, and O. M. Castle, "The implementation of a multi-view autostereoscopic display," *IEE Colloquium on Stereoscopic Television*, 15 Oct 1992, IEE, London, UK Digest No. 1992/173, 4/1–4/16, 1992.
- [35] A. R. L. Travis, S. R. Lang, J. R. Moore and N. A. Dodgson, "Time-multiplexed three-dimensional video display," *J. Soc. for Information Display*, vol. 3, no. 4, pp. 203–205, 1995.
- [36] N. A. Dodgson, J. R. Moore, S. R. Lang, G. Martin and P. Canepa, "A 50" time-multiplexed autostereoscopic display", *Proc. SPIE 3957, SPIE Symposium on Stereoscopic Displays and Applications XI*, 23-28 Jan 2000
- [37] Bahram Javidi, Fumio Okano, *Three-Dimensional Video and Display: Devices and Systems*, SPIE-International Society for Optical Engine, 2001.
- [38] J. S. Jang, B. Javidi, "Three-dimensional projection integral imaging using micro-convex-mirror arrays," *Optics Express*, vol. 12, no. 6, pp. 1077-1083, 2004.
- [39] Sung-Wook Min, Minsoo Hahn, Joohwan Kim and Byoungcho Lee, "Three-dimensional electro-floating display system using an integral imaging method," *Optics Express*, vol. 13, no. 12, pp. 4358-4369, 2005.
- [40] Eun-Soo Kim, Sun-Ju Jang, *Three-demensional display device for terminal*, PCT Patent, No. PCT/KR2005/000718, March 11, 2005.
- [41] Eun-Soo Kim, Sun-Ju Jang, *Three-dimensional display device*, PCT Patent, No. PCT/KR2005/000719, March 11, 2005.
- [42] Sun-Joo Jang, Seung-Cheol Kim, Jung-Sik Koo, Jung-Ill Park, Eun-Soo Kim, "100 inch 3D real image rear-projection display system based-on Fresnel lens," *Proc. of SPIE*, vol. 5618, pp. 204-211, 2004.
- [43] Jae-Woong Moon, Gil-Taek Lim, Sun-Joo Jang, Jung-Sik Koo, Jung-II Park, Eun-Soo Kim, "Compensation of image distortion in Fresnel lens-based 3D projection display system using a curved screen", *Proc. of SPIE*, vol. 6016, 2004, In publish.
- [44] H. Kakeya and Y. Arakawa, "3D autostereoscopic display with real-image screen," *Proc. of 3D Image Conference*, pp. 3-68, 2000.
- [45] D. Ezra, B. A. Omar, and G. J. Woodgate, "Autostereoscopic directional displays apparatus," *European Patent Application*, EP 0 602 934 A2, 1993.
- [46] Michael Huebschman, Bala Munjuluri, and Harold R. Garner, "Dynamic holographic 3-D image projection," *Opt. Express*, vol. 11, pp. 437-445, 2003.
- [47] L.J. Hornbeck, "Digital light processing for high-brightness, high-resolution applications," *Electronic Imaging, EI '97, Projection Displays III*, SPIE, San Jose, CA, 10 February 1997.

- [48] L. Yoder, et al., *DLP Technology: Applications in Optical Networking*, Texas Instruments White Pages, 2001,
http://www.dlp.com/dlp_technology/dlp_technology_white_papers.asp.

Chapter 12

THREE-DIMENSIONAL DISPLAY AND INFORMATION PROCESSING BASED ON INTEGRAL IMAGING

Byoungho Lee¹, Jae-Hyeung Park¹, and Sung-Wook Min²

¹*School of Electrical Engineering, Seoul National University, Kwanak-gu Shinlim-dong, Seoul 151-744, Korea;* ²*Digital Media Lab, Information and Communications University, 517-10 Dogok-dong, Gangnam-gu, Seoul 135-854, Korea*

Abstract: We discuss the three-dimensional (3D) display technique using integral imaging and its application to 3D information processing. First, the fundamental concept of integral imaging as a 3D display technique is presented. Then recent researches to improve the quality of integral 3D imaging are overviewed. The extension of the integral imaging technique for the acquisition and correlation of 3D information is also described.

Key words: Integral imaging, integral photography, three-dimensional display, depth extraction, three-dimensional correlation

1. INTRODUCTION

1.1 Integral imaging for three-dimensional display

Integral photography is a three-dimensional (3D) display technique that was first proposed by Lippmann in 1908.¹ Since then, the technique has been further developed and improved by many research groups, especially after the demonstration of real-time 3D image displays.² The technique is now usually referred to as integral imaging rather than integral photography to emphasize its use in the area of dynamic image display. Integral photography is analogous to holography in that 3D information is recorded

on a two-dimensional (2D) medium. Although integral photography was proposed about 40 years earlier than holography, it has some attractive points compared with holography: it does not require any coherent light source and current integral imaging techniques enable the display of full natural color dynamic 3D images.

Figure 12-1(a) shows the basic concept of Lippmann's system. The objective of Lippmann's system was to create a directly viewable (autostereoscopic) 3D image. Integral photography resembles dragonfly's eyes. A 3D object is recorded on a film (or photographic plate) surface in the form of a 2D image array, called an elemental image array, by means of a lens array. The lens array consists of a number of small identical (or inhomogeneous in some cases) lenses, called elemental lenses. This recording process of integral photograph is called a 'pickup'. The recorded elemental images contain parallax information, and thus each elemental image has a different view observed from a different viewing direction. If the object distance is much greater than the focal length of the lens array, the photographic emulsion is located approximately in the focal plane of the lens array. After exposure, the photographic plate is developed in reverse. The developed plate and the decoding lens array are then reassembled with proper alignment and illuminated from the rear by diffused light. This process is called a 'reconstruction' or 'display'. Because the light rays passing through a decoding lens array retrace the original routes, they converge at the point where the original object had been originally located. These rays form an autostereoscopic 3D image that exhibits full natural color (if a color film was used) and continuous parallax within a viewing zone. However, in time of Lippmann, high-resolution lens arrays were not commercially available. Even in the 1960's, a microlens array was difficult to fabricate and not economically feasible for practical use.

Unfortunately, all the methods of using photographic plate or film for recording and displaying an image are not suitable for moving objects. This static feature has been limiting in applications of integral photography. In the 1990's, using electronic pickup devices instead of the original idea of employing photographic film, real-time 3D integral imaging, such as Fig. 12-1(b), became possible, opening the possibility for use in 3D television or broadcasting. Okano *et al.* proposed and implemented this real-time integral photography,² i.e., integral imaging, that offers 3D autostereoscopic images of moving objects by use of a high-definition television camera and a display panel. Unlike classical integral photography, this method enables us to shoot moving 3D pictures in real-time. In real-time integral imaging, both the pickup of a moving 3D object and its real-time transmission are possible. With the development of active devices such as high-resolution cameras and flat panel display devices, the technical and economic problems can be

overcome. Furthermore, fabrication technologies for refractive micro-lens arrays have also advanced significantly over the past few years.

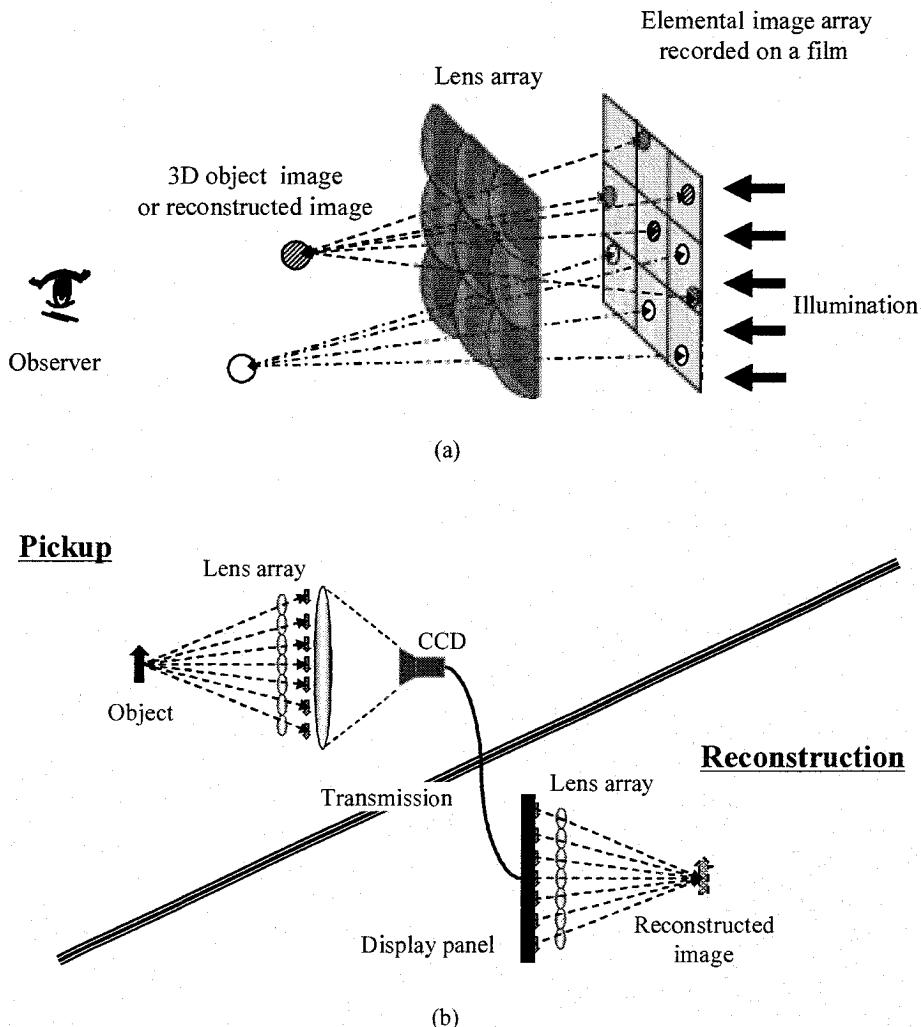


Figure 12-1. Basic concept of (a) Lippmann's system and (b) real-time integral imaging.

In an integral imaging system, the reconstructed image is a volumetric image, which exists in space at its generated location independent of the viewer's position. This occurs because, unlike stereoscopic techniques which present planar perspective views to the viewer's eyes, each point within the volume of an integral image is generated by the intersection of ray pencils that are projected by the individual elemental lenses. An additional

advantage of this technique is that the reconstructed image retains its correct scale, independent of viewing distance, that is, it appears to exist in space at its original location. In stereoscopic display systems, the magnification of the image in depth direction is changed while viewing distance is varying. For example, the magnification of the depth of objects increases while observer moves farther from the display device.

In summary, the technique of integral imaging has many advantages as follows.

- It requires no special viewing-aids (autostereoscopy).
- It provides continuous viewpoints within the viewing angle.
- It provides full parallax (both horizontal and vertical parallax).
- It provides natural depth perception with relatively low eye-fatigue.
- It can display full-color and real-time 3D animated (dynamic) images.
- Multiple observers can see 3D images freely within the viewing angle.
- Display devices of current 2D technology can be easily adopted.

However, conventional integral imaging technology also suffers from the following drawbacks.

- Limited viewing-resolution
- Limited viewing-angle
- Limited image-depth range
- Difficulty in compatibility with 2D images

Various studies have been investigated to overcome or mitigate these problems as will be reviewed in the following sections.

1.2 Integral imaging for three-dimensional information processing

3D information processing provides a direct method for handling 3D information of the world. Additional dimensionality places excessive data in the user's hand, conferring real benefits. Integral imaging is a powerful method in realizing 3D information processing. Originally, as explained previously, integral imaging was a 3D display technique using a lens array. With recent rapid developments, however, the range of applications of integral imaging has now been expanded to include the acquisition and the correlation of 3D information as well as the display.

Integral imaging handles 3D information by manipulating directional ray distribution using a lens array. Directional ray distribution is radiance as a function of position and direction in regions of space free of occluders. It is a five-dimensional function (three for space and two for directions), but it can be reduced to four dimensions (two for space and two for directions) in the case of no occluders. Since each point on the 3D object emits or reflects light

rays in all directions with the corresponding color or intensity of that point, the directional ray distribution contains complete 3D information concerning the objects. Because each elemental lens captures rays passing through its principal point, the lens array is optically equivalent to multiple cameras that are distributed spatially. An abundant number of the elemental lenses of a lens array provide an efficient method for sampling the directional ray distribution in a dense grid.

1.3 Outline

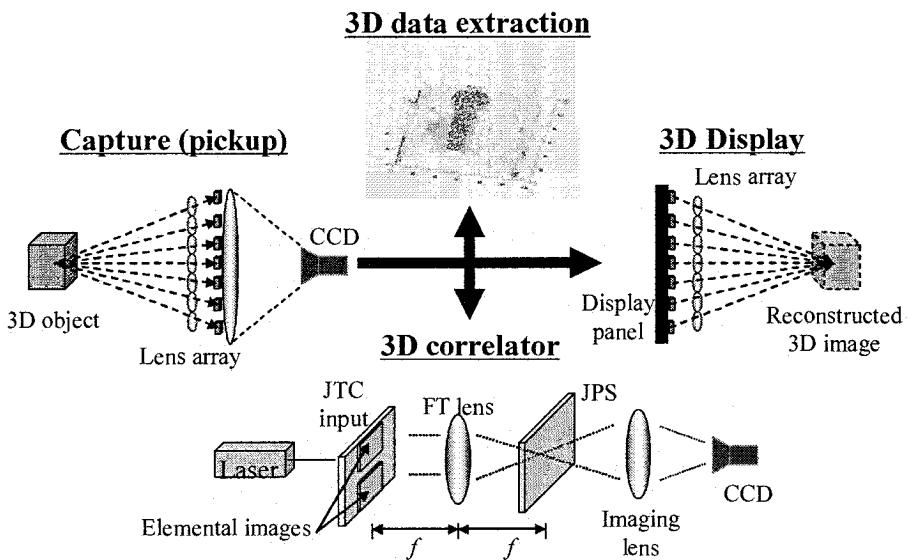


Figure 12-2. Optical 3D information processing using lens array. (JTC: joint transform correlator, JPS: joint power spectrum, FT: Fourier transform)

In this chapter, we review the 3D display technique using integral imaging and also present applications of integral imaging to the processing of 3D information. First, the fundamental concept of integral imaging as a 3D display technique is presented. With the basic principle in hand, its viewing parameters and endeavors to enhance them are introduced. Recent developments directed toward overcoming or mitigating the aforementioned drawbacks of integral imaging for 3D display will be discussed. Its extension for the acquisition and correlation of 3D information are then described. For the acquisition of 3D information, a method for extracting depth information of an object by detecting the disparity between the elemental images is explained. In addition to this elemental image based method, a one-

dimensional (1D) sub-image based method is also presented. This method removes ambiguity in the correspondence analysis and solves the small field of view problem by different disparity characteristics along the horizontal and vertical directions. For the processing of the 3D information, an optical 3D correlator is described. Using 2D sub-images, not only the 3D location but also the out-of-plane rotation can be detected optically. The overall concept of optical 3D information processing based on integral imaging is depicted in Fig. 12-2.

2. FUNDAMENTALS OF INTEGRAL IMAGING

2.1 Computer generated integral imaging

In some applications, it may be necessary to display an object that does not actually exist and whose structural information is stored in the form of electrical signals. Integral imaging mixed with computer graphics is highly attractive due to its simple structure and various application areas. For this purpose, analogous to computer-generated holography, computer-generated integral imaging, which displays a 3D scene that does not actually exist, has been developed.^{3,4} The basic idea of this method is to generate elemental images of any object by a computer.

We assume an imaginary object that contains 3D information. If we define the lateral plane which is parallel to the display panel as the x - y plane, the z direction represents information relative to depth. An imaginary object possesses information concerning a 3D location (x, y, z). This information can be transformed into the form of a 2D elemental image array assuming observation from different viewing directions through different elemental lenses in a lens array. Therefore, a point in 3D space is mapped to an array of 2D dots. We refer to this process as 3D to 2D mapping. This mapping is performed for points that are used to build up a 3D object, following the steps in Fig. 12-3.

First, the imaginary object is assumed as a set of points on multiple planes with different depths along the depth direction. A set of elemental image points can be obtained for a single point of an object. The procedure is as follows. Information on the x - y coordinates of a point in an imaginary object determines the center of the corresponding elemental image points on the imaginary pickup or display device plane. Based on information on the z coordinate of an imaginary object, the positions and the number of elemental images can be calculated and, as a result, a set of elemental image points can be obtained. It is noteworthy that this process is performed only when the

condition for the existence of elemental images is satisfied. In order to calculate the elemental images of the whole object, the above process is performed repeatedly until a set of elemental images for all points of an imaginary object are generated. By summing up total elemental image sets for all point objects, a complete elemental image set of an arbitrary imaginary object can be obtained. Figure 12-4 shows an example of the computer-generated elemental images and their 3D reconstruction.

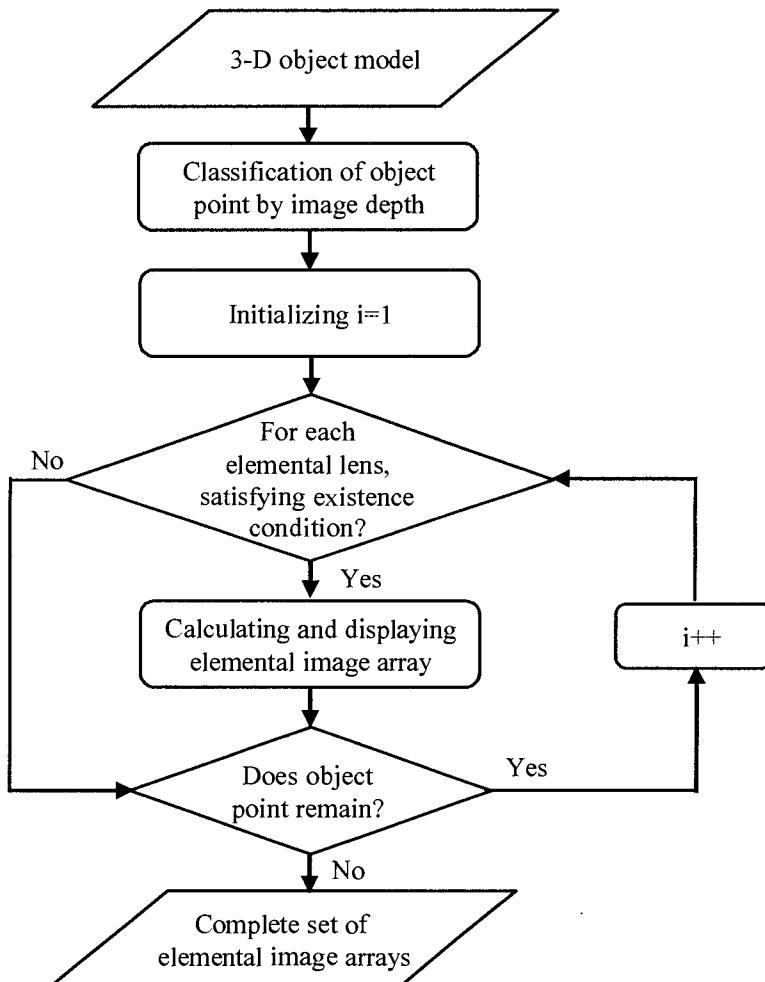


Figure 12-3. The procedure for generating elemental images in computer-generated integral imaging.

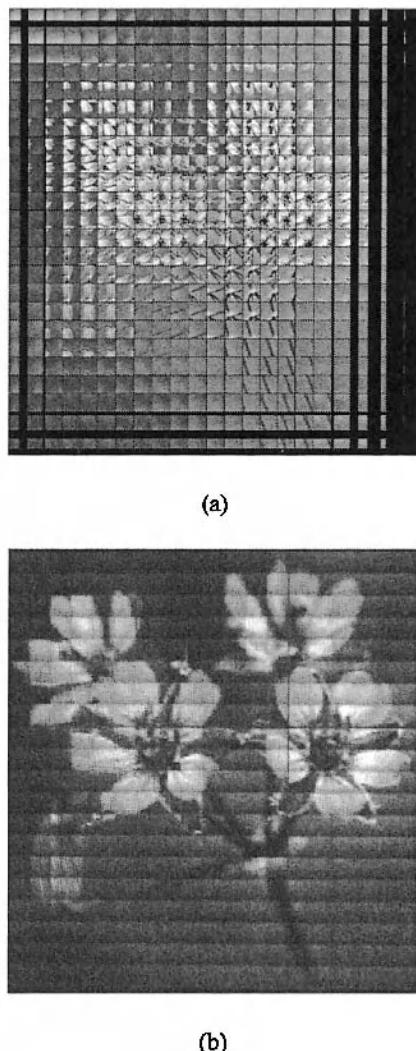


Figure 12-4. (a) Elemental images generated by computer graphics following the steps in Fig. 12-3, and (b) the reconstructed image.

2.2 Viewing characteristics of integral imaging

The integral imaging system has unique viewing characteristics because the reconstructed images of the system must be observed through a lens array consisting of elemental lenses. The integral imaging display system has three different display modes; real, virtual and focused modes, which are induced by the characteristics of the convex lens.⁵⁻⁸ Figure 12-5 shows the concept of image integration for different display modes. When the gap g

between the lens array and the display device is longer than the focal length f of the elemental lens, i.e., $g > f$, the image focused by the elemental lens is located in front of the lens array. Here, the focal plane (image plane for each elemental lens) of the lens array is named as the central depth plane (CDP). The separation l of the CDP from the lens array is given by the Gauss lens law (imaging equation):

$$\frac{1}{g} + \frac{1}{l} = \frac{1}{f}. \quad (12-1)$$

Although the highest resolution image can be reconstructed at the CDP as can be seen in Fig. 12-5(a), we wish to implement a volumetric image around the CDP rather than a planar image, as shown in Fig. 12-5(b). In this figure, the integrated image is located on the plane where the rays which start from each elemental image and pass through the focal plane (CDP) of the corresponding elemental lens create the cross sections, referred to as the integration plane. The integrated image is too misty or broken to be observed when the distance between location of the integration plane and the central depth plane exceeds a certain value. Accordingly, the integration plane cannot be much distant from the central depth plane and the solidity of the reconstructed image is restricted within some boundary of the integration planes. As shown in Fig. 12-5(b), the CDP and the integration planes of the real mode are located in front of the lens array. Although Fig. 12-5(b) shows only one integration plane, a volumetric image has various integration planes across the volume around the CDP. When the gap between the lens array and the display device is shorter than the focal length, i.e., $g < f$, the integrated image is displayed in the virtual mode, i.e., $l < 0$ in Eq. (12-1), where the CDP and the integration planes are located behind the lens array, as shown in Fig. 12-5(c). When the gap is almost equal to the focal length, the integrated image can be simultaneously displayed in front of and behind the lens array. This display mode is referred to as the focused mode, where the rays from each elemental point image are parallel and the focused point of each elemental lens does not exist, as shown in Fig. 12-5(d). Although the CDP of the focused mode does not exist (located at an infinite distance), we can consider it as the elemental lens plane as will be explained later.

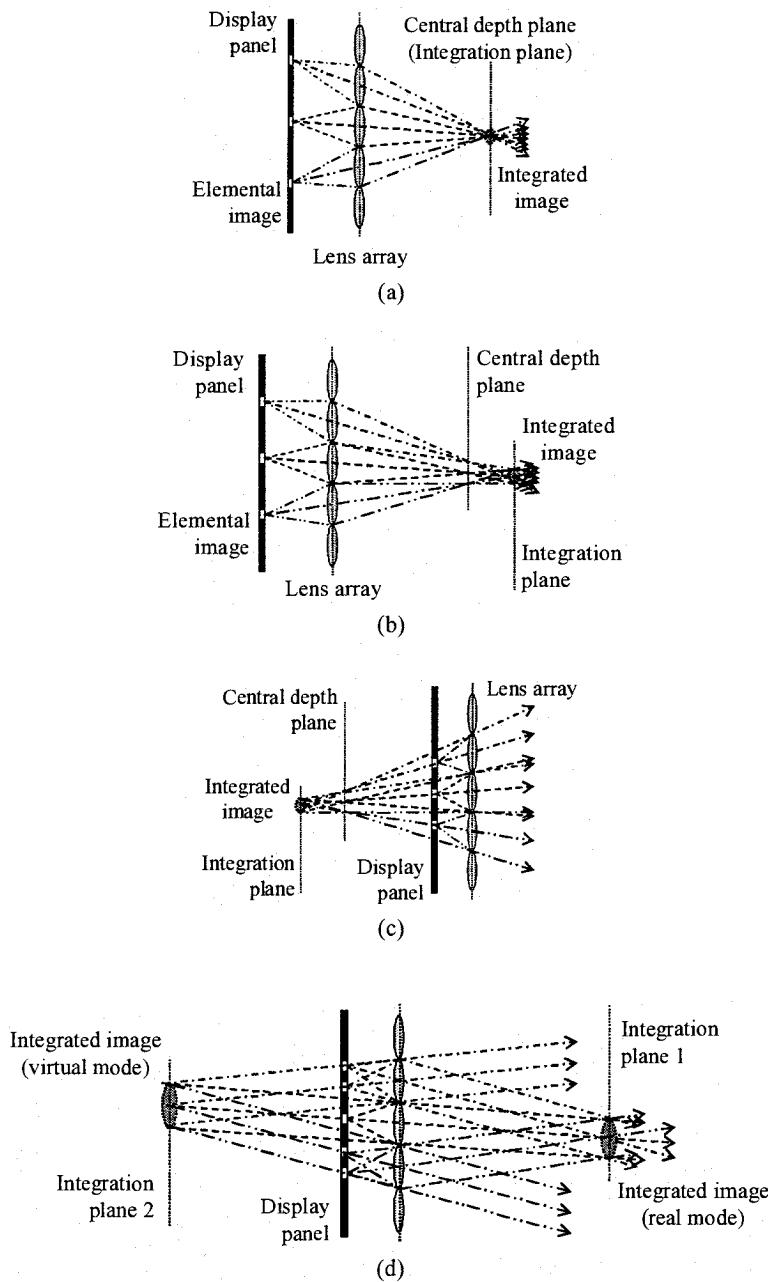


Figure 12-5. Concept of image integration for different modes: (a) real mode when the integration plane is located on CDP, (b) real mode when the integration plane is different from CDP, (c) virtual mode, and (d) focused mode.

The viewing characteristics of the integral imaging display system are different depending on the display modes. When the other conditions of the system are similar, the viewing angle of the real mode is narrower than that of the virtual mode. However, the feel of depth of the real mode is certainly superior to that of the virtual mode because the integrated image of the real mode is closer to the observer. In the focused mode, the depth of the integrated image can be expressed more deeply than the other modes, but the resolution of the integrated image is degraded because the pixel size of the image is fixed as the size of elemental lens. The purpose of the system should be taken into consideration when choosing the display modes of the integral imaging display system.

The viewing characteristics of the integrated image can be analyzed using three viewing parameters: (1) image resolution, (2) viewing angle, and (3) image depth.

(1) Image resolution is the resolution of the reconstructed image, which is determined by the resolution of the display device, the gap between the lens array and the display device, and the focal length of the lens array. Image resolution is the reciprocal of the implemented pixel size on the CDP where the image is focused by the lens array.

(2) Viewing angle is the angular region where the integrated image can be observed without cracking and flipping. The limitation in viewing position occurs because both the size of the elemental lens and that of the pixel area for displaying each elemental image (which corresponds to each elemental lens) are limited. If elemental images are observed not through corresponding elemental lenses but through neighboring lenses, they lead to image flipping, the duplication of the original integrated image. A detailed explanation of image cracking and its minimization in computer-generated integral imaging can be found in References⁹⁻¹¹.

(3) Image depth is defined as the thickness of the integrated image around the CDP. The integrated image, however, becomes more broken and distorted as it becomes located farther from the CDP. The term image depth is sometimes used to simply refer to the distance between the integrated image and the lens array.

Viewing angle is influenced by the size of the elemental lens and the gap. The viewing angle increases as the gap decreases and the size of the elemental lens increases. However, a short gap degrades the image resolution, and a large elemental lens reduces the depth of the image. Accordingly, to increase each of the above-mentioned viewing parameters of an integral imaging system, the other viewing parameters must be degraded. That is, a trade-off exists among the viewing parameters of the integral imaging system.

The exact analysis of the relationship among the viewing parameters requires mathematical definitions of the viewing parameters. The definitions obtained by ray-optical approximation, which can give, not the exact solution, but the tendency of the viewing parameter, are sufficiently adequate for a relationship analysis.⁸ Wave-optic analysis on integral imaging can be found in Refs. 6, 12 and 13.

Figure 12-6 shows the scheme for defining the viewing parameters of the integrated image. Since the pixel size of the integrated image is determined by the product of the pixel size of the display device and the magnification of the elemental lens, the image resolution R_I can be approximately expressed as Eq. (12-2), disregarding the diffraction effect caused by the lens array.

$$R_I = \frac{1}{P_I} = \frac{g}{lP_X} = \frac{g}{l} R_X, \quad (12-2)$$

where P_I is the pixel size of the image, P_X the pixel size of the display device, R_X the resolution of the display device, g the gap, and l the position of the central depth plane, as shown in Fig. 12-6. According to Eq. (12-2), image resolution is proportional to the gap and the resolution of the display device, and inversely proportional to the position of the CDP. The maximum image resolution cannot exceed the reciprocal of the minimum spot size of the elemental lens. When the pixel size of the image is larger than the size of the elemental lens, the image resolution is fixed as the reciprocal of the lens pitch. This is because the pixel size, which is represented by the lens array, cannot exceed the size of the elemental lens. In this case, image resolution is given by

$$R_I = \frac{1}{P_L} \quad (\text{when } R_I \geq P_L), \quad (12-3)$$

where P_L is the size of the elemental lens.

The tangent of one-half of the viewing angle can be roughly given by

$$\tan\left(\frac{\Omega}{2}\right) = \frac{P_L}{2g}, \quad (12-4)$$

where Ω is the viewing angle (see Fig. 12-6). As expressed in Eq. (12-4), the viewing angle increases with the increase in the elemental lens pitch and with the decrease in the gap. Because Eq. (12-4) is derived under the simple assumption that the integrated image is only one point and one elemental

lens is substituted for the lens array, the actual viewing angle, which can be obtained by a complex calculation including more detailed specifications of the system such as the size of the integrated image, is smaller than that obtained using Eq. (12-4).¹⁴ However, the tendency of this actual viewing angle follows that of the calculated viewing angle.

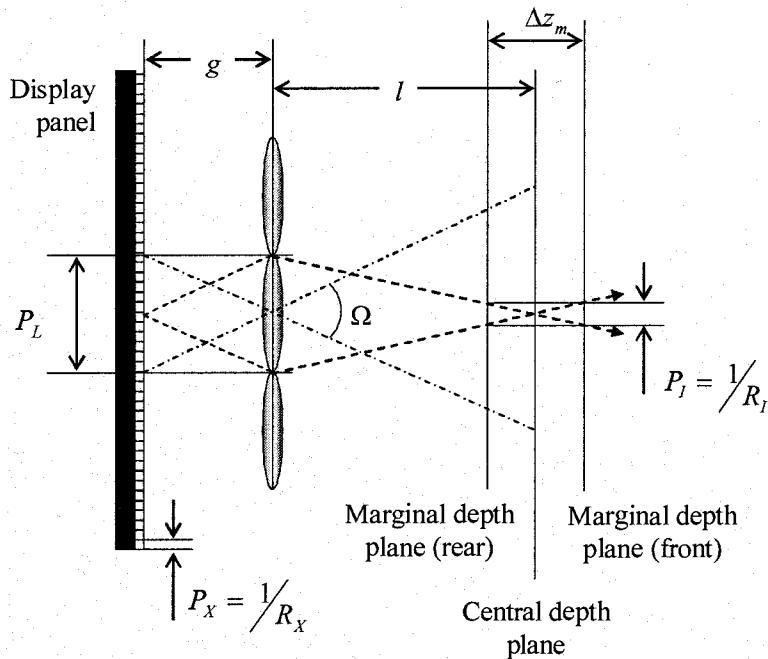


Figure 12-6. Scheme of viewing parameters in an integrated image.

As mentioned above, the integrated image is too broken to be observed when the image is located quite far from the CDP. The margin of the image depth can be set as the distance between the CDP and the marginal depth plane where the ray-optical focusing error becomes equal to the size of the image pixel (at CDP) because it can be assumed that the error of image integration occurs due to the overlap of the image pixel. The image depth Δz_m (see Fig. 12-6) of the integral imaging system can then be expressed as

$$\Delta z_m = 2 \frac{l}{P_L} P_I, \quad (12-5)$$

where the coefficient 2 indicates that the marginal depth planes are located in front and on the rear side of the CDP. The diffraction effect of the lens

array can also be disregarded in the derivation of Eq. (12-5). The image depth is proportional to the position of the CDP and the pixel size of the image, and inversely proportional to the elemental lens pitch, as shown in Eq. (12-5).

When the gap is close to the focal length, the pixel size of the integrated image is fixed as the pitch of the elemental lens, as shown in Eq. (12-3). In this case, the integrated image can be displayed in the focused mode. We may assume, for convenience, that the CDP is located at the elemental lens plane because the plane distinguishes between real and virtual images. The position of the marginal depth plane is determined by the lens pitch multiplied by the ratio of the gap to the pixel size of the display. Thus, in this case, the equation of the image depth is expressed as

$$\Delta z_m = 2 \frac{g}{P_x} P_L \quad (\text{when } R_I = \frac{1}{P_L}). \quad (12-6)$$

According to Eqs. (12-2), (12-4) and (12-5), it is certain that the viewing parameters of the integral imaging system exhibit a trade-off relationship. The mathematical relationship of the viewing parameters can be expressed as

$$R_I \Delta z_m \tan\left(\frac{\Omega}{2}\right) = R_X. \quad (12-7)$$

Equation (12-7) is named as the characteristic equation of the integral imaging system.⁸ This characteristic equation shows that the product of the square of the image resolution, the image depth, and the tangent of half of the viewing angle is equal to the resolution of the display device. Even when the gap is close to the focal length, i.e., when Eqs. (12-3) and (12-6) are valid instead of Eqs. (12-2) and (12-5), the characteristic equation for the system, i.e., Eq. (12-7), is still preserved. The effective range of the characteristic equation is the range where all of the approximations and assumptions concerning the viewing parameters are valid. The characteristic equation certifies the trade-off among the viewing parameters and reveals that the overall viewing quality cannot be improved if the resolution of the display is fixed. Therefore the type of lens array can affect the relative values of the viewing parameters, but a trade-off exists among the parameters, which can be improved by improving the resolution of the display.

3. 3D DISPLAY BASED ON INTEGRAL IMAGING

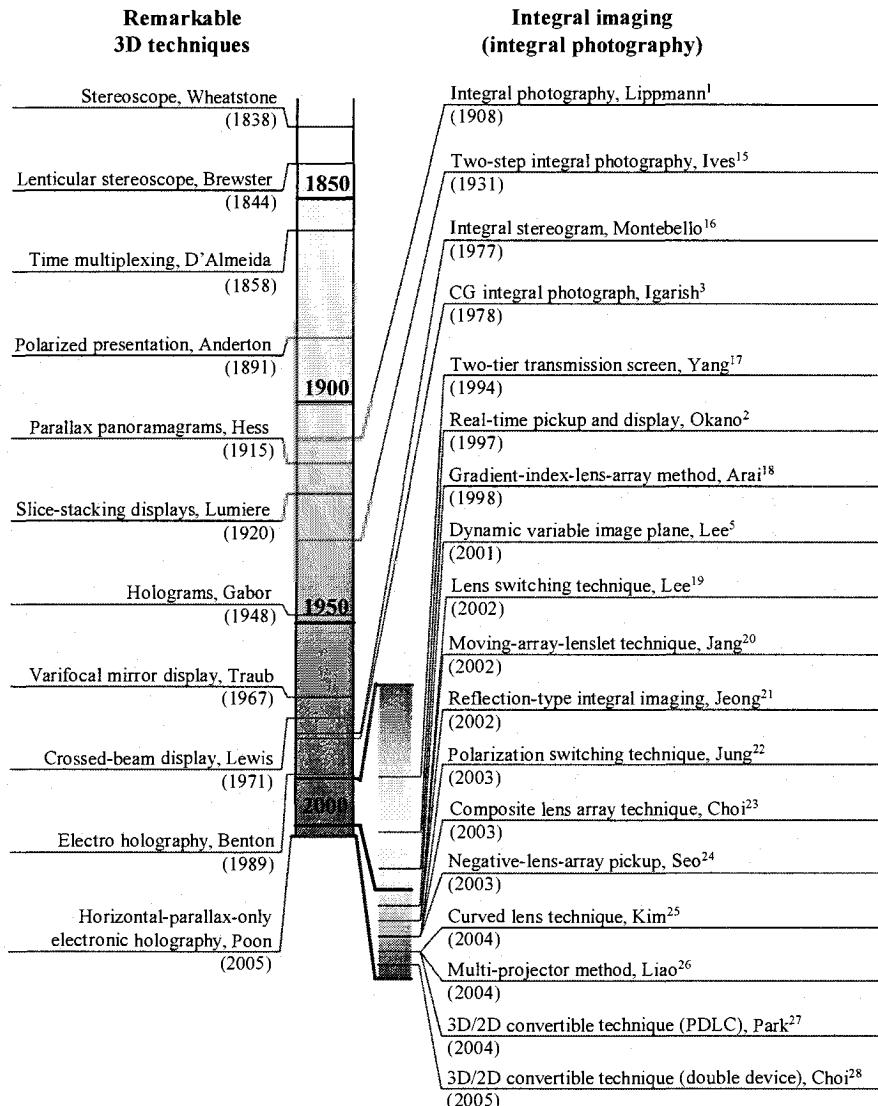


Figure 12-7. Progress of technical developments of integral imaging.

In the early stage of integral imaging, studies concerning this technique were focused on resolving the pseudoscopic problem, a situation in which the depth sequence of the reconstructed image is reversed, and on realizing a

motion picture. Since these problems have been largely resolved by previous studies, recent studies have concentrated on enhancing the viewing characteristics of an integral imaging system, such as the widening of the viewing angle and the increase in image depth. Figure 12-7 shows the process of some of the important technical developments related to 3D display and integral imaging.^{1-3,5,15-28}

3.1 Pseudoscopic image problem

The pseudoscopic problem occurs because the directions of the pickup and the display are opposite. Figure 12-8 shows the origin of the pseudoscopic problem. The interval between the elemental images for the object is inversely proportional to the distance between the object and the lens array and the location of the integrated image. As shown in Fig. 12-8(a), because object A is located behind object B, interval a for object A is shorter than interval b for object B, and the integrated image A' is located in front of the integrated image B' . Therefore, in this case, the integrated image is observed with a reversed feel of depth. The pseudoscopic problem can be resolved by obtaining elemental images from the pseudoscopic image, which can be performed using another pair of lens arrays. In this case, the elemental images in which intervals are converted are obtained, as shown in Fig. 12-8(c). Another method for solving the pseudoscopic problem is the change in the display mode. The display mode of the integrated image can be changed by providing elemental images that are rotated from the original elemental images by 180° with respect to the center of each elemental image region. In the virtual mode, the integrated image in which elemental image interval is longer, is located farther from the observer. Therefore, the depth sequence of the image is not changed in the virtual mode. In other words, elemental images which raise a pseudoscopic problem in the real mode can be integrated as an orthoscopic image in the virtual mode after real-virtual conversion. Figure 12-8(d) shows such elemental images, after conversion.

Two-step integral photography was proposed by Ives to solve the pseudoscopic problem.¹⁵ Ives concluded that the re-recording step via a second integral photograph in the convergent wave would correct the spatial sense because the pseudoscopic copy of a pseudoscopic image is orthoscopic. Consequently, the method is a meaningful solution for obtaining an orthoscopic virtual image of the original scene. However, this two-stage process inevitably leads to image degradation because of diffraction, light loss due to the multiple steps involved, and the turbidity of the emulsion. A method using two-tier transmission screen is a solution for the pseudoscopic problem, the scheme for which is similar to two-step integral photography.¹⁷

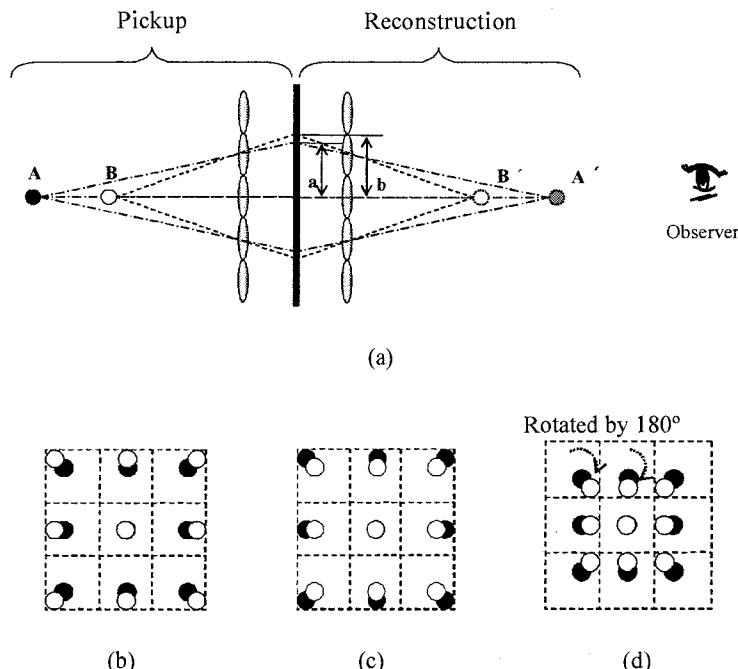


Figure 12-8. Origin of the pseudoscopic problem. (a) Scheme of pseudoscopic problems (b) elemental images of a pseudoscopic image in the real mode (c) elemental images of an orthoscopic image in the real mode (d) elemental images of an orthoscopic image in the virtual mode.

Thereafter, several methods such as a method using a gradient-index lens array and a negative lens array in the pickup process have been proposed to resolve the pseudoscopic problem. The method in which a gradient-index lens array is used is a system using a special lens array of which the elemental lens is a gradient-index lens.¹⁸ The negative lens array is a lens array in which the elemental lens is a concave lens and a charge-coupled device (CCD) camera with a capturing lens system is used in the pickup process using the negative lens array.²⁴ In these methods, each elemental image is shifted centrosymmetrically (i.e., rotated by 180°) about the center of the elemental image. As mentioned above, the display mode of integrated image is also changed from the real mode to the virtual mode as the result of the pseudoscopic-orthoscopic conversion.

3.2 Viewing angle or depth enhancement

Attempts to enhance viewing limitations such as the image depth and the viewing angle have been very active in recent years.

The lens switching method uses a movable mask¹⁹ or a movable non-uniform lens array²³ to enhance the viewing angle or image depth, respectively. When the lens array is movable in the longitudinal direction (lens axis direction), the system can represent integrated images located in both real and virtual modes with time multiplexing.⁵ A moving-array-lenslet technique has also been proposed in which two-step integral imaging and a lens switching method are combined.²⁰ In this method, a pair of lens arrays is moved in the transverse direction to produce clearer elemental images which are degraded in the two-step integral imaging process. In these movable lens array methods, the systems must include mechanical moving parts, which is a drawback for use in a practical system.

The method in which polarization switching is used expands the effective area of an elemental image to increase the viewing area without the need for any mechanical moving part.²² A system using the polarization switching is a type of time-multiplexing system using two orthogonal polarization states to produce an afterimage effect with an enhanced viewing angle. Figure 12-9 shows a schematic diagram of the proposed method. Two different polarization states whose directions are orthogonal relative to each other are used. The polarizing sheet consists of vertical polarizer arrays whose interval is the same as the width of the elemental lens. The directions of polarization of the neighbor arrays are orthogonal as shown in Fig. 12-9. The elemental lenses are opened or closed in turn (time multiplexing) by modulating the polarization states of the elemental images into two states by the use of a polarization shutter. This polarization multiplexing method can be extended to a system in which both horizontal and vertical viewing angles are improved.²⁹

There are other methods for expanding the effective area of an elemental image, in which novel components such as a curved lens array and an embossed screen are used.^{25,30} In these methods, the systems use a projection type display device which consists of a projector and a diffuser screen. Figure 12-10 shows the scheme for a system using a curved lens array and curved screen.³¹ Multiple projectors can be used for a fully curved system. Instead of the projection system, a flexible display device may be adopted in the future.

There are some methods in which multiple display devices or multiple mirrors are used.³² Figure 12-11 shows an enhanced depth system in which two display devices are used, one is emissive and the other one transmissive, used as a spatial light modulator (SLM).³³ Two CDPs are generated with this method and the occlusion problem can be solved by time multiplexing.

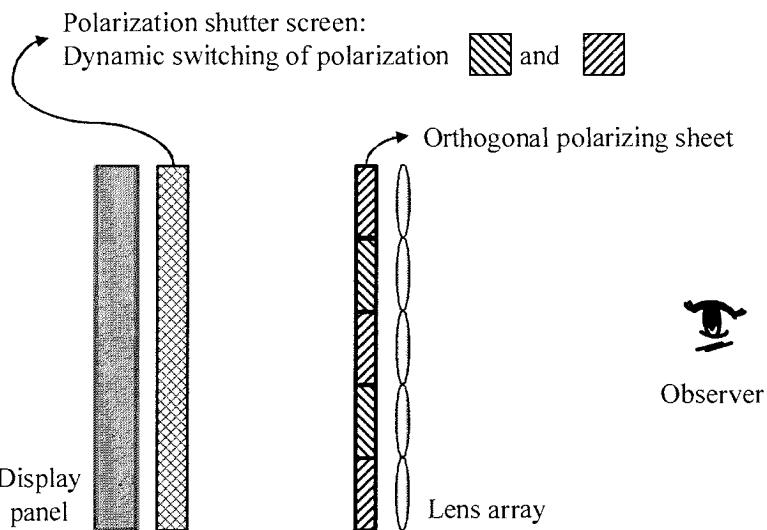


Figure 12-9. Configuration of a viewing-angle-enhanced scheme using polarization multiplexing.

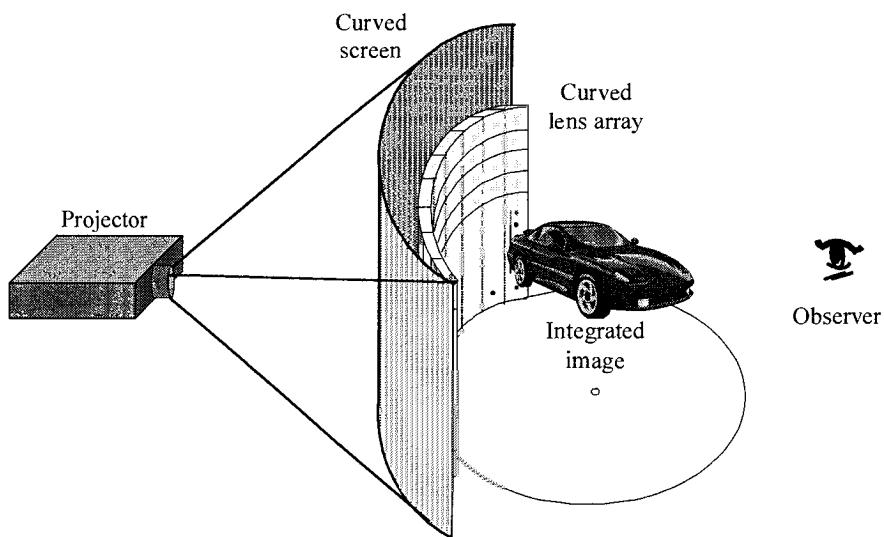


Figure 12-10. Viewing angle enhancement using a curved lens array and a curved screen.

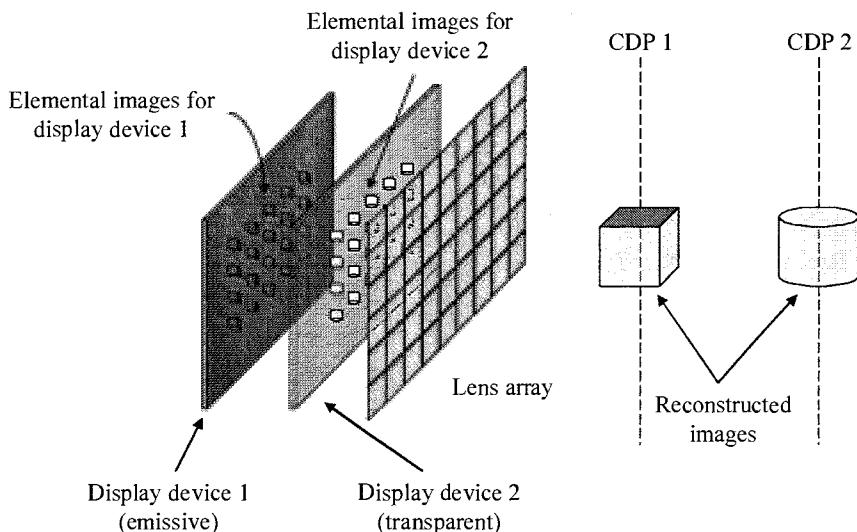


Figure 12-11. Integral imaging in which multiple display devices are used to generate multiple central depth planes.

A combination of floating display and integral imaging was recently proposed, as shown in Fig. 12-12, to generate a much improved feel of depth of 3D objects.³⁴

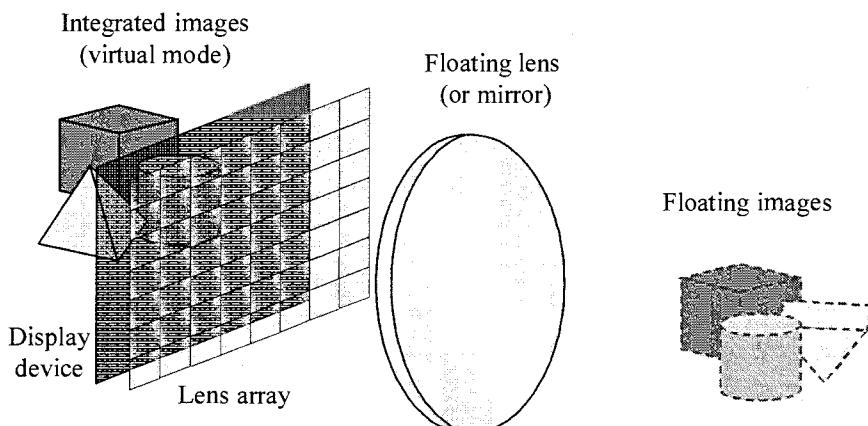


Figure 12-12. Combination of floating display and integral imaging.

3.3 3D/2D convertible integral imaging

3D/2D convertibility is crucial in realizing widely-used 3D display systems such as 3D televisions. A 2D image can be displayed with the conventional integral imaging by integrating a planar image on an integration plane. However, the resolution and viewing angle of this 2D image is inferior to the 2D image of a conventional television. Two methods have been proposed for 3D/2D convertible integral imaging systems.

A schematic diagram of one of the methods is shown in Fig. 12-13.²⁷ In the 3D mode, a polymer-dispersed liquid crystal (PDLC) located just behind the lens array is electrically controlled so as to be transparent. The collimated illumination light is then focused by the lens array to form a point light source array at the focal plane of the lens array. The SLM, which is located in front of the lens array at a distance of twice the focal length of the lens array, modulates the intensity of the light rays properly and 3D image is displayed in an increased depth range by principles that are explained below. In the 2D mode, the PDLC is set to be diffuse. The illumination light is scattered by the PDLC and relayed by the lens array to illuminate the SLM with its diffusing field. Since each pixel on the SLM is effectively illuminated in all directions, the observers see a 2D image on the SLM with full resolution and viewing angle of the SLM. Although the PDLC can be located at other positions including the front face of the SLM or in the gap between the lens array and the SLM, locating it on the back plane of the lens array is preferred, considering the effectiveness of diffusing and readiness of implementation.

The main feature of this 3D/2D convertible integral imaging is that the lens array is located behind the SLM, unlike conventional integral imaging where the lens array is placed in front of the SLM (display panel). This feature increases the expressible depth range of the system and enables the introduction of the PDLC to accomplish 3D/2D conversion.

Due to the inverted order of the lens array and the SLM, the principle of displaying a 3D image is slightly different from conventional integral imaging. In conventional integral imaging, each elemental image is imaged at the CDP by the corresponding elemental lens and the 3D image is formed in the vicinity of the CDP where the light rays from the elemental images converge, as shown in Fig. 12-5(b). Since the formation of a 3D image is based on the imaging property of the elemental lenses, as discussed previously, the 3D image cannot be located at a large distance from the CDP of the elemental lenses due to the large off-focusing and the 3D image resolution degrades as the 3D image depth increases due to the large magnification involved. Therefore the expressible depth range is significantly limited in conventional integral imaging. Moreover the gap

between the display panel and the lens array should be determined by the Gauss lens law according to the desired central depth, which prevents the conventional integral imaging system from simultaneously displaying real and virtual 3D images with good resolution.

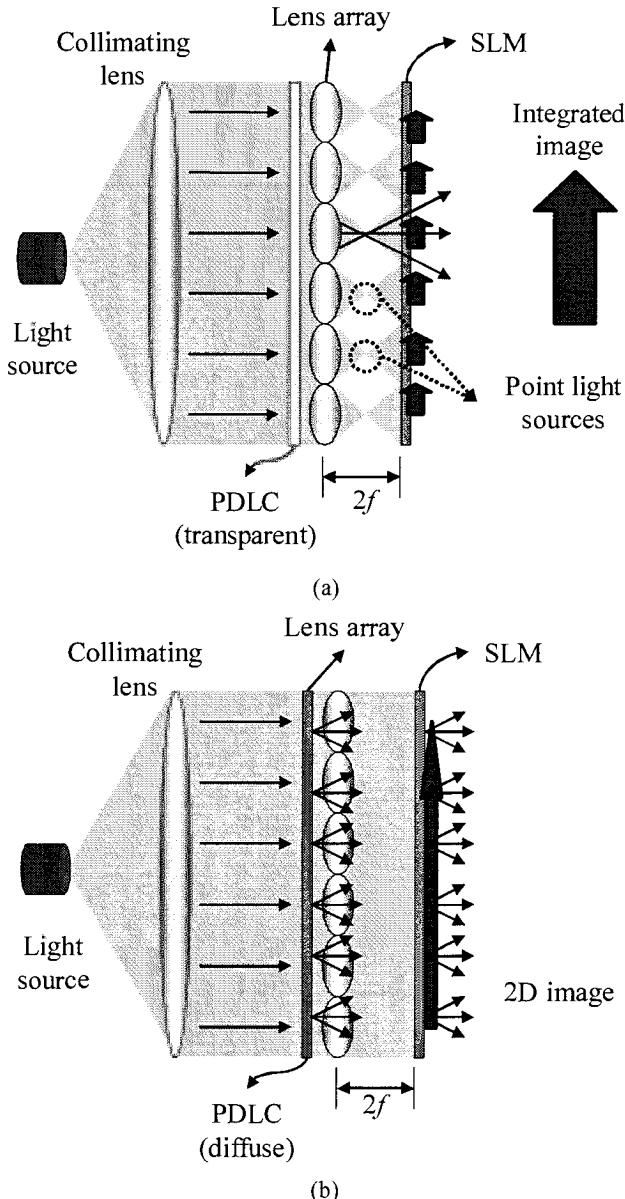
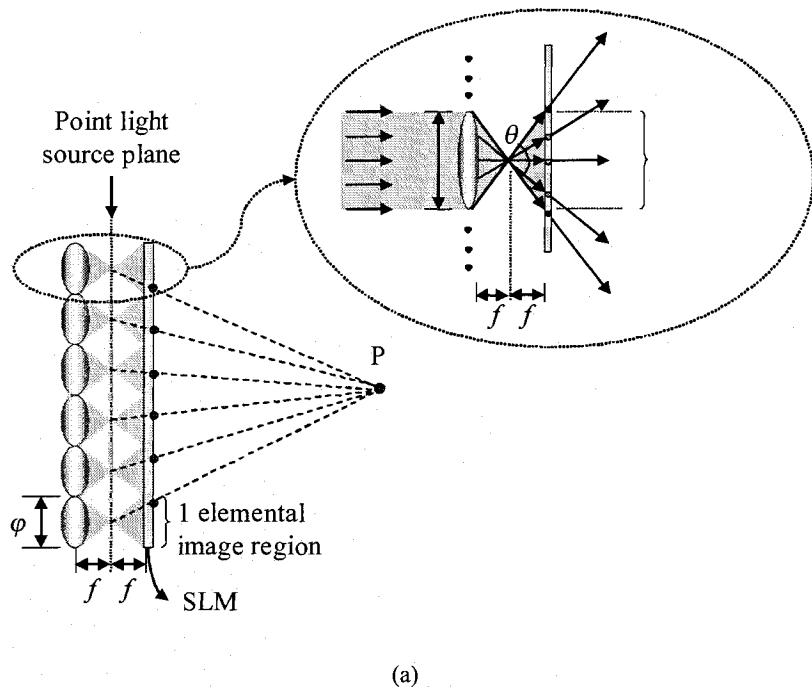


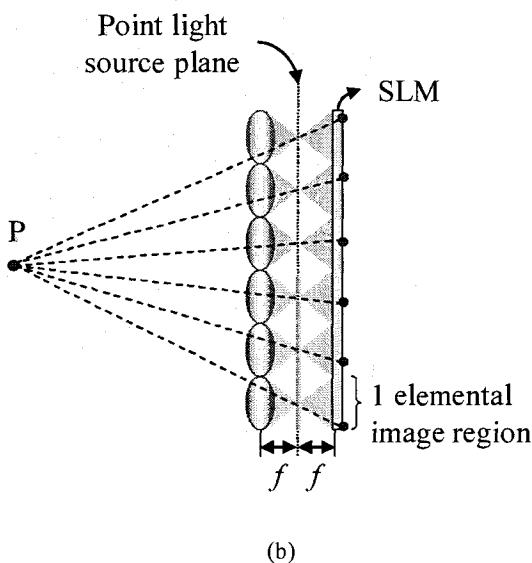
Figure 12-13. Concept of 3D/2D convertible integral imaging using PDLC (a) 3D mode, (b) 2D mode.

The 3D/2D convertible integral imaging of Fig. 12-13, however, overcomes this limitation since the formation of 3D in the proposed method is not based on the imaging of elemental lenses. Figures 12-14(a) and (b) show the principle of the 3D image formation in 3D/2D convertible integral imaging. In 3D/2D convertible integral imaging, the elemental image region corresponding to each point light source is defined as the extent of diverging rays falling on the display panel from the point light source. Since the gap between the lens array and the display panel is fixed as twice the focal length of the lens array without regarding the 3D image depth, elemental image regions do not overlap with each other and its width is the same as an elemental lens width φ . In order to display a real 3D image point at P in Fig. 12-14(a) or a virtual 3D image point at P in Fig. 12-14(b), the elemental image points of P are displayed where the lines joining P and point light sources meet the SLM. The elemental image point that exceeds the corresponding elemental image region is discarded, to avoid image-flipping effects, as conventional integral imaging. The light rays from the point light sources passing through the corresponding elemental image points then converge at the 3D image point P, and hence an observer can see the 3D image point P. Since 3D image formation is not based on the imaging of elemental images, there is no restriction on the object depth and real and virtual 3D images can be simultaneously displayed in this type of 3D/2D convertible integral imaging. An experimental demonstration of this method is shown in Fig. 12-15, detailed information on which can be found in Ref. 27.

Another 3D/2D convertible integral imaging system that uses two display devices is shown in Fig. 12-16.²⁸ In the 3D mode, the rear display panel shows elemental images, which are integrated through a lens array. In the 2D mode, the rear display panel is in the white mode to function as a backlight for the front transmission-type display device, which functions as an SLM to show 2D images. In this case, the lens array does not do anything because it is located between the uncollimated white light source (rear display panel) and the front panel.

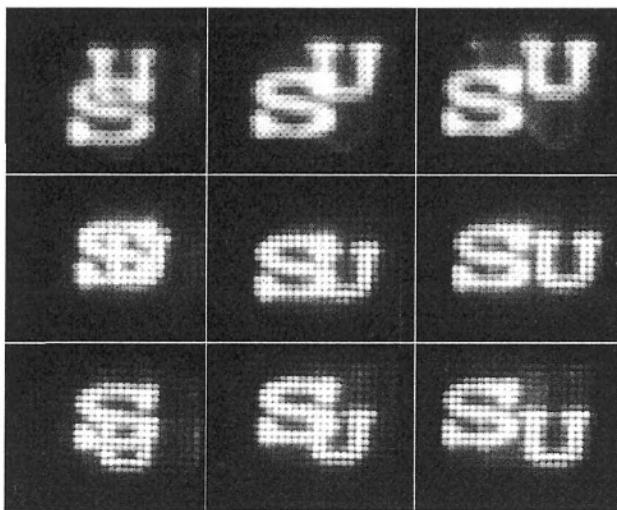


(a)



(b)

Figure 12-14. Principle of 3D image formation for (a) a real field and (b) a virtual field in 3D/2D convertible integral imaging using a point light source array and PDLC.

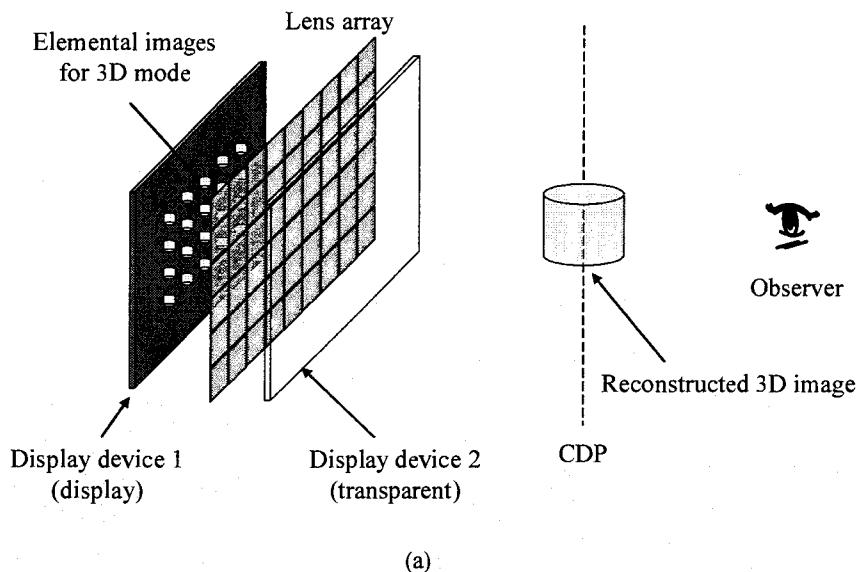


(a)

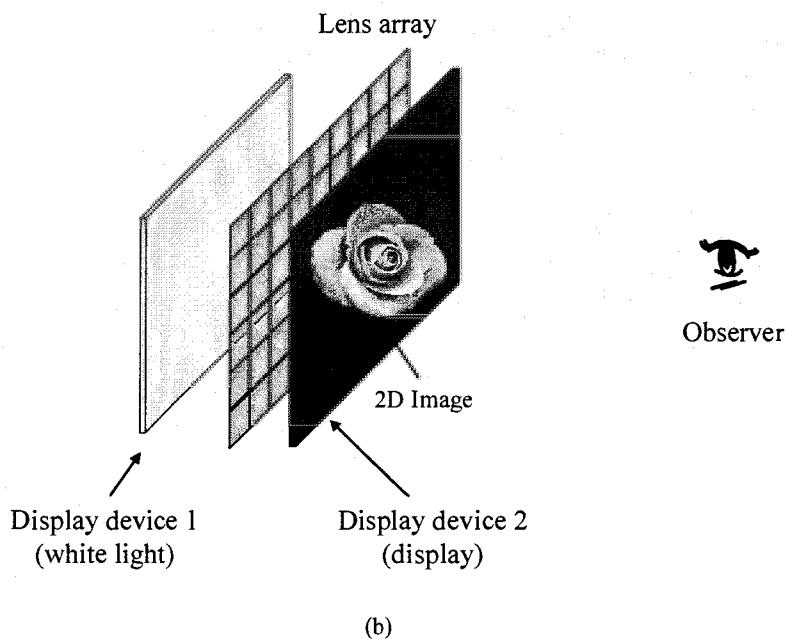


(b)

Figure 12-15. Experimental results: (a) 3D image observed from different viewing directions with the transparent PDLC state, and (b) 2D image with the diffuse PDLC state.



(a)



(b)

Figure 12-16. 3D/2D convertible integral imaging system using two display devices: (a) 3D mode (b) 2D mode.

4. DEPTH EXTRACTION

In the pickup process of integral imaging, elemental images are actually perspectives of the object at specific viewing directions which are determined by the relative positions of the corresponding elemental lenses and the object. Hence 3D information concerning the object is stored as a form of an elemental image array and it is possible to extract stored 3D information from it.³⁵

4.1 Geometry of the pickup system of the integral imaging

Figure 12-17 shows the geometry of the pickup system used in integral imaging.³⁶ Although we consider a 1D lens array here for simplicity, the extension to the 2D lens array is straightforward. In Fig. 12-17, a large convex lens (depth control lens) is also used in addition to the 2D lens array, and we also consider ray refraction by the camera lens.

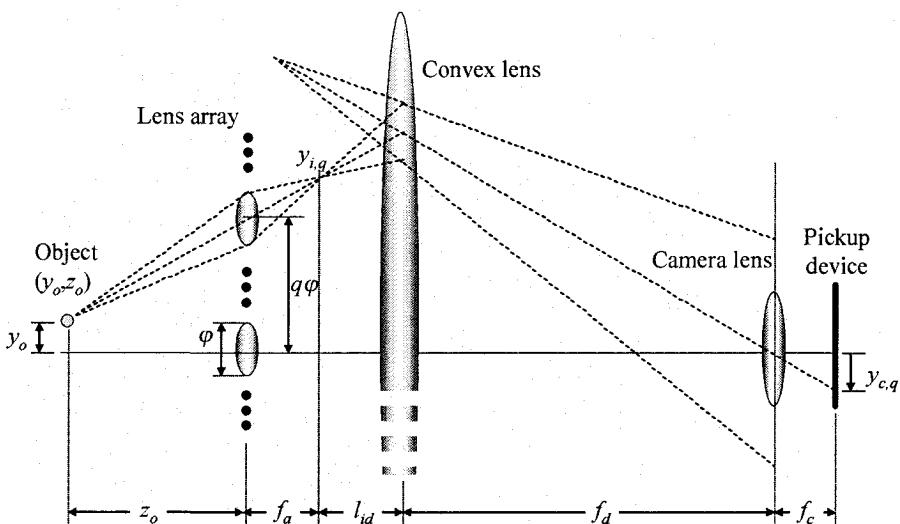


Figure 12-17. Geometry of the pickup process of the integral imaging.

The object point (y_o, z_o) , imaged by the q -th elemental lens at $y_{i,q}$ in the elemental image plane, passes through the lens system and is captured at $y_{c,q}$.

in the CCD plane. The position $y_{c,q}$ of the q -th elemental image of the object point (y_o, z_o) is given by

$$y_{c,q} = -\frac{f_a f_c (q\varphi - y_o)}{f_d (z_o - f_a)} - \frac{f_c q\varphi}{f_d}, \quad (12-8)$$

where f_a is the focal length of the elemental lens, f_c the focal length of the camera lens, f_d the focal length of the depth control lens, and φ the elemental lens pitch. If we define the center of the q -th elemental image $y_{o,q}$ as

$$y_{o,q} = -\frac{f_c q\varphi}{f_d}, \quad (12-9)$$

the position $y_{r,q}$ of the q -th elemental image of the object point with respect to the center of the q -th elemental image becomes

$$y_{r,q} = -\frac{f_a f_c (q\varphi - y_o)}{f_d (z_o - f_a)}. \quad (12-10)$$

The disparity between the q_1 -th and q_2 -th elemental images is given by

$$\Delta y_{q1,q2} = y_{r,q2} - y_{r,q1} = \frac{f_a f_c \varphi (q_2 - q_1)}{f_d (z_o - f_a)} = (q_2 - q_1) d, \quad (12-11)$$

where d denotes the fundamental disparity. Since d is a function of the object depth z_o , depth information on the object can be extracted by detecting the disparity. Once we know the depth z_o , its transverse information y_o can also be extracted using Eq. (12-10). Although, in this section, we consider a pickup configuration with the depth control lens that makes the pickup directions of the elemental lenses be aligned in parallel, the pickup configuration without a depth control lens can also be analyzed in the same manner and the process yields a similar result with slightly different disparity d .

4.2 Extraction of the 3D information from elemental images

Since the transverse information of the object is linked to depth information and depth information is inscribed in the disparity, precise disparity detection is a key step for the acquisition of 3D information on an

object. Techniques for detecting the disparity between two or more perspectives by searching for the corresponding points in every perspective are generally termed stereo matching and have been developed. Among various stereo matching techniques, the multi-baseline technique can be directly applied to elemental images since more than two elemental images can be used cooperatively for this purpose. Figure 12-18 shows an example of the detected disparity map by applying the multi-baseline technique to elemental images.

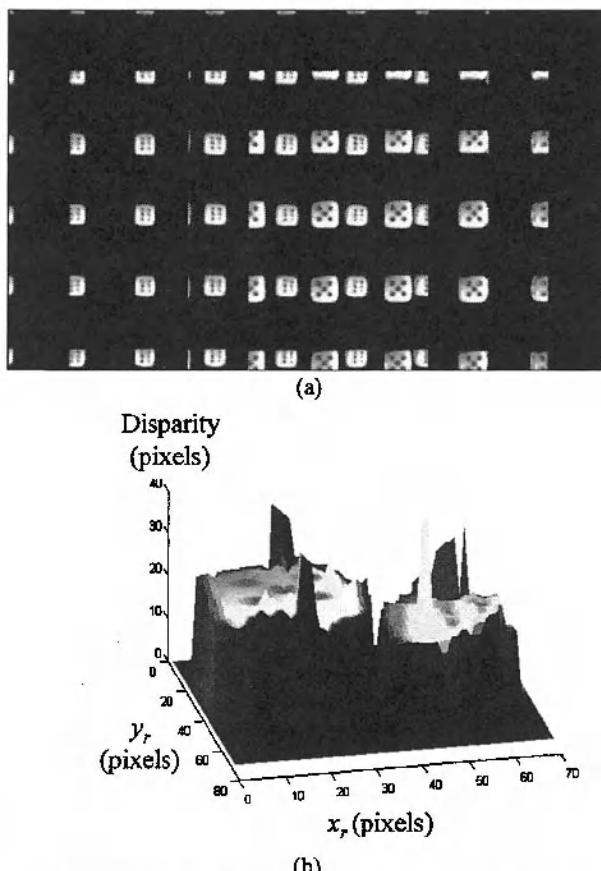


Figure 12-18. Example of disparity detection using the multi-baseline technique: (a) Elemental images, (b) detected disparity map.

4.3 Extraction of 3D information from 1D sub-images

The main difference between the usual stereo matching configuration and the integral imaging pickup system is that stereo matching utilizes several cameras to capture the perspective of the object, while the integral imaging

pickup system incorporates a lens array and single camera. Therefore the integral imaging pickup system provides much more perspectives (as many as the number of the elemental lenses used) than the usual stereo matching system. Since all the perspectives are captured by one camera, however, the resolution of each perspective is low. Generally, the accuracy of disparity detection depends on the number of input images and the resolution of each. Hence two counteracting characteristics of the integral imaging pickup system – the large number of perspectives, the low resolution of each perspective – determine the reliability of disparity detection and 3D information extracted for the object. To fully utilize the large number of elemental images in integral imaging, a 1D sub-image method has been developed, where elemental images are transformed to 1D sub-images before disparity detection.³⁷ The sub-image is a collection of pixels at the same position in all elemental images. Figure 12-19 illustrates the generation of 1D sub-images.

Pixel columns that are on the same horizontal position in each elemental image are assembled to form a 1D sub-image along the horizontal direction. For example, the first pixel columns of the elemental images on the fourth row of the array of elemental images are assembled to form the [1,4]-th 1D sub-image, if we designate the 1D sub-image whose position is i -th from the left and j -th from the top as $[i, j]$. In general, the i -th pixel columns of elemental images of the j -th row of the array of elemental images are assembled to form the $[i, j]$ -th 1D sub-image. Therefore if the lens array consists of N_x (horizontal) by N_y (vertical) elemental lenses and thus there are N_x by N_y elemental images and each elemental image consists of N_{xp} by N_{yp} pixels, then modification of the elemental images yields N_{xp} by N_y 1D sub-images, each of which consists of N_x by N_{yp} pixels. From the viewpoint of pixels that comprise the elemental images, after modification, only their horizontal positions are changed while their vertical positions remain unchanged.

The main effect of this modification is the inversed dependency of disparity on object depth. Suppose that an object point (x_o, y_o, z_o) in the object scene is imaged by the lens array and captured by CCD. From Eq. (12-10), the position of the elemental image point that is imaged by $[q_x, q_y]$ -th elemental lens is given in pixel units with respect to the origin of $[q_x, q_y]$ -th elemental image as follows:

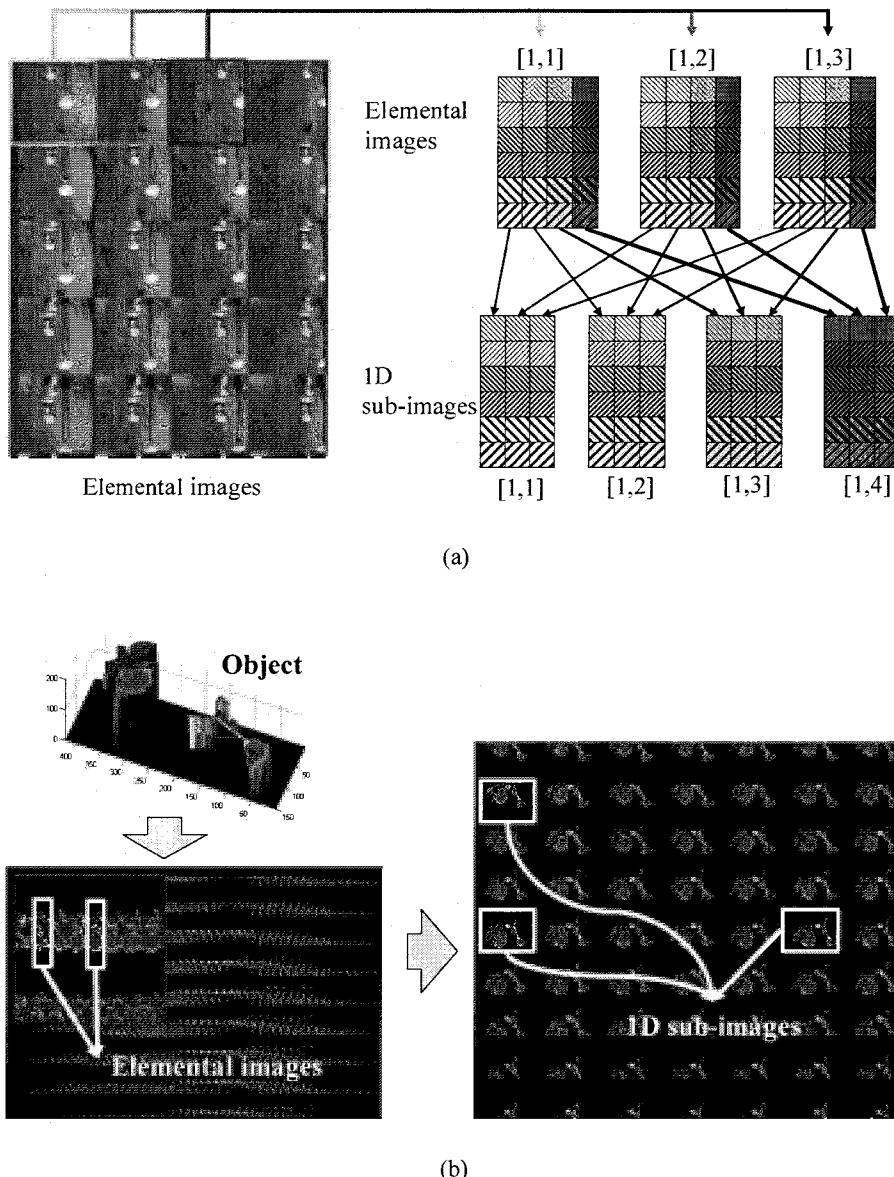


Figure 12-19. Generation of 1D sub-images. (a) Method of forming 1D sub-images, (b) example of elemental images (left) and generated 1D sub-images (right).

$$(x_{q_x}, y_{q_y}) = \left(\frac{f_c f_a}{f_d s z_o} (q_x \varphi_x - x_o), \frac{f_c f_a}{f_d s z_o} (q_y \varphi_y - y_o) \right), \quad (12-12)$$

where s is the pixel pitch of the CCD. By generating a 1D sub-image, the elemental image point (x_{q_x}, y_{q_y}) in $[q_x, q_y]$ -th elemental image is transformed to (q_x, y_{q_y}) in $[x_{q_x}, q_y]$ -th 1D sub-image. Therefore, considering Eq. (12-12), the coordinates (u_{px}, v_{py}) in the $[p_x, p_y]$ -th 1D sub-image can be written as:

$$(u_{px}, v_{py}) = \left(\frac{1}{\varphi_x} \left(\frac{sz_o}{m} p_x + x_o \right), \frac{m}{sz_o} (p_y \varphi_y - y_o) \right), \quad (12-13)$$

where m is $f_d f_d / f_d$. From Eq. (12-13), the horizontal and vertical disparities between $[p_{x1}, p_{y1}]$ -th and $[p_{x2}, p_{y2}]$ -th 1D sub-images are given by:

$$\Delta u_{px1, px2} = \frac{sz_o}{\varphi_x m} (p_{x2} - p_{x1}), \quad (12-14)$$

$$\Delta v_{py1, py2} = \frac{\varphi_y m}{sz_o} (p_{y2} - p_{y1}). \quad (12-15)$$

Equations (12-14) and (12-15) reveal that the horizontal disparity is proportional to the object distance z_o , while the vertical disparity is inversely proportional to the object distance z_o . This inversed dependency brings about two major advantages – elimination of ambiguities in disparity detection and a constant quantization error in the disparity-to-depth conversion.

When the object scene includes a periodic texture, the usual stereo matching techniques frequently fail to discriminate the correct disparity since the evaluation function, like the sum of the squared difference (SSD), has multiple minima rather than one global minimum. Although the multi-baseline technique shows good performance in eliminating ambiguities, its performance is limited if it is directly applied to elemental images obtained by a circular or square lens array because the baseline of the lens array is uniform. In the 1D sub-image method, however, the inversed dependency of the disparity distributes the position of the minima of the evaluation function differently along the horizontal and vertical directions so that the evaluation function has only one minimum at the correct position. Consequently ambiguity is eliminated completely regardless of the period of the texture. Figure 12-20 shows an example of the detected disparity map of an object with a periodic pattern. We can see that the disparity map detected by both of the vertical and horizontal disparities in the 1D sub-image method reflects the object exactly while disparity map obtained by only vertical disparity, which is equivalent to the conventional multi-baseline technique, suffers

from many errors due to the ambiguities.³⁷ This is confirmed by the evaluation functions at a certain image point, as shown in Fig. 12-20(c). Although the horizontal (dotted line) and vertical (solid line) evaluation functions have their respective multiple minima, their sum ('x' marked line) has one well defined global minimum at the correct position.

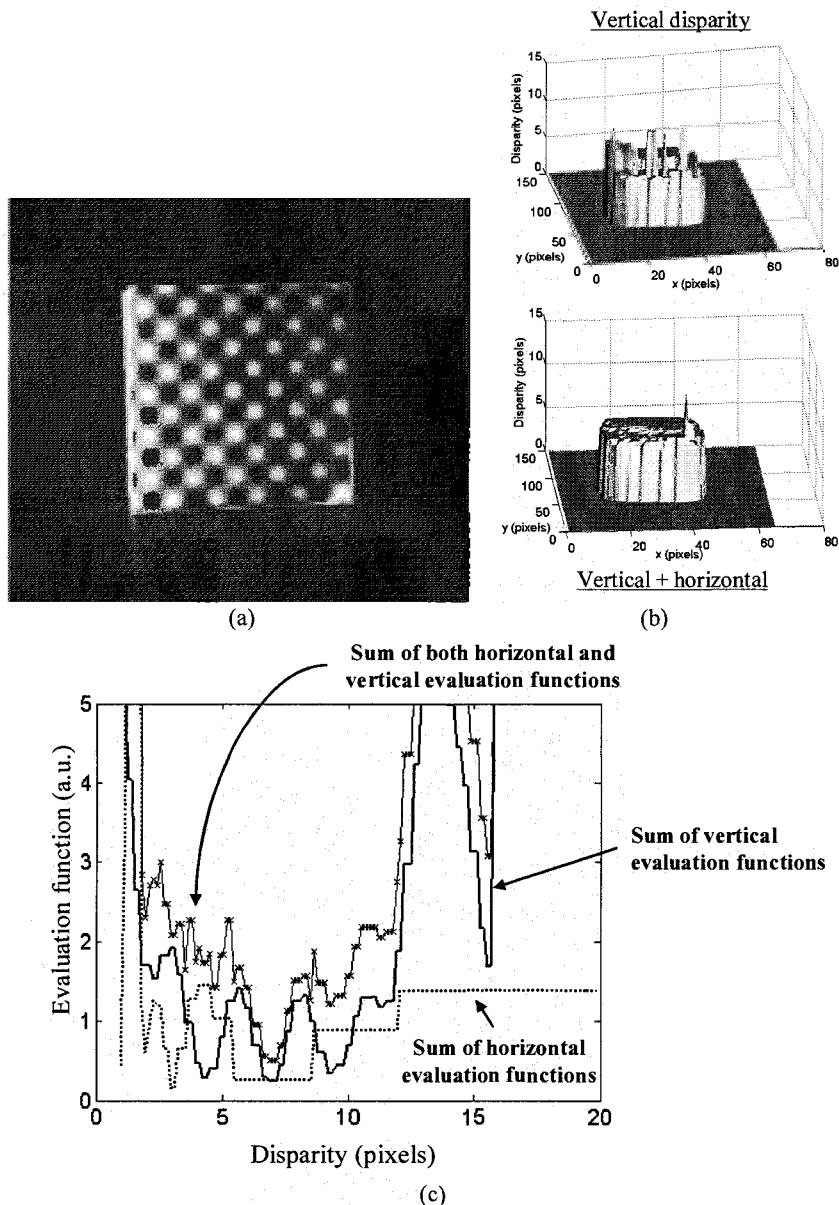


Figure 12-20. Elimination of ambiguities in the 1D sub-image method. (a) Flat object, (b) detected disparity map, (c) example of evaluation functions.

Another advantage of the inversed dependency of disparity is the constant quantization error. The quantization error is an error in disparity-to-depth conversion due to the discrete pixel structure of the CCD. In conventional stereo matching techniques, the quantization error increases as the object depth increases and becomes infinity, thus imposing a fundamental limit on the object depth when the object depth is too large so that the disparity is smaller than the half of one pixel. Although the fundamental limit for object depth cannot be alleviated in the 1D sub-image method, the quantization depth error of the 1D sub-image method can be maintained constant.³⁷ In the 1D sub-image method, the conditional expectation value of the quantization error from the vertical disparity and the expectation value from the horizontal disparity are given, respectively, by

$$E(\delta z_y \| z) = \frac{sz^2}{3m\varphi_y |p_{y1} - p_{y2}|}, \quad (12-16)$$

$$E(\delta z_x) = \frac{m\varphi_x}{3s |p_{x1} - p_{x2}|}. \quad (12-17)$$

From Eqs. (12-16) and (12-17), it can be seen that the expectation value of the quantization error from the vertical disparity is proportional to the square of the object depth in conventional stereo matching techniques, while that from the horizontal disparity is constant regardless of the object depth. Therefore it is advantageous to use vertical disparity when the object depth is small and to use horizontal disparity when the object depth is large.

In Fig. 12-21, the step height represents the quantization error. Thus, the step height from the vertical disparity is smaller than that from the horizontal disparity when the object depth is small and vice versa when the object depth is large. Therefore the quantization error in the 1D sub-image method can be reduced by selectively using the disparities between the horizontal and vertical directions.

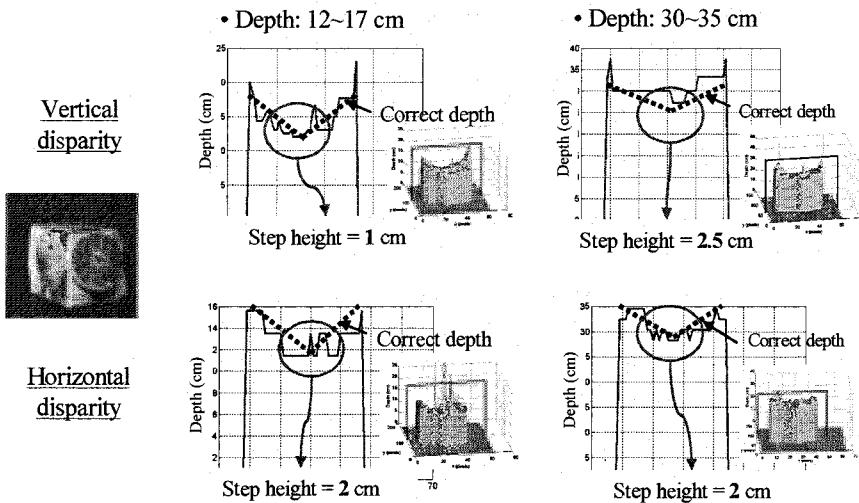


Figure 12-21. Quantization error in the 1D sub-image method.

5. 3D CORRELATOR

A 3D optical correlator provides a direct way to recognize and locate 3D objects distributed in a 3D space. Several schemes have been proposed to realize a 3D optical correlator using holography^{38,39} and 3D optical Fourier transform.⁴⁰⁻⁴² A schematic diagram of a 3D correlator based on integral imaging is shown in Fig. 12-22.^{43,44} The reference and signal 3D objects are imaged by lens arrays to form elemental images. The elemental image itself is an ordinary picture of the object space. The captured perspective of the object and its position in the elemental image depend on the position of the corresponding elemental lens relative to the position of the object. By imaging through the lens array, these elemental images are produced and captured by CCDs. The captured elemental image arrays for reference and signal objects are digitally transformed into 2D sub-image arrays of each. Some 2D sub-images of the reference object are then selected and correlated with every 2D sub-image of the signal object by means of a conventional joint transform correlator (JTC) scheme, thus yielding information on the 3D shift and out-of-plane rotation of the signal object with respect to the reference object.

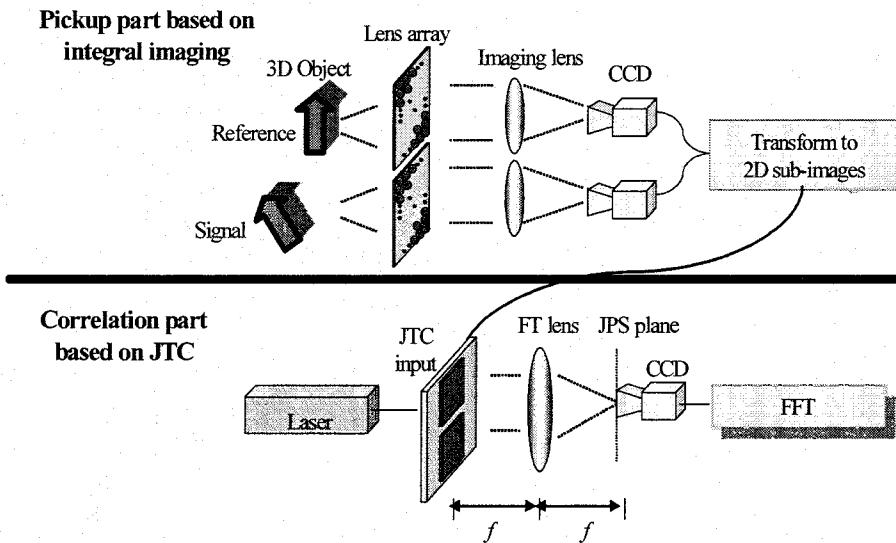


Figure 12-22. Schematic diagram of a 3D correlator based on integral imaging. (JTC: joint transform correlator, JPS: joint power spectrum, FT: Fourier transform, FFT: fast Fourier transform)

The most unique part is that a correlation operation is performed on the 2D sub-images instead of elemental images.⁴⁴ The 2D sub-image is an extension of the 1D sub-image discussed in the previous section, and is a collection of pixels at the same position in all of the elemental images, or equivalently at the same relative position with respect to the optic axis of the corresponding elemental lens. Figure 12-23 illustrates the generation of 2D sub-images. In Fig. 12-23(a), 5 elemental lenses are shown, generating 5 elemental images. The pixels at the same location in the elemental images are collected to form the corresponding 2D sub-image. For example, pixels at positions corresponding to a black dot in every elemental image form one 2D sub-image and pixels at white dot positions form another 2D sub-image. Since there are 5 elemental images in Fig. 12-23(a), each 2D sub-image will consist of 5 pixels. Figure 12-23(b) shows a front view. Pixels located at [1,1] in every elemental image (in Fig. 12-23(b), 4(H)×5(V) elemental images are shown) are collected to form the [1,1]-th 2D sub-image and pixels at [i,j] form the [i,j]-th 2D sub-image. Each 2D sub-image in Fig. 12-23(b) consists of 4(H)×5(V) pixels since there are 4(H)×5(V) elemental images.

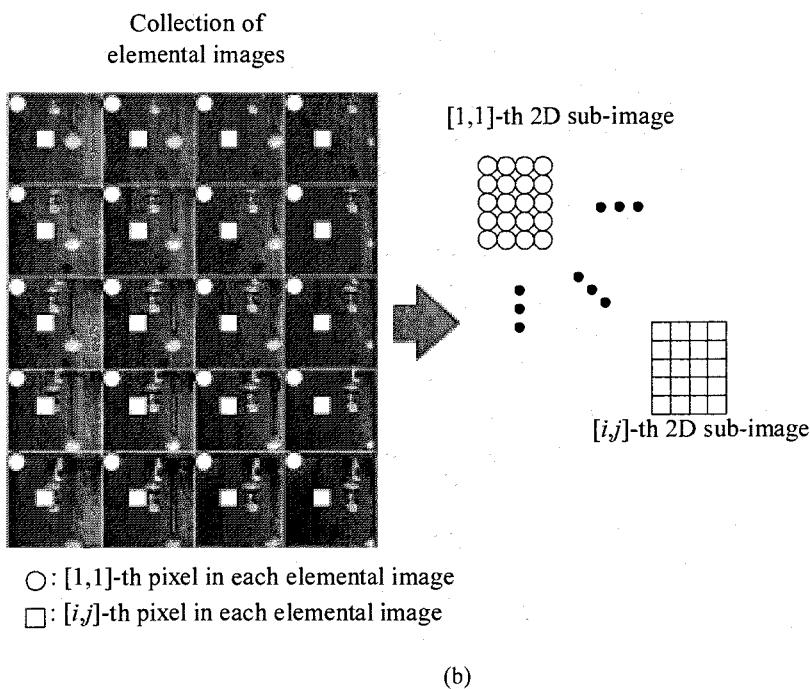
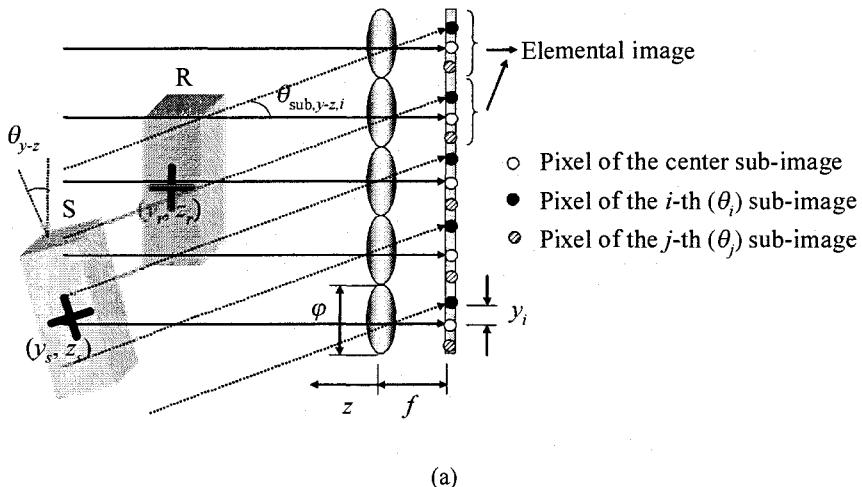


Figure 12-23. 2D sub-image: (a) geometry (b) generation.

The 2D sub-image has two useful features. One is that each 2D sub-image represents a specific angle from which the object is observed regardless of the 3D position of the object. For example, in Fig. 12-23(a), i -

th 2D sub-image (collection of black dots in Fig. 12-23(a)) contains the perspective of the object observed in an angle given by

$$\theta_{\text{sub},y-z,i} = \tan^{-1}\left(\frac{y_i}{f}\right), \quad (12-18)$$

where y_i is the position of the i -th pixel with respect to the optic axis of the corresponding elemental lens. Note that, in an ordinary imaging system, the angle of observation is determined by the relative position of the imaging lens with respect to the object. This observation-angle dependency on object position, however, is removed in the case of a 2D sub-image. Figure 12-24(a) demonstrates this point. In the case of an ordinary imaging system, the captured perspective of the object changes as the object moves from position 1 to position 2. In the 2D sub-image, however, the perspective of the object contained in each 2D sub-image is the same regardless of the object shift as shown in Fig. 12-24(a): the sub-image corresponding to black pixels observes the object from an angle of 0° and the 2D sub-image corresponding to white pixels observes the object from $\theta_{\text{sub},i}$ for both positions 1 and 2. The angle-invariance of the 2D sub-image makes it possible to select certain angle of observation deterministically regardless of the position of the object.

Another useful property of the sub-image is that the perspective size is invariant, regardless of the object depth. In an ordinary imaging system, the perspective size is inversely proportional to the object depth. Therefore if the object moves farther from the imaging lens, the object perspective in the captured image becomes smaller. In the sub-image, however, the object is captured in the form of parallel lines with the sampling period of elemental lens pitch ϕ , as shown in Fig. 12-23(a), and, thus, the size of the object perspective in the 2D sub-image is constant. When the object depth changes, only the position of the object perspective is changed in each 2D sub-image but the size itself is not. For example, suppose that an object whose transverse size covers 5 elemental lenses is imaged by the lens array shown in Fig. 12-24(b). The size of the object perspective in the 2D sub-image is determined by the number of the 2D sub-image parallel lines that intersect the object. In Fig. 12-24(b), it is the 5 pixel size for the 2D sub-image corresponding to the white dots. When the object moves longitudinally, as shown in Fig. 12-24(b), the number of parallel lines intersecting the object is still 5 for the white dots and, thus, the size of the object perspective in such 2D sub-images is not changed. Only the position of the object perspective in the 2D sub-image is changed (by 2 pixels for the 2D sub-image corresponding to the white dots in Fig. 12-24(b)). This size-invariant feature eliminates the need for any scale-invariant detection techniques such as a

Mellin transform,⁴³ even though the signal object shifts in the depth direction.

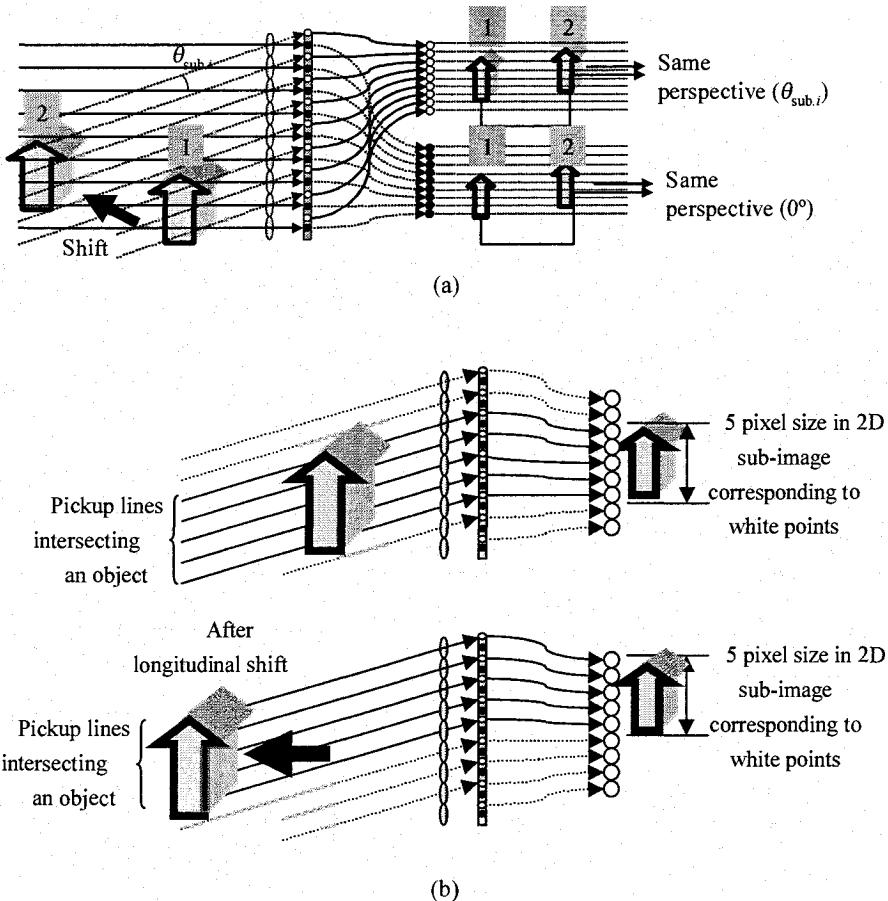


Figure 12-24. Property of 2D sub-image: (a) observing angle invariance (b) perspective size invariance.

Using these two features of 2D sub-images, i.e. size-invariance and angle-invariance, 3D shift and the out-of-plane rotation can be detected using a JTC scheme as follows. Suppose that the reference object is located at (y_r, z_r) and the signal object is located at (y_s, z_s) as shown in Fig. 12-23(a). First, let us assume that the signal object has no out-of-plane rotation for the sake of simplicity; i.e. $\theta_{y-z} = 0^\circ$ in Fig. 12-23(a). Since there is no out-of-plane rotation, the perspective of the object contained in the i -th 2D sub-image of the signal object is the same as that contained in i -th 2D sub-image of the reference object. Note that this is true irrespective of where the signal object is located with respect to the reference object due to the observation

angle invariance property of the 2D sub-image. Also note that the sizes of the perspectives in these two 2D sub-images for the reference and signal objects are the same due to the size-invariance property. The position of the perspective in the i -th 2D sub-image is given by $u_{r,i} = (1/\varphi)(y_r + z_r \tan \theta_{\text{sub},y-z,i})$ for the reference object and $u_{s,i} = (1/\varphi)(y_s + z_s \tan \theta_{\text{sub},y-z,i})$ for the signal object. The difference in their position $\Delta u_{r,i,s,i}$ can be written as

$$\Delta u_{r,i,s,i} = u_{r,i} - u_{s,i} = \frac{y_r - y_s + (z_r - z_s) \tan \theta_{\text{sub},y-z,i}}{\varphi}. \quad (12-19)$$

Since the i -th 2D sub-images of the reference and signal objects contain the same perspective of an object with the same size, the difference in position $\Delta u_{r,i,s,i}$ can be detected by correlating the i -th 2D sub-images of the reference and the signal objects using JTC. In Eq. (12-19), only y_s and z_s are unknowns and thus, the 3D shift in the signal object can be found by applying two correlation operations with different i 's.

When there is an out-of-plane rotation θ_{y-z} in the signal object, the 3D shift cannot be found by correlating the reference object sub-image with the signal object 2D sub-image of the same index because they will, in general, contain different perspectives of the object. In this case, the 2D sub-image pair that contains the same perspective of the object should be found first, in other words, the out-of-plane rotation must be detected first. The 3D shift can then be found by considering the out-of-plane rotation. The out-of-plane rotation angle θ_{y-z} of the signal object is detected by successively correlating one arbitrarily chosen 2D sub-image for the reference object with every 2D sub-image of the signal object. Among them, the 2D sub-image pair yielding the strongest correlation peak will satisfy $\theta_{\text{sub},y-z,i} - \theta_{\text{sub},y-z,j} = \theta_{y-z}$ where $\theta_{\text{sub},y-z,i}$ is the angle of observation of the i -th 2D sub-image of the reference object and $\theta_{\text{sub},y-z,j}$ is that of the j -th 2D sub-image of the signal object, since they have the same perspective of the object. Therefore, by finding the 2D sub-image pair that produces the strongest correlation peak, the out-of-plane rotation angle θ_{y-z} can be detected. After θ_{y-z} is detected, the 3D position of the signal object can also be detected by correlating two more 2D sub-image pairs as a no out-of-plane rotation case. In this case, however, we correlate the i -th 2D sub-image of the reference object with the j -th 2D sub-image of the signal object where $\theta_{\text{sub},y-z,i} - \theta_{\text{sub},y-z,j} = \theta_{y-z}$, since they have the same perspective. The difference in position $\Delta u_{r,i,s,j}$ of the object perspectives in i -th reference 2D sub-image and j -th signal 2D sub-image is given by

$$\begin{aligned}\Delta u_{r,i,s,j} &= \frac{y_r - y_s + z_r \tan \theta_{\text{sub},y-z,i} - z_s \tan \theta_{\text{sub},y-z,j}}{\varphi} \\ &= \frac{y_r - y_s + z_r \tan \theta_{\text{sub},y-z,i} - z_s \tan(\theta_{\text{sub},y-z,j} + \theta_{y-z})}{\varphi}.\end{aligned}\quad (12-20)$$

Therefore the 3D position of the signal object can be found through Eq. (12-20) by selecting two 2D sub-image pairs corresponding to $\theta_{\text{sub},y-z,i}$ for the reference object and $\theta_{\text{sub},y-z,i} + \theta_{y-z}$ for the signal object and measuring the positions of their correlation peaks.

In the detection of out-of-plane rotation, the minimum resolvable angle is determined by the difference between the observation angles of neighboring 2D sub-images. Specifically, the angular resolution $\Delta\theta$ is given by $\Delta\theta = \theta_{\text{sub},i+1} - \theta_{\text{sub},i}$. Since the observation angle of the i -th 2D sub-image $\theta_{\text{sub},i}$ is given by Eq. (12-18), the angular resolution $\Delta\theta$ becomes

$$\Delta\theta = \tan^{-1}\left(\frac{y_{i+1}}{f}\right) - \tan^{-1}\left(\frac{y_i}{f}\right) \approx \frac{y_{i+1} - y_i}{f} = \frac{s}{f}, \quad (12-21)$$

where s is the pixel pitch at the image plane of the lens array. The detectable angular range Ω is determined by the range of the observation angles of the 2D sub-images (range of θ_{sub}). Since y_i is restricted by $-\varphi/2 \leq y_i \leq \varphi/2$, the angular range Ω becomes

$$\Omega = 2 \tan^{-1}\left(\frac{\varphi}{2f}\right) \approx \frac{\varphi}{f}, \quad (12-22)$$

where Eq. (12-18) is used.

Figure 12-25 shows an example of the object, captured elemental images, and 2D sub-images that are generated. The lens array used here consists of 50×50 rectangular elemental lenses with a 3.3 mm focal length and a 1 mm lens pitch. Figure 12-26 shows an experiment in which the capability of the process for detecting out-of-plane rotation is verified. We fixed the reference object at $(x_r, y_r, z_r) = (0 \text{ mm}, 0 \text{ mm}, 25 \text{ mm})$ and rotated the signal object located at $(x_s, y_s, z_s) = (5 \text{ mm}, 0 \text{ mm}, 40 \text{ mm})$ by $\theta_{x-z}=6^\circ$ while $\theta_{y-z}=0^\circ$. One reference 2D sub-image is correlated with each 2D sub-image of the signal object. In the correlation operation, the joint power spectrum was obtained optically using a He-Ne laser, an SLM with a 0.036 mm pixel pitch and a CCD, and it was then Fourier transformed digitally to produce the correlation peak. As can be seen in Fig. 12-26, the profile of the correlation

peak intensity correctly reflects the out-of-plane rotation of the signal object. By finding the signal 2D sub-image that yields the maximum correlation peak, the out-of-plane rotation angle (θ_{x-z} , θ_{y-z}) can be precisely detected.

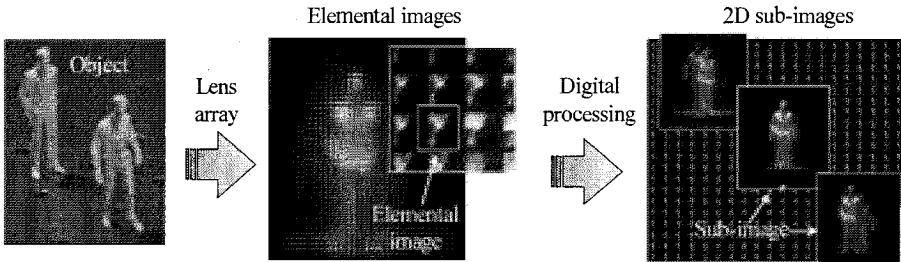


Figure 12-25. Examples of experimentally obtained elemental images and 2D sub-images.

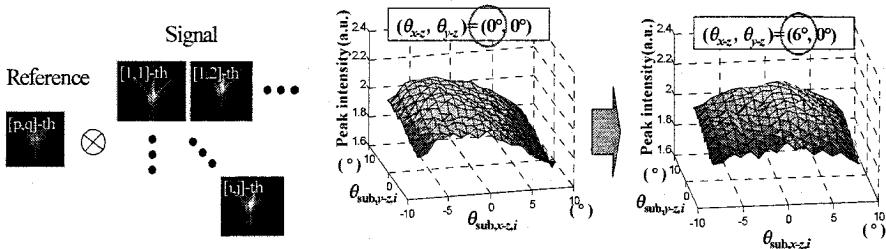


Figure 12-26. Experimental result: intensity profile of the correlation peaks between one sub-image for a reference object located at $(x_r, y_r, z_r) = (0 \text{ mm}, 0 \text{ mm}, 25 \text{ mm})$ and each sub-image of a signal object located at $(x_s, y_s, z_s) = (5 \text{ mm}, 0 \text{ mm}, 40 \text{ mm})$ with $\theta_{x-z} = 0^\circ$ and 6° while $\theta_{y-z} = 0^\circ$.

After the rotation angle is detected, the 3D location of the signal object can be detected by finding the correlation peak positions of two 2D sub-image pairs using Eq. (12-20). Figure 12-27 shows the detected positions of the correlation peak with various locations of the signal object. Equations (12-19) and (12-20) indicate that the slope of the $\tan\theta_{\text{sub},x-z,i}$ -versus- Δu line corresponds to $(z_s - z_r)$ and its Δu -offset corresponds to $(x_s - x_r)$ and z_s . The experimental results for the case of no out-of-plane rotation shown in Fig. 12-27 demonstrate this point clearly. The Δu -offset changes reflecting $(x_s - x_r)$ as shown in Fig. 12-27(a), and the slope increases as the signal object moves farther from the reference object longitudinally as shown in Fig. 12-27(b). This provides convincing support for the 3D shift detection capability of the 3D correlator based on integral imaging.

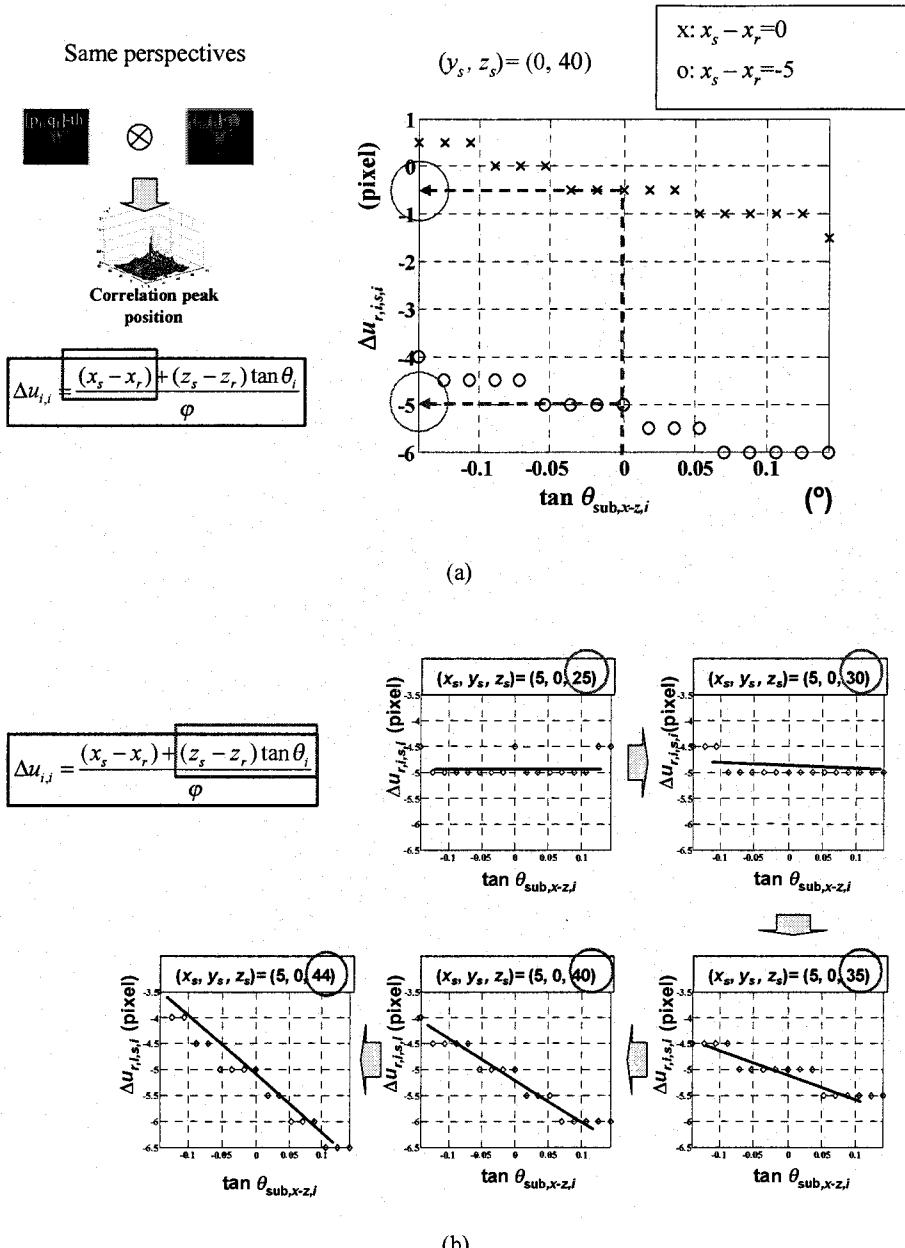


Figure 12-27. Experimental result: detected positions of the correlation peak for various locations of the signal object when the reference object is located at $(x_r, y_r, z_r) = (0 \text{ mm}, 0 \text{ mm}, 25 \text{ mm})$ and the signal object has no out-of-plane rotation. (a) Transverse position detection, (b) longitudinal position detection.

6. SUMMARY

Integral imaging is a useful method for realizing 3D displays and the optical 3D information processing. The use of a 2D lens array enables us to capture and reproduce the directional ray distribution that contains the 3D information of the object. By capturing the directional ray distribution using the lens array, 3D information of the object can be extracted or correlated. By reproducing the directional ray distribution using the lens array, 3D images can be displayed in space. Based on this principle, we reviewed recent schemes for obtaining, correlating and presenting 3D information of an object. Integral imaging is a feasible candidate for 3D television because it can provide a 3D view for multiple observers in the viewing angle even if they lie down when they watch television. Optical 3D information processing is a new and exciting field of research. Although we could not review all works from various research groups on integral imaging, we hope this chapter provides a fundamental understanding of this technique and will stimulate attempts to realize 3D television and information processing using integral imaging.

REFERENCES

1. G. Lippmann, "Épreuves réversibles. Photographies intégrales," *Comptes Rendus de l'Académie des Sciences* **146**, 446-451 (1908).
2. F. Okano, H. Hoshino, J. Arai, and I. Yuyama, "Real-time pickup method for a three-dimensional image based on integral photography," *Appl. Opt.* **36**, 1598-1603 (1997).
3. Y. Igarishi, H. Murata, and M. Ueda, "3D display system using a computer-generated integral photograph," *Japanese J. Appl. Phys.* **17**, 1683-1684 (1978).
4. S.-W. Min, S. Jung, J.-H. Park, and B. Lee, "Three-dimensional display system based on computer-generated integral photography," *The 2001 Stereoscopic Displays and Applications Conference, Proc. SPIE* **4297**, pp. 187-195, San Jose, CA, Jan. 2001.
5. B. Lee, S. Jung, S.-W. Min, and J.-H. Park, "Three-dimensional display by use of integral photography with dynamically variable image planes," *Opt. Lett.* **26**, 1481-1482 (2001).
6. J.-H. Park, S.-W. Min, S. Jung, and B. Lee, "Analysis of viewing parameters for two display methods based on integral photography," *Appl. Opt.* **40**, 5217-5232 (2001).
7. J.-S. Jang, F. Jin, and B. Javidi, "Three-dimensional integral imaging with large depth of focus by use of real and virtual image fields," *Opt. Lett.* **28**, 1421-1423 (2003).
8. S.-W. Min, J. Kim, and B. Lee, "New characteristic equation of three-dimensional integral imaging system and its applications," *Japanese J. Appl. Phys.* **44**, L71-L74 (2005).
9. J. Hong, J.-H. Park, J. Kim, and B. Lee, "Elemental image correction in integral imaging for three-dimensional display," *Proceedings of 2004 IEEE Lasers and Electro-Optics Society Annual Meeting*, Puerto Rico, paper ML6, pp. 116-117, Nov. 2004.
10. J. Hong, J.-H. Park, J. Kim, and B. Lee, "Analysis of image depth in integral imaging and its enhancement by correction to elemental images," *Novel Optical Systems Design and Optimization VII, Proc. SPIE* **5524**, Denver, Colorado, pp. 387-395, Aug. 2004.

11. J. Hong, J.-H. Park, S. Jung and B. Lee, "A depth-enhanced integral imaging by use of optical path control," *Opt. Lett.* **29**, 1790-1792 (2004).
12. B. Lee, S.-W. Min, and B. Javidi, "Theoretical analysis for three-dimensional integral imaging systems with double devices," *Appl. Opt.* **41**, 4856-4865 (2002).
13. S.-W. Min, J. Hong, and B. Lee, "Analysis of an optical depth converter used in a three-dimensional integral imaging system," *Appl. Opt.* **43**, 4539-4549 (2004).
14. H. Choi, Y. Kim, J.-H. Park, S. Jung, and B. Lee, "Improved analysis on the viewing angle of integral imaging," *Appl. Opt.* **44**, 2311-2317 (2005).
15. H. E. Ives, "Optical properties of a Lippmann lenticulated sheet," *J. Opt. Soc. Am.* **21**, 171-176 (1931).
16. R. L. de Montebello, "Wide-angle integral photography – the integrang system," *Proc. 1977 SPIE Annual Tech. Conf.*, San Diego, CA, seminar 10, no. 120-08, Tech. Digest, pp. 73-91, Aug. 1977.
17. L. Yang, M. McCormick, and N. Davies, "Discussion of the optic of a new 3-D imaging system," *Appl. Opt.* **27**, 4529-4534 (1988).
18. J. Arai, F. Okano, H. Hoshino, and I. Yuyama, "Gradient-index lens-array method based on real-time integral photography for three-dimensional images," *Appl. Opt.* **37**, 2034-2045 (1998).
19. B. Lee, S. Jung, and J.-H. Park, "Viewing-angle-enhanced integral imaging by lens switching," *Opt. Lett.* **27**, 818-820 (2002).
20. J.-S. Jang and B. Javidi, "Moving array lens technique (MALT) for improved resolution of all-optical three-dimensional projection," *Proc. SPIE* **4789**, 10-15 (2002).
21. Y. Jeong, S. Jung, J.-H. Park, and B. Lee, "A reflection-type integral imaging scheme for displaying three-dimensional images," *Opt. Lett.* **27**, 704-706 (2002).
22. S. Jung, J.-H. Park, H. Choi, and B. Lee, "Wide-viewing integral three-dimensional imaging by use of orthogonal polarization switching," *Appl. Opt.* **42**, 2513-2520 (2003).
23. H. Choi, J.-H. Park, J. Hong, and B. Lee, "Depth-enhanced integral imaging with a stepped lens array or a composite lens array for three-dimensional display," *The 16th Annual Meeting of the IEEE Lasers & Electro-Optics Society*, Tucson, Arizona, **2**, pp. 730-731, Oct. 2003.
24. J. Seo, S. Park, J. Jang, S. Cha, and S.-H. Shin, "Orthoscopic integral imaging 3D display by use of negative lens array," *Photonics Conference 2003*, Jeju, Korea, pp. 797-798, Nov. 2003.
25. Y. Kim, J.-H. Park, H. Choi, S. Jung, S.-W. Min, and B. Lee, "Viewing-angle-enhanced integral imaging system using a curved lens array," *Opt. Express* **12**, 421-429, (2004).
26. H. Liao, M. Iwahara, N. Hata, and T. Dohi, "High-quality integral videography using a multiprojector," *Opt. Express* **12**, 1067-1076 (2004).
27. J.-H. Park, H.-R. Kim, Y. Kim, J. Kim, J. Hong, S.-D. Lee, and B. Lee, "Depth-enhanced three-dimensional-two-dimensional convertible display based on modified integral imaging," *Opt. Lett.* **29**, 2734-2736 (2004).
28. H. Choi, J.-H. Park, J. Kim, S.-W. Cho, and B. Lee, "An improved 2D/3D convertible integral imaging with two parallel display devices," *Proc. 2005 International Meeting on Information Display*, Seoul, Korea, vol. I, pp. 46-49, July 2005.
29. S. Jung, J.-H. Park, H. Choi, and B. Lee, "Viewing-angle-enhanced integral three-dimensional imaging along all directions without mechanical movement," *Opt. Express* **11**, 1346-1356 (2003).
30. S.-W. Min, J. Kim, and B. Lee, "Wide-viewing projection-type integral imaging system using an embossed screen," *Opt. Lett.* **29**, 2420-2422 (2004).

31. Y. Kim, J.-H. Park, S.-W. Min, S. Jung, H. Choi, and B. Lee, "Wide-viewing-angle integral three-dimensional imaging system by curving a screen and a lens array," *Appl. Opt.* **44**, 546-552 (2005).
32. S. Jung, J. Hong, J.-H. Park, Y. Kim, and B. Lee, "Depth-enhanced integral-imaging 3D display using different optical path lengths by polarization devices or mirror barrier array," *J. Society for Information Display* **12**, 461-467 (2004).
33. H. Choi, Y. Kim, J.-H. Park, J. Kim, S.-W. Cho, and B. Lee, "Layered-panel integral imaging without the translucent problem," *Opt. Express* **13**, 5769-5776 (2005).
34. S.-W. Min, M. Hahn, J. Kim, and B. Lee, "Three-dimensional electro-floating display system using an integral imaging method," *Opt. Express* **13**, 4358-4369 (2005).
35. J.-H. Park, Y. Kim, J. Kim, S.-W. Min, and B. Lee, "Three-dimensional display scheme based on integral imaging with three-dimensional information processing," *Opt. Express* **12**, 6020-6032 (2004).
36. J.-H. Park, S.-W. Min, S. Jung, and B. Lee, "A new stereovision scheme using a camera and a lens array," *Conference on Algorithms and Systems for Optical Information Processing V, Proc. SPIE* **4471**, San Diego, CA, pp. 73-80, Jul.-Aug. 2001.
37. J.-H. Park, S. Jung, H. Choi, Y. Kim, and B. Lee, "Depth extraction by use of a rectangular lens array and one-dimensional elemental image modification," *Appl. Opt.* **43**, 4882-4895 (2004).
38. T.-C. Poon and T. Kim, "Optical image recognition of three-dimensional objects," *Appl. Opt.* **38**, 370-381 (1999).
39. B. Javidi and E. Tajahuerce, "Three-dimensional object recognition by use of digital holography," *Opt. Lett.* **25**, 610-612 (2000).
40. J. Rosen, "Three-dimensional optical Fourier transform and correlation," *Opt. Lett.* **22**, 964-966 (1997).
41. J. Rosen, "Three-dimensional joint transform correlator," *Appl. Opt.* **37**, 7438-7544 (1998).
42. J. Esteve-Taboada, D. Mas, and J. Garcia, "Three-dimensional object recognition by Fourier transform profilometry," *Appl. Opt.* **38**, 4760-4765 (1999).
43. J.-H. Park, S. Jung, H. Choi, and B. Lee, "Detection of the longitudinal and the lateral positions of a three-dimensional object using a lens array and joint transform correlator," *Optical Memory and Neural Networks* **11**, 181-188 (2002).
44. J.-H. Park, J. Kim, and B. Lee, "Three-dimensional optical correlator using a sub-image array," *Opt. Express* **13**, 5116-5126 (2005).

Chapter 13

AUTOSTEREOSCOPIC, PARTIAL PIXEL, SPATIALLY MULTIPLEXED, AND OTHER 3D DISPLAY TECHNOLOGIES

Monish R. Chatterjee¹ and Shih-Tun Chen²

¹*University of Dayton, Dayton, OH 45469;* ²*Teradyne Inc., Agoura Hills, CA 91301*

Abstract: The technologies and advances in the field of three-dimensional (3D) displays within the past two decades are reviewed. Specifically, the developments in diffractive, refractive, reflective and occlusive 3D display strategies are discussed, highlighting the advantages and limitations of the associated systems. The partial pixel and the partial object pixel architectures associated with autostereoscopic displays are discussed in some detail, while other techniques are briefly introduced. It is shown that major breakthroughs occurred with the development of computer-generated holography (CGH), liquid crystal arrays (LCA) and other spatial light modulators (SLM), and also the discrete fabrication of diffractive and holographic optical elements (DOE and HOE), micromirror arrays, and microelectromechanical systems (MEMS). These techniques and components have enabled the implementation of spatially and angularly multiplexed 3D images and scenes with potentially dynamic (time-varying) features.

Key words: 3D displays; partial pixels; spatial multiplexing; holographic gratings; autostereoscopic displays; holographic stereograms.

1. INTRODUCTION

In recent years, the use of three-dimensional (3D) effects has experienced a resurgence. Many advertising labels use 3D stickers, and are increasingly seen in 3D audio and visual displays. In addition, a great number of medical and technological applications can be found in a variety of industries, including magnetic resonance imaging, computerized axial

imaging (CAT scan) imaging, architectural plan development and flight simulation. In theory, the 3D visual effect is generated by a slight spatial difference between the eyes of an observer. However, standard display technologies such as the cathode ray tube (CRT), liquid crystal display (LCD), and electric plasma display cannot offer a realistic 3D effect. The 3D pixels for a 3D display system, which diffract the incident light into an intensity distribution over the spatial domain, may be seen as a hologram consisting of a number of (superposed) gratings. Yet, techniques for the dynamic control or programming of the erasure and recording (OFF and ON switching) of these gratings are not readily available. This chapter will review the current state-of-the-art in 3D display technologies, providing an overall picture of the progress made in this field in the past decade or more. Brief introductions will be given to some of the well-known methodologies; in a few cases, such as with the *partial pixel*, or *spatial multiplexing architecture*, a more detailed description will be provided. Developed in 1995, the partial object pixel recording technique discussed in this chapter records a pixelated hologram which emulates a composite 3D pixel array for 3D display systems. The partial object pixel technique allows one to record a hologram that yields the predefined intensity distribution during the recording phase. Computer-generated object beam patterns for specific intensity distributions are recorded on a pixel area with a traditional two-beam recording setup. In section 2, modern 3D/stereoscopic display methodologies are reviewed. There are several approaches producing a 3D effect; these include the Virtual Reality (VR) technique, red-blue stereoscopic 3D display, polarized stereoscopic 3D displays, and LCD stereoscopic 3D displays with holographic/diffractive elements. In the VR technique, two LCDs are used to feed the left and right images (generated by a computer 3D image synthesizer program) to the eyes of an observer. The observer has to wear a special optical viewing device in order to correctly intercept the two images. Such devices are now widely used in the computer games industry because of their total capability in replacing the observer's surrounding scene with a virtual one. In contrast to the VR technique, the red-blue stereoscopic 3D display technique is not a recent innovation. Movies based on this technique have been made decades ago. New commercial products have been developed for scientific and medical visualization applications. For example, Silicon Graphics® Computer Inc. has recently announced a three-dimensional image-rendering system using stereoscopic 3D display techniques. However, a more prevailing 3D methodology is to use computer software in order to generate 3D scenes; this does not involve any specialized 3D output device. By interacting the user with vividly animated/synthesized 3D images, a low cost, *pseudo* 3D effect

may be realized. Although each of these techniques has its own advantages in rendering 3D effects, the displays may improve greatly if a true 3D effect, i.e., a holography-like effect is implemented, since holograms have long been regarded as true 3D images. However, because of the difficulties involved in dynamically recording and reconstructing holographic images, a fundamental breakthrough in the holographic display technique needs to be made in order to renovate this technology. Recently, a *partial pixel* technique has been developed for 3D displays utilizing the available LCD technology in conjunction with diffractive optical elements (Nordin et al., 1995). Similar devices using holographic optical elements have also been reported elsewhere (DeMeis, 1996). However, these efforts have some limitations leading to reduced pixel density on the display, and they only generate a limited number of parallax images for multiscopic display. A synthetic aperture holography was reported by Benton *et al.* in 1992 (St.Hilaire et al., 1992). This synthetic aperture approach projects the fringe image from an AOM towards an image plane, and scans the image plane to display a 3D image. A 2D to 3D conversion scheme is proposed later in the same section. This scheme is used to transform conventional videos or motion pictures for stereoscopic presentations. This conversion technique provides the (auto)stereoscopic display systems access to the available 2D format media, and expands the scope of applications beyond animated computer pictures. An autostereoscopic scheme to utilize the available stereoscopic technique and viewing angle detection is discussed briefly. Section 2 concludes with a short discussion on a novel light pipe approach to generate 3D pixel arrays. A spatially multiplexed, partial-pixel architecture to generate 3D images of multiple scenes over a range of viewing angles, was developed (Chen and Chatterjee, 1997). This architecture is discussed in some detail in section 3. Finally, section 4 takes a look at the emerging 3D display technologies worldwide since 1995, outlining significant achievements, underscoring limitations, and predicting futuristic trends in this technology.

2. SOME PERSPECTIVES ON 3D DISPLAY METHODOLOGIES PRIOR TO 1995

Three-dimensional stereoscopic display technology may be described in terms of three major constituents, viz., a three dimensional display medium, electronic hardware, and associated software. The most demanding aspect of these constituents was the search for a three dimensional display medium since a dynamic, holographic 3D display medium was not readily available. Although work on holographic stereograms, including a conical holographic

stereogram and a partial object holographic stereogram (Murillo-Mora et al., 1995; Chen and Chatterjee, 1995) had been reported prior to 1995, these techniques were based on static holographic gratings. Dynamic, erasable or programmable holographic gratings were still in the materials research phase, and not commercially available.

Conventional technologies including liquid crystal displays, electric plasma displays, and cathode ray tubes greatly improved during the preceding decade (1985-1995). Liquid crystal displays (LCDs) underwent improvements in viewing angle, array size, pixel contrast, and pixel color in the past ten years; the capabilities of electrical plasma displays were extended to include color pixels and large display areas; moreover, conventional CRTs were available with larger sizes than ever before. In the mid-1990s, an international agreement was reached regarding high definition television (HDTV) standards. On the other hand, technological advances in integrated electronics and electronic computing were increasingly more ready to support the hardware and software developments of the next generation of three-dimensional displays. Research on 3D technology had support simultaneously from the electronic, optical and software fields.

The effort to develop an electronically controllable holographic grating encompasses several optical and materials research areas, including the nonlinear optical properties of photorefractive materials, nematic liquid crystals, and photo-polymers. Khoo (1995) reported the holographic grating formation in fullerene-C₆₀-doped liquid crystals. Meerholz et al. (1994) successfully recorded high diffraction efficiency holographic gratings in photorefractive polymers. Some of the dynamic, 3D display techniques developed prior to the mid-1990s included holographic movies, aperture synthetic 3D display systems, partial pixel stereoscopic display systems, LCD stereoscopic display systems with diffractive optical elements, polarized stereoscopic display systems, color-coded (red-blue) stereoscopic display systems, and virtual reality (VR) rendering systems (Joachimowicz and Gold, 1990; Palais and Miller, 1996; Ernst et al., 1996; Kulick et al., 1995; Travis, 1990; MacFarlane 1994; St. Hilaire et al., 1992; Takeda et al., 1995). Some of these techniques have been applied to developing graphical interfaces for computer games such as VR simulation systems and color-coded stereograms applied to a traditional CRT monitor. Some are still in the laboratory stage, including aperture synthetic and partial pixel stereoscopic systems. The results vary from technique to technique depending on cost, complexity, and resource requirements.

We next examine 3D display techniques as they existed prior to 1995 (Chen, 1996). These techniques include color-coded stereoscopic display,

polarized stereoscopic display, LCD 3D display systems, spatially multiplexed 3D display systems, and synthetic aperture holographic display.

2.1 Color-coded stereoscopic display

This technique was among the first few to become commercially available. It requires only a nominal cost in simple hardware. In a color-coded stereoscopic system, as depicted in Fig. 13-1, the two stereo scenes are colored with different primary colors (commonly red and blue) and are to be seen through a pair of special glasses with two lenses colored with the two colors.

The two scenes with different color tones are individually enhanced by the filtering effect from two colored lenses. The observer may perceive the different color enhancements at the left and right eyes when a color-coded picture and special glasses are used. As a result, if the picture is properly color-coded, the system renders a three-dimensional experience to an observer. The advantages of such a system are low cost and ease of implementation on conventional CRT television sets. However, the discoloration of the original color images is one disadvantage of the technique. One may still have a feel for the distinctive depth differences between different parts of the objects in the picture; however, the colors of the objects may suffer from unevenness. The resulting stereoscopic display does not have vivid colors, and the viewer sometimes experiences binocular rivalry from extended use (Boff et al., 1986).

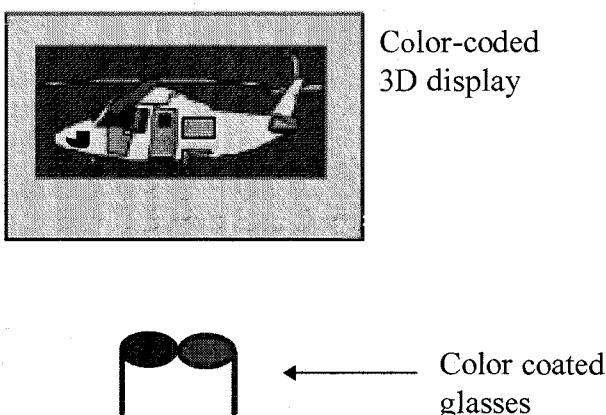


Figure 13-1. Color-coded stereoscopic pictures may be viewed with a regular TV/monitor and a special pair of glasses.

2.2 Polarized stereoscopic display systems

Another important research sub-topic in 3D display systems pre-1995 was one that embedded the polarization difference to the two images and renders stereoscopic pictures through a pair of polarized glasses. Typically, two orthogonally polarized light beams (usually vertical and horizontal) are generated using either two light sources passing through two different polarizers, or an electronically-controlled half-wave retarder to change the polarization according to the vertical synchronization for time-multiplexing (Jachimowicz and Gold, 1990).

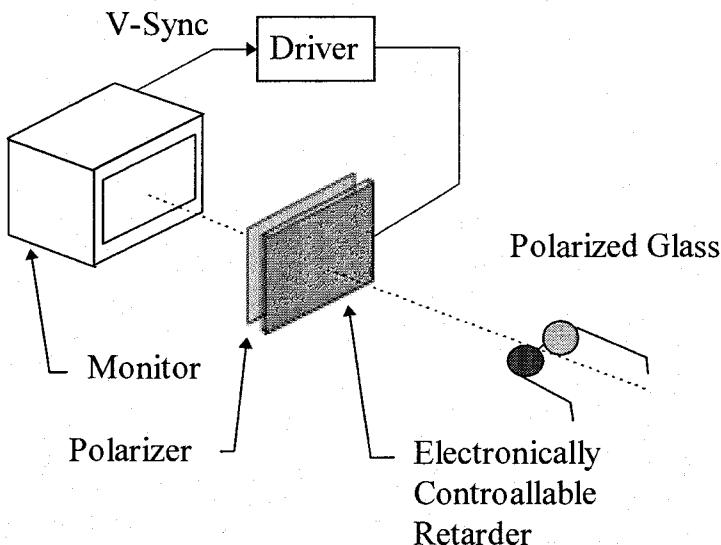


Figure 13-2. Time-multiplexed polarization stereoscopic display system (after Jachimowicz and Gold, 1990).

In a time-multiplexing polarization stereoscopic system as depicted in Fig. 13-2, image frames of each side have to be presented 60 frames per second in order to eliminate the flickering effect; therefore, for the two sides, the V-Sync is set at 120 Hz, as is the electronically controllable retarder control signal. This time-multiplexing polarization technique requires twice the bandwidth for the display module (Jachimowicz and Gold, 1990).

Two orthogonal polarized light sources may be used and eliminate the need for time-multiplexing. The light sources may be video projection CRTs or liquid crystal projectors with polarizers used to control the light polarization. Both images are projected onto a screen and the observer wears a pair of polarized glasses as depicted in Fig. 13-3.

The disadvantage for a double-wave polarization stereoscopic display system is the need to maintain the polarization for the light waves once the light hits the projection screen. A diffusion effect which scrambles the polarization of the incident light on the projection screen has to be taken care of. One way to minimize such a problem is to use a reflecting metal screen. The other problem associated with the system is that the screen has to be curved in order to keep the spot images from reflecting the incident light away from the viewer. Although the selection of a reflecting metal projection screen may be beneficial the maintenance of the polarization, it is not the best choice for projecting images.

Projection Screen

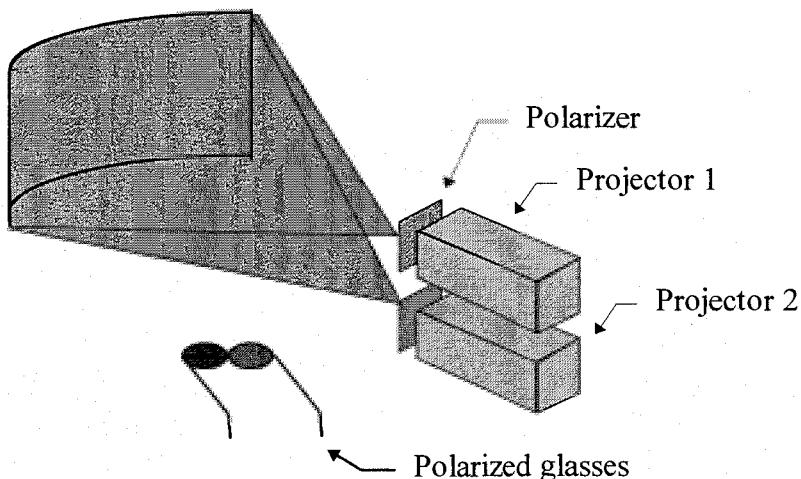


Figure 13-3. A polarized double projection stereoscopic display system. Note that to avoid the hot spot effect and maintain polarization of the light, a reflecting, curved metal projection screen is required.

2.3 Time-multiplexed shutter stereoscopic display

The use of the concept of switching image frames between right and left eyes is limited to time-multiplexed polarization stereoscopic displays only.

A time-multiplexed shuttering stereoscopic display employs a similar methodology (Palais and Miller, 1996). A pair of electronically controllable LC shutter glasses, shown in Fig. 13-4, is used for the image separation between the two frames.

The LC shutter acts as a pair of active glasses, and allows one image to be seen by one eye in a time slot, and then the other eye, alternatively. The special glasses are equipped with an IR receiver to obtain the synchronization signal for maintain the timing with the images presented on the monitor.

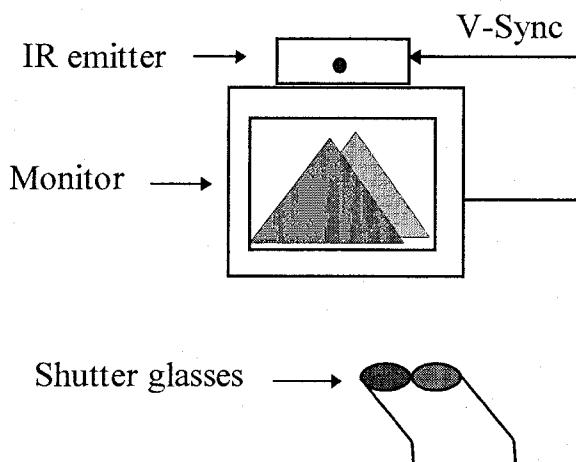


Figure 13-4. A shuttering stereoscopic display system. The shutter glasses is made with LC and is controlled by the emitter IR signal to shut off one of the glasses at a time.

Such a system is well suited for the current development of 3D graphical presentation on a computer, since it has an integrated optical interface and requires a minimum optical alignment setup. However, it would be more beneficial for such a system to become autostereoscopic.

2.4 LCD 3D display systems with diffractive (holographic) optical elements

Since LCD technologies around the mid-1990s were more advanced and used much more widely than the preceding decade, interest was also focused on modifying the existing LCD technology such that LCDs might be used in the 3D displays of the future. Jones et al.(1995) reported a partial pixel architecture which uses diffractive optics in conjunction with an LCD module to generate a 3D effect. This technique uses a concept similar to that

of the pixelated hologram as the partial object pixel technique described in section 3; there, each 3D pixel is responsible for the diffraction pattern over its projection space.

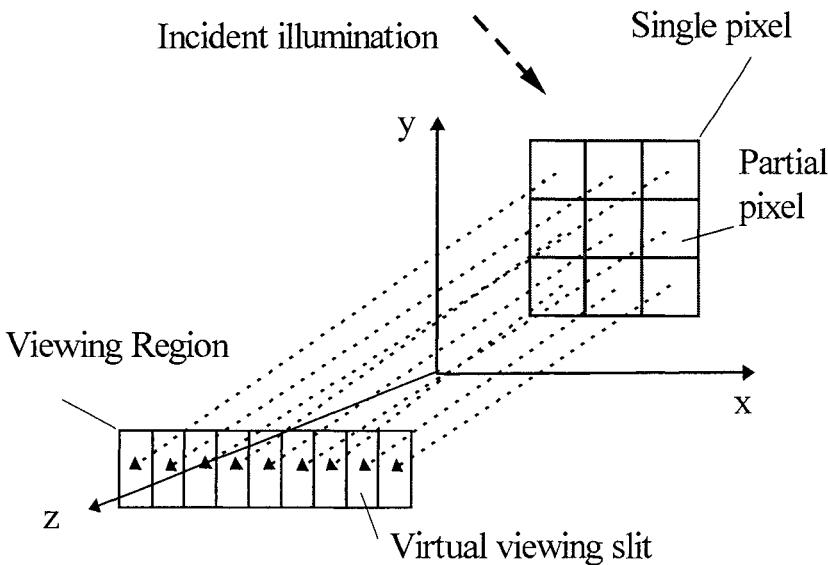


Figure 13-5. Schematic representation of a single pixel composed of multiple partial pixels, each of which directs light to a single virtual viewing slit. (after Kulick et al., 1995).

The partial pixel approach was developed because of the difficulties encountered with a composite approach. A composite pixel, somewhat like the partial object pixel described later in section 3, consists of layers of diffractive components (gratings) which are responsible for diffracting light into space. Dynamic gratings embedded in a nematic liquid crystal were under investigation pre-1995, and were not readily useful in constructing a composite pixel. In the partial pixel approach, the available display area of the pixel is divided into several sub- or partial pixels. These sub-pixels diffract light into an array of viewing windows distributed over an angular range, as depicted in Fig. 13-5. The diffractive elements on the partial pixels vary according to their location on the module.

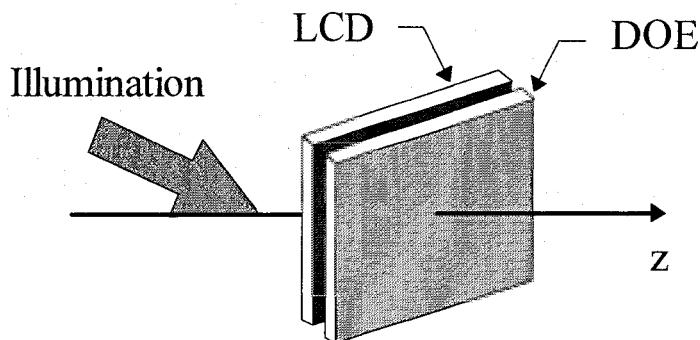


Figure 13-6. An LCD configuration with diffractive optical elements as a 3D display (after Jones et al., 1995).

The diffractive optical elements are placed in front of the LCD as depicted in Fig. 13-6. Each of the LCD pixels corresponds to one partial pixel of a resulting 3D pixel, or a partial pixel array. One such 3D display with 135×90 pixels, of which each 3D pixel consists of a 4×4 partial-pixels array, has been reported by Jones et al. (1995). The device was designed to multiplex sixteen images over the same number of virtual viewing slits or windows. The advantage of this technique is that it combines the available LCD technology with diffractive optics. However, the pixels seen by the observer in a partial pixel approach suffer from pixel density reduction. The pixel densities of the original LCD and the resulting 3D display have a reduction ratio of sixteen to one, i.e., from 540×360 to 135×90 pixels. Even though the information-space bandwidth product remains the same, the information per spatial area has been reduced.

Through the mid-1990s, other groups used similar spatial multiplexing techniques; these included Trayner et al. at Richmond Holographic Studios, London, England (holograms), and Takahashi and co-workers at Toppan Printing, Tsukuba, Japan (diffractive optics). These groups reported their work using either a holographic plate or diffractive optical elements in front or behind an LCD to generate a multiplexing effect (DeMeis, 1996; Bains, 1996; Takahashi et al., 1991). Some of the advantages and disadvantages of using diffractive optics or holograms are listed in Table 13-1.

Table 13-1. Comparison charts for optical elements selection of LCD 3D display systems in 3D applications.

LCD with DOE	LCD with Hologram
<ul style="list-style-type: none"> • Hard to manufacture • Limited to small sizes • High cost • Clearer & brighter views 	<ul style="list-style-type: none"> • Easier to manufacture • Fits large display area • Lower cost • Lower contrast & efficiency • Possible use in portables • Coherent illumination (better performance)

The manufacturing of diffractive optical elements (gratings) is more difficult compared with that of holograms. It costs even more when the area of the 3D display is increased. Because of this, the partial pixel approach, or Takahashi's color 3D display is kept smaller in size than 2×2 sq. inches (40×40 mm 2).

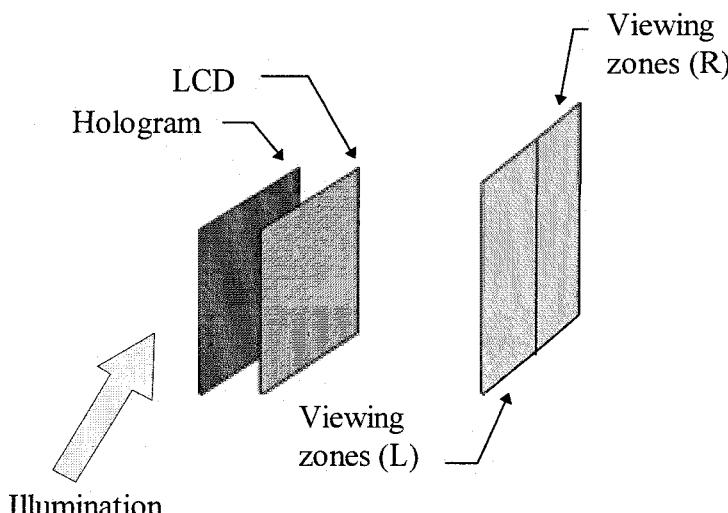


Figure 13-7. 3D display using LCD/hologram combination (Bains, 1996).

In the LCD and hologram approach as depicted in Fig. 13-7, a holographic plate is placed before the illuminating light reaches the LCD module to steer the light along the predefined viewing zones. It is also possible to use a substrate hologram in such a 3D display system to minimize the physical size.

In the Takahashi approach, as depicted in Fig. 13-8, the illuminating white light is first diffracted into three primary colors (red/green/blue) by a color filter for the three sub-pixels. The light from these three color filter cells is then diffracted by three optical gratings (designed for each color) along a predefined direction. The three liquid crystal sub-pixels after the gratings decide whether to let each color go through or not.

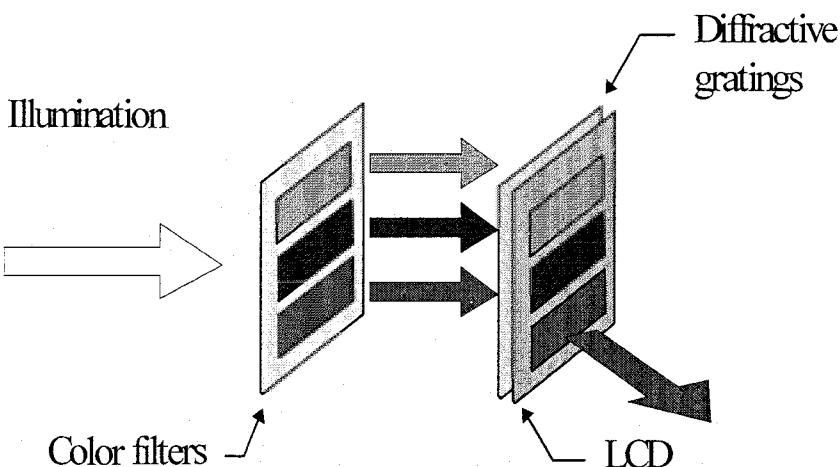


Figure 13-8. 3D display using LCD/DOE combination, showing a sub-pixel (DeMeis, 1996).

2.5 Spatially multiplexed autostereoscopic display systems

This technique is similar in concept to a time-multiplexing technique, in that a time-multiplexed LCD is used. However, the spatial multiplexing capability comes from the sequentially separated illuminations depicted in Fig. 13-9. While the illumination comes from alternately changed illuminating angles, the LCD module displays images corresponding to the illuminating angles.

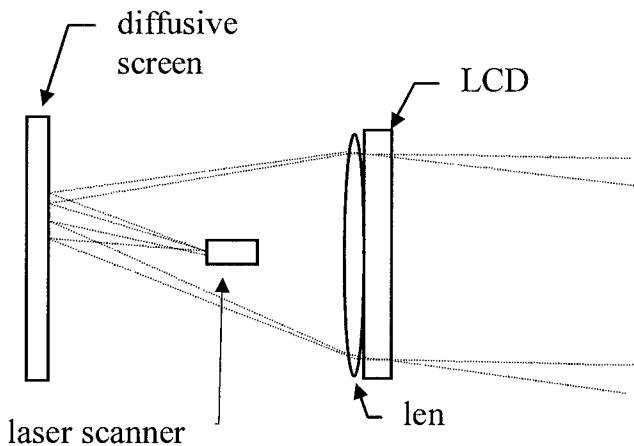


Figure 13-9. A spatially multiplexed 3D display system developed at Cambridge University (Travis, 1990).

The display medium of such a spatially 3D display has been extended to include conventional CRTs from the original LCD design (Dodgson, 1996). However, the fact that the current system design cannot fit into space-restricted applications or portable environments is a potential disadvantage.

2.6 Virtual Reality (VR) stereoscopic display

Virtual Reality 3D displays apply two liquid crystal displays to cast two separated images toward the user in its setup, as shown in Fig. 13-10. VR displays are usually manufactured as optical goggles, in which the VR display system is masked by the goggles worn by the user. Simulated scenes generated by an appropriate 3D animation software replace the “reality”, such that a virtual reality may be felt.

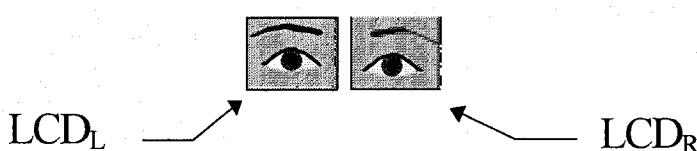


Figure 13-10. In the virtual reality technique, each eye sees a different LCD display.

2.7 Synthetic aperture holographic 3D display system

This technique was developed at the Media Laboratory of MIT by St. Hilaire et al. (1992). Acousto-optic modulators are used as dynamic holograms and the images of the acoustic fringes within these modulators are projected to the image plane through a mechanical scanning setup. The horizontal scanning geometry, which is to maintain the stationarity of the acoustic fringe patterns, is depicted in Fig. 13-11 for illustrating its basic operations.

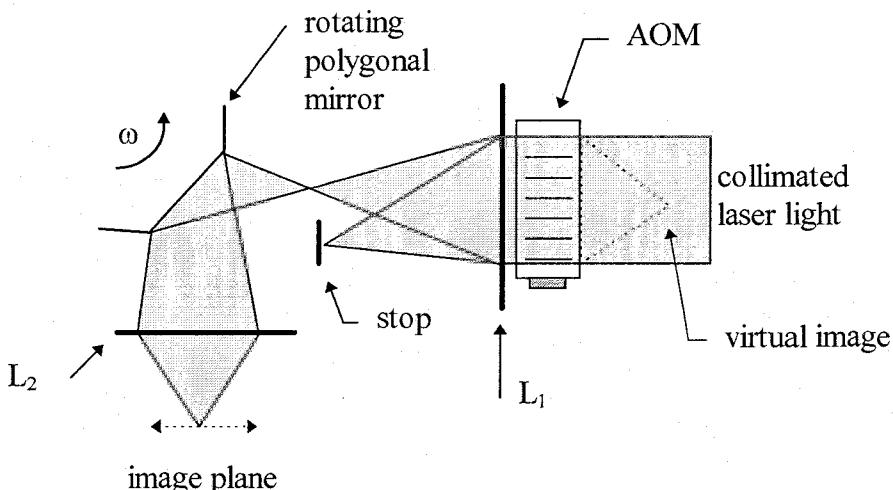


Figure 13-11. The MIT electronic holographic 3D display system showing the horizontal scanning geometry.

Different horizontal lines are generated by the adjustment of the vertical scanning mechanism. A maximum of 192 lines was demonstrated (limited by the memory size of the image frame buffer) by the MIT group. A three-channel acousto-optic modulator is also applied for a color application. Although this technique presents a near-holographic approach, the required resources to attain this goal is very costly. The viewing angle of the MIT system is 15°, the display area size is 5cm by 5 cm, and each image frame contains 4.1×10^6 samples. The resulting bandwidth is 2Gbits/s for 60 frames per second. The required computation was achieved by using a Thinking Machines Connection Model 2 massively parallel minisupercomputer (16,384 processors), with an image frame buffer.

3. SPATIALLY MULTIPLEXED PIXELATED 3D MESSAGE DISPLAYS USING A HOLOGRAPHIC OPTICAL ELEMENT (HOE) ARRAY

In this section, a pixelated holographic stereogram generated by using a computer is proposed and studied for emulation as a composite 3D pixel spatially multiplexed display (Chen, 1996). This approach utilizes pixelated holograms to compose spatially multiplexed images. Each holographic pixel has a diffraction pattern which scatters light into pre-defined spatial directions. Under reconstruction, such a holographic pixel generates different intensity patterns along different viewing angles. Eventually, when a holographic pixel array is assembled, it has the capacity to potentially deliver 3D effects. The novel recording scheme is designed to synthesize a computerized 3D display system.

3.1 Background

In traditional display systems, a pixel represents the intensity of the light emitted over a small area defined by the pixel size. The emitted light, in general, will be projected forward with a broad angular spread. In such a system, a viewer cannot intercept or perceive any change in the image when his/her viewing angle is changed.

A spatially multiplexed technique is sometimes called a stereogram when two or more images are viewed (or projected) at the same time from a single “picture”. These images may be spatially multiplexed to construct a 3D effect to the human eyes. There are several techniques available for attaining a 3D effect, such as a computer-controlled virtual reality simulator (as was discussed in section 2), and a maze-like optical illusion stereogram. The holographic technique discussed in this section provides a more natural way of viewing a 3D image, without staring at illusion stereograms or requiring the assistance of optical instruments.

In this section, an HOE array is used to produce an effect similar to that of a stereogram. The concept of using an HOE array as a holographic stereogram is different from the traditional holographic stereograms investigated by King (1968), King et. al. (1969) and Cross (1977), or the so-called multiple photo generated holograms investigated by Spierings and van Nuland (1992). A conventional stereogram technique exposes different segmented slits on the photographic film with prerecorded scenes from different perspectives, i.e., each slit is responsible for a single perspective.

This approach has been proven to be a successful methodology to record computer-generated images (King et al., 1969; Spierings and van Nuland, 1992). However, in the approach discussed here, a *partial object pixel (POP)* recording technique is applied to record the HOE array. An HOE array is used to display a set of spatial multiplexed and discretized (digitized) images. Each pixel in the array contains information from the same pixel position in every perspective; in other words, the images of a given object *in its entirety* from different perspectives are not directly used in the recording. Hence, instead of recording with total pictures, each time we use partial pictures from all images to record each pixel (where each pixel becomes an HOE).

3.2 Holographic Stereogram Based On Computer Generated Holography

In general, holograms are known for their ability to record lightwaves entirely, i.e., record both the amplitude and phase information of the lightwaves. Computer generated holograms may be classified into two categories: Fourier transform holograms (commonly regarded as computer generated holograms or CGH) and holographic stereograms (King, 1968; King et al., 1969).

A computer generated Fourier transform hologram can be made by calculating the Fourier transform of an object wavefront and embedding it as a transmittance function on a film (Yaroslavski and Merzlyakov, 1980). For simplicity, computer generated holograms are often targeted for 2D instead of 3D object patterns. Three-dimensional objects often have complex wavefront functions. These require more complicated computations for creating the corresponding computer generated holograms. On the other hand, a holographic stereogram is composed of a series of pictures from different aspects of an object scene, either by using a camera or a computerized image synthesizing program. These pictures are recorded on small vertical slits of a film using conventional two-beam recording. As a result, a viewer may see a different picture when his/her viewing angle is varied relative to the hologram. To a certain degree, a conventional hologram may be regarded as a linear/continuous version of a holographic stereogram, which continuously reconstructs different images along linearly varied viewing angle. Alternatively, the holographic stereogram can be considered as a digitized/segmented conventional hologram, which reconstructs different images over a segmented viewing angular range. A

simple two-image stereogram projects two images along spatially multiplexed directions, as depicted in Fig. 13-12.

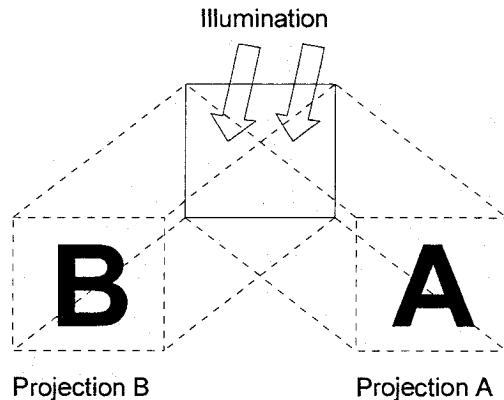


Figure 13-12. A spatially multiplexed display (stereogram) projects two patterns in different directions.

3.3 Recording of a Conventional Holographic Stereogram

The holographic multiplexing technique for stereograms was developed by King et al. (1968). The technique was originally designed to spatially multiplex different scenes of a particular object for 0° - 360° perspectives. In this technique, the holographic stereogram is recorded in a slit-by-slit fashion, as depicted in Fig. 13-13.

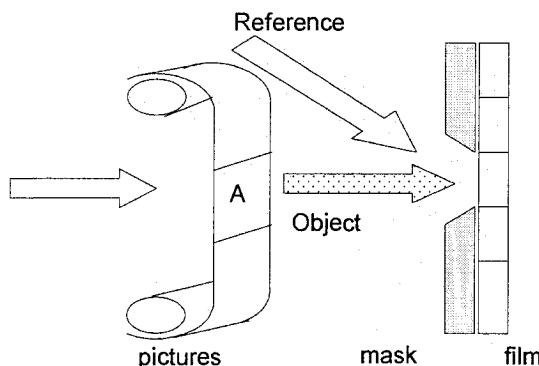


Figure X-13. Recording of holographic stereogram (top view), in which the film is recorded slit by slit (after Iizuka, 1985).

The vertical slits on the hologram are exposed to different scenes so the composite hologram can display a different scene when the viewer moves his viewing angle or perspective.

3.4 Partial object pixel computer generated holography

The partial object pixel hologram is recorded with relatively large object scenes (compared to the size of the hologram), so that the reconstructed scene from the recorded hologram does not cover the whole object. Every time the hologram is viewed, only a small part of the object can be observed. One analogy of this technique is to observe scenes by looking into a small opening with a large object (relative to the opening) placed close behind, as in Fig. 13-14. Similar to the partial object recording, only a small part of the object may be seen from any given viewing angle.

To record a partial object pixel, we expose the film with the object scene projected onto a screen, as depicted in Fig. 13-15. In the figure, a recording for a horizontally multiplexed pixel with nine viewing windows is depicted. The information embedded in the pixel is determined by the arrangement of the intensity distribution on the screen. Therefore, the diffraction pattern of a resulting pixel over the viewing space is determined by the intensity distribution on the projection screen during the recording.

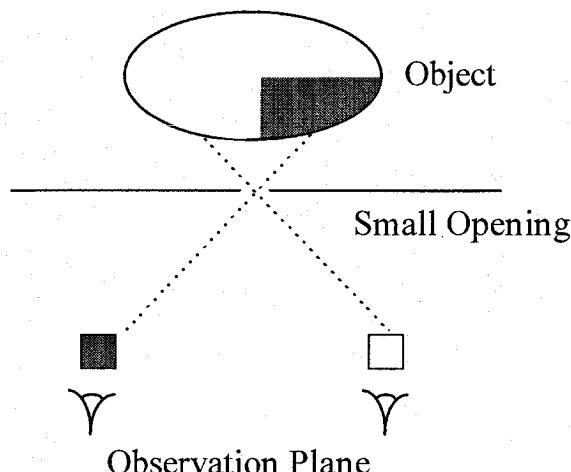


Figure 13-14. The partial object pixel technique is used to make a hologram with properties similar to looking into a small opening in front of a large object at a close range.

Once the pixel is recorded with the partial object scene, its diffraction profile over the viewing angular range is defined according to the content of the screen.

The object beam, which consists of the divided segments on the projection screen, is added with a reference beam for recording an HOE. The two beams are applied to expose the film in a holographic recording setup, as depicted in Fig. 13-16.

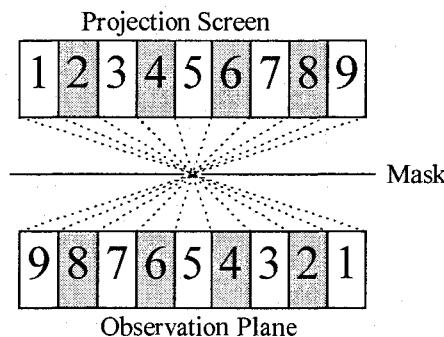


Figure 13-15. The images on the projection screen and its views in the observation plane through a small opening on the mask.

In the figure, the film is placed behind the mask which defines the pixel size and shape. The opening on the mask allows light beams to expose an area for the pixel hologram with the size of and shape of the opening.

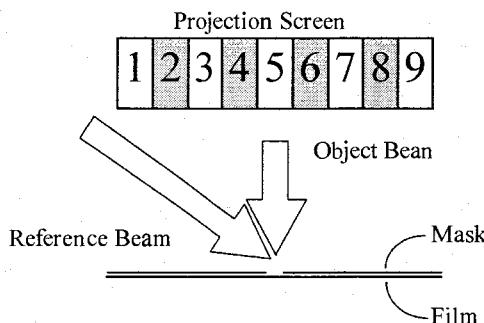


Figure 13-16. Recording scheme for a horizontally multiplexed partial object pixel. The object beam consists of nine object segments.

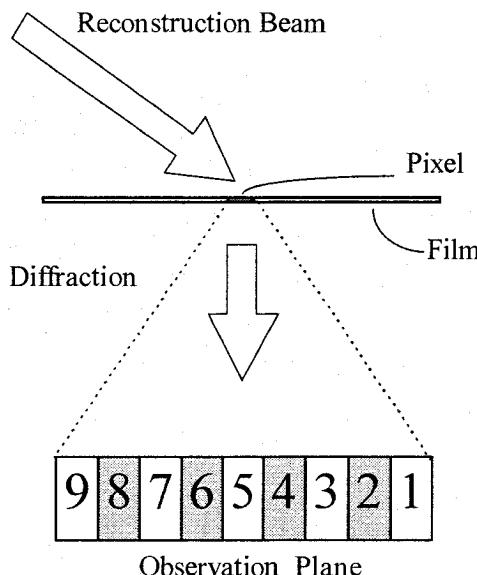


Figure 13-17. The reconstruction of information recorded on a holographic pixel.

After the recording is complete and the film is developed, the recorded information may be reconstructed by using the same reference beam. The pixel diffracts the reference beam along the object beam direction. As the object beam is reconstructed, information on the projection screen during the recording is retrieved, but in a reverse order, as shown in Fig. 13-17.

The left-to-right numbered segments on the projection screen during the recording are reconstructed in the reverse order of right-to-left. During the reconstruction, the viewing windows at the observation plane are multiplexed according to the recorded projection screen. Each window may see a different light intensity on the partial object pixel. The pixel area has a higher intensity when looking toward an object segment recorded bright, and a lower intensity when looking toward one recorded dark one.

3.4.1 Diffraction from an HOE array

The approach here is to use the partial object pixel HOEs as the pixels in a display array. The pixels render individually recorded intensity distributions over the multiplexing space. When the viewing angle of the viewer is varied, the composed image changes accordingly. Therefore, by spatially arranging the multiplexed image pixels, a 3D scenario may be observed.

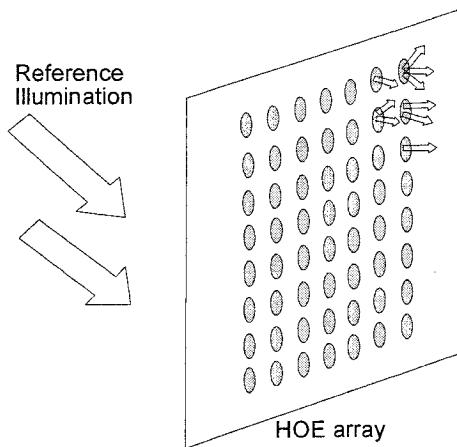


Figure 13-18. An HOE array used for diffracting the reference light along different directions.

In Fig. 13-18, a transmission type HOE array display is shown. As described, each element (as a 3D pixel) in the display is assumed to diffract light into a predefined angular intensity distribution and share the same reference beam for reconstruction with the other elements. This HOE array may compose different images along the multiplexing directions using these 3D pixels.

As an illustration, suppose that two different images are to be viewed from two different directions, e.g., a vertical bar toward the left-hand side, and a horizontal bar to the right-hand side. The light diffracted by the HOE array consists of two images, as depicted by views A and B in Fig. 13-19. A 4×4 holographic optical element array is used as a display for the images. The elements of the 4×4 array are numbered from 1 to 4 counterclockwise. Unlike conventional holographic stereogram techniques, in which the images are recorded by using the whole prerecorded scenes, each of the elements in this example is recorded with information from the same pixel location in every scene. Each element has its own diffraction pattern as shown in Fig. 13-19. A viewer looking into the resulting hologram from the left-hand side sees pixels #1 and #2 being turned on and pixels #3 and #4 being turned off. Similarly, a viewer looking from the right-hand side would see pixels #2 and #3 being turned on and pixels #1 and #4 being turned off.

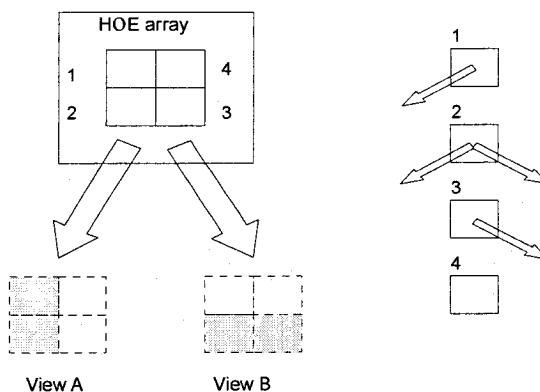


Figure 13-19. Two images projected by the HOE array along two directions. The individual diffraction pattern of each element is shown on the right.

The recording of the array is performed in a pixel-by-pixel manner. For a pixel at a given position, its recording consists of collecting all the information from the same position on all sampled images. In the 4×4 array example, there are two horizontal views at the observation plane; hence there are two segments at the projection screen for recording. To record element no.1, which has a single diffraction toward the left-hand direction, the right of the two segments on the projection screen has to be turned on in accordance with the scheme shown in Fig. 13-16. Similarly, for element no.2, both the segments on the projection screen have to be turned on in order to generate diffraction patterns along both right and left directions. The projection screen configuration for recording individual pixels is shown in Fig. 13-20.

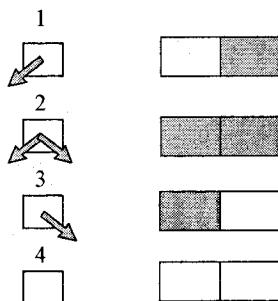


Figure 13-20. The individual diffraction patterns and the recording projection screen configurations. The (common) reference beam is not shown.

3.4.2 Factors to consider for optimal recording

(a) Problem of uniform angular multiplexing for all pixels in the HOE array

The individual pixel in an HOE array diffracts light along pre-defined spatial divisions. Similarly, an HOE array also projects images toward the same pre-defined spatial divisions. However, the angular relationships between the diffraction patterns from pixels in the array have to be carefully designed. If the mask and the projection screen are held fixed during the recording, and the film is moved in the x-direction, one obtains a diffracted output which does not completely overlap at the corresponding partitions on the observation plane; hence the reconstructed pixels from the same scene are angularly distributed, so that adjacent scenes suffer spatial misalignment. The spatial misalignment of two pixels at different locations is shown in Fig. 13-21.

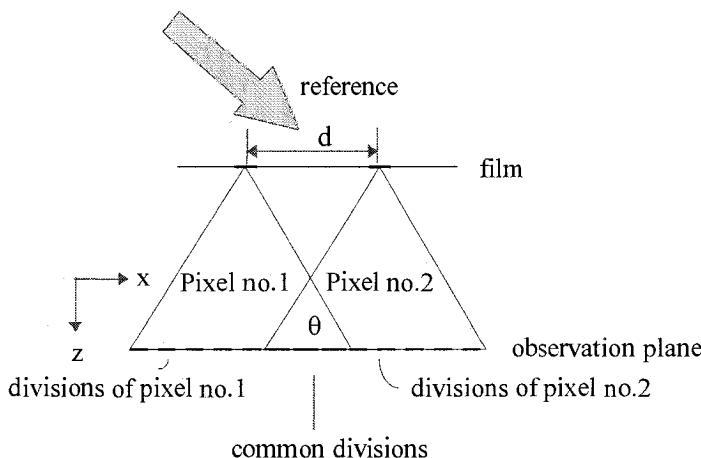


Figure 13-21. The spatial division misalignments from two pixels with uniform multiplexing divisions during reconstruction.

We note that in Fig. 13-21, the diffraction patterns of the two pixels do not overlap each other entirely. Both pixels have their own angularly multiplexed ranges. As a result, in the observation plane one can see the diffracted light from both pixels only within the common, overlapped divisions shown in the figure. The angular range of the common divisions is determined by the parameters during the recording, such as the projection

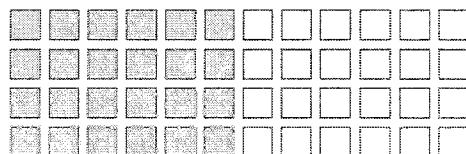
screen size in the multiplexing dimension (x-axis), the distance of the most distant pair of pixels along the x-axis in the array, and the distance from the projection screen to the film.

(b) The “running” effect

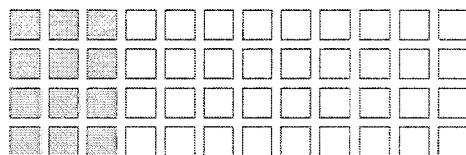
There is a “running” effect associated with the uniform angular multiplexing of all the pixels, as shown in Fig. 13-22. The “running” effect appears from the fact that the angular multiplexing divisions of each individual pixel are slightly separated from those of the neighboring pixels. When the observing point is moved along the multiplexing dimension (x-axis in Fig. 13-21), the “switching” of the multiplexed scenes may only be seen gradually. One way to show such an effect is to record an array with two fixed divisions of dark and bright (pixel no.1 in the previous example). When the array is under reconstruction, the dark and bright interface position changes as the viewing angle is varied.

(c) Removal of the “running” effect

A set of well-multiplexed scenes must possess abrupt switching characteristics, i.e., the multiplexed divisions of individual pixels have to be spatially synchronized. The divisions of pixel no.1 must overlap the divisions of pixel no. n, as shown in Fig. 13-23. If a spatial synchronization is achieved, the running effect is expected to disappear.



(a)



(b)

Figure 13-22. The “running” effect associated with fixed, uniform angular multiplexing. Shown in (a) and (b) are images observed from two different locations on the x-axis.

From Fig. 13-23, we note that the angular ranges of any two pixels are not identical to each other. Therefore, to generate perfect alignment of the diffraction patterns, the configuration of the projection screen has to be modified in order to accommodate the misalignment between the angular ranges of the pixels. The configuration of the projection screen for each pixel may be adjusted according to the desired angular multiplexing range under reconstruction.

Considering the recording setup shown in Fig. 13-24, the film is first moved along the x-axis and then is moved one step along y-axis. The film is placed on an x-y carriage which allows the film to move in both directions, controlled by the computer. The HOE array is recorded in a pixel-by-pixel manner. The light source is a 10 mW HeNe laser. The light beam is then spliced into reference and object beams. The object beam is then expanded and projected onto a liquid crystal display (LCD), which is also controlled by the computer, to configure the object beam. The configured object beam is dispersed by a diffuser thereafter to generate a broader angular range. The broader angular range allows light emerging from any point on the LCD to reach the opening on the mask.

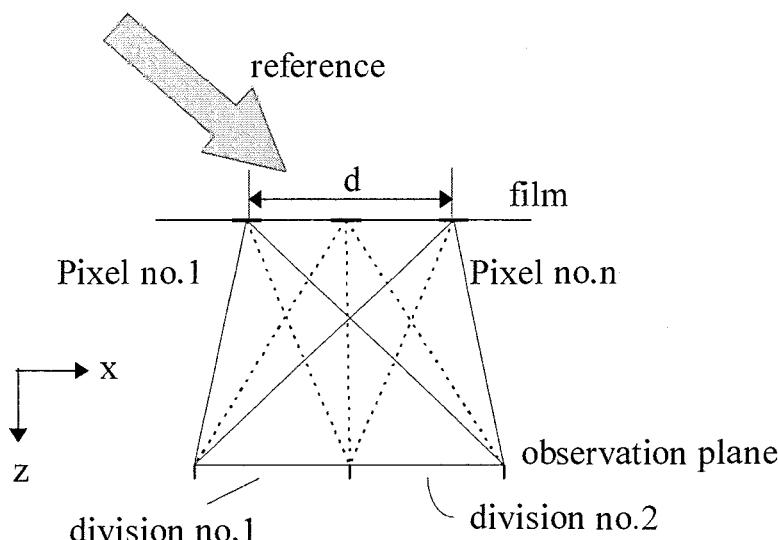


Figure 13-23. Perfectly aligned diffraction of pixels at the observation plane, where d is the (maximum) distance of the leftmost and rightmost pixels.

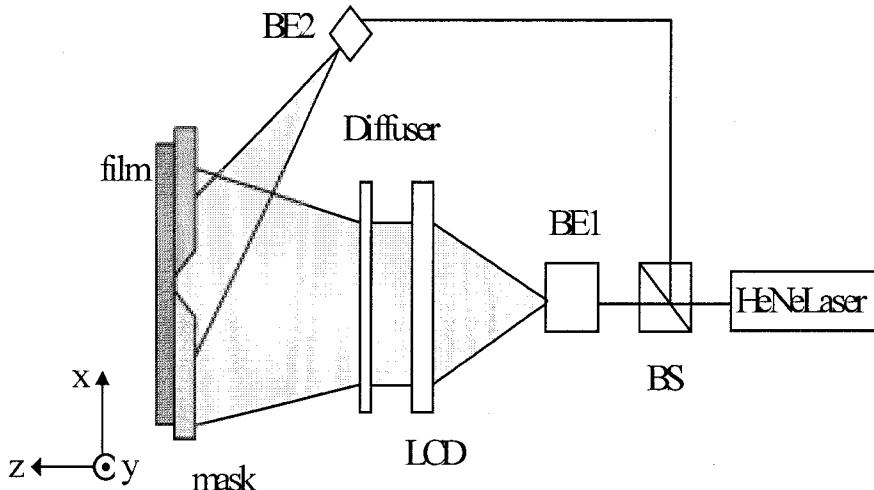


Figure 13-24. Recording schematics for the HOE array.

3.4.3 Geometrical analysis

The recording geometry is analyzed based on one-dimensional multiplexing, and may be extended to 2D multiplexing to include both x and y axes. Assuming x-axis multiplexing, the recording geometry may be depicted as in Fig. 13-25(a) and (b).

Note in Fig. 13-25(a) that the available display area of the LCD is not 100% utilized, since compromises have been made to accommodate available common multiplexing division. l_{IF} is the distance from the image plane on the LCD to the film, l_{FO} is the distance from the film to the observation plane, w_d is the viewing window width, and l_{LCD} is the width the LCD.

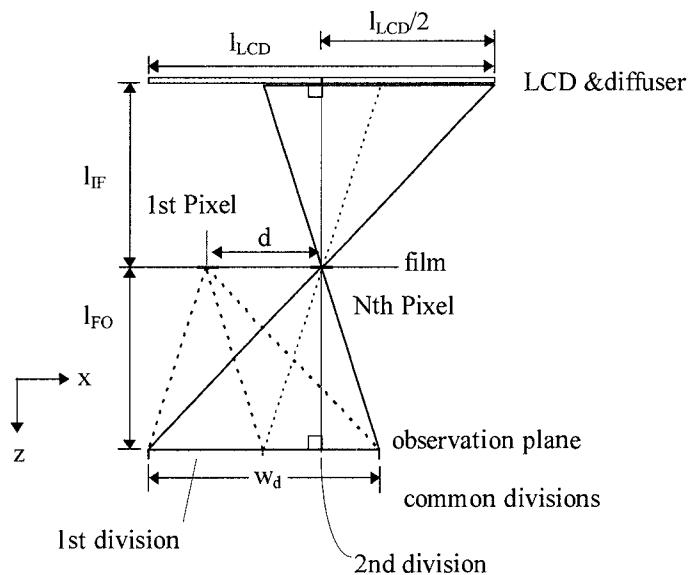


Figure 13-25(a). The recording geometry for the rightmost pixel no. n is shown.

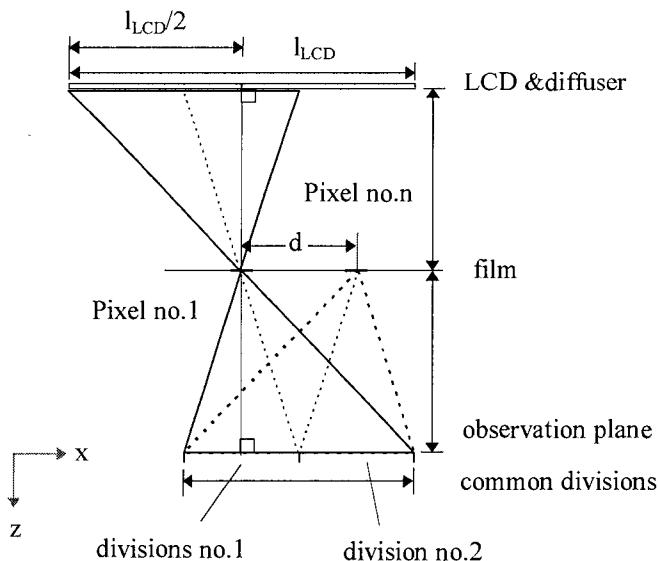


Figure 13-25(b). As in (a), but for the leftmost (1st) pixel.

The width of the common multiplexing division (i.e. the width of the viewing window), w_d , is determined by

$$w_d = l_{LCD} \frac{l_{FO}}{l_{IF}} - d. \quad (13-1)$$

The common division on the observation plane is mapped through the individual pixel to a portion of the available area on the LCD, i.e., the “window” of that particular pixel. The LCD is programmed to present a specific pattern within the window during the recording.

The boundaries of the window on the available display area have to be found and be mapped to numerical numbers of dots out of the LCD’s N dots along the multiplexing axis (640 for the LCD used in the experiment). In Fig. 13-26, the two boundary points $n_R(x)$ and $n_L(x)$ (in number of dots on the LCD) may be found for the i th pixel located at x .

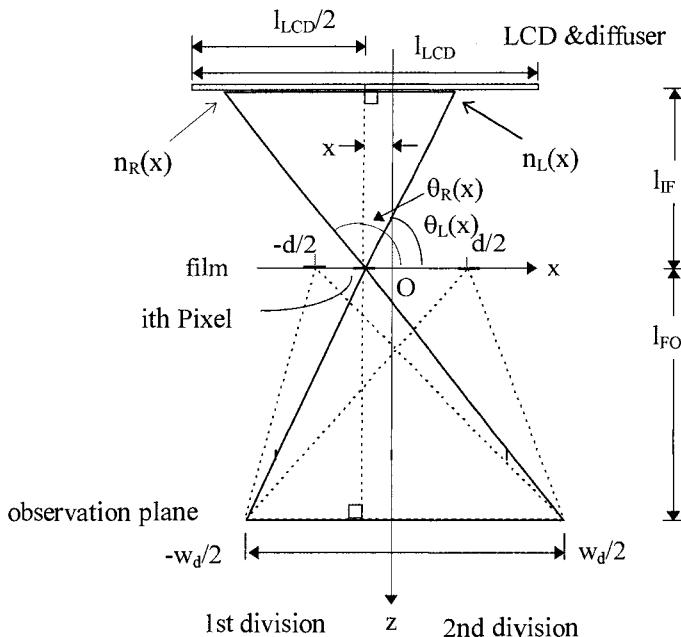


Figure 13-26. The recording geometry for the i th pixel located at x . Within the range from R to L ,the LCD displays the pattern which is to be reconstructed later during replay.

$\theta_R(x)$ and $\theta_L(x)$ are angles from the pixel to pattern edges on the projection screen, which are defined as:

$$\theta_R(x) = \pi - \tan^{-1} \frac{\frac{wd}{2} - x}{l_{FO}}, \quad (13-2a)$$

$$\theta_L(x) = \tan^{-1} \frac{\frac{wd}{2} + x}{l_{FO}}. \quad (13-2b)$$

Translating the coordinate positions to the pixels on the LCD, the pixel number of the right and left boundaries are

$$n_R = \left\lfloor \frac{N}{2} + N \frac{l_{IF}}{l_{LCD}} \cot \theta_R(x) \right\rfloor, \quad (13-3a)$$

$$n_L(x) = \left[\frac{l_{IF}(\cot \theta_R(x) + \cot \theta_L(x))}{l_{LCD}} \right] N + n_R(x), \quad (13-3b)$$

where $n_R(x)$ and $n_L(x)$ are the dot positions on the LCD to the right and left of the recording window on the projection screen, respectively, and N is the total number of pixels along the multiplexing axis (x-axis in this case). The range n_R to n_L is divided into m divisions for multiplexing m images. The two dimensional multiplexing technique applies the similar geometrical calculations, but divides the area into $(m_x \times m_y)$ chessboard-like divisions. The recording may be done with binary levels, i.e., dark and bright, or with varying gray levels.

3.4.4 Experiments

An experimental setup similar to the one in Fig. 13-24 has been implemented. The LCD used has 640 (H) by 480 (V) pixels with 16 gray levels. The recording film/plate is placed on a two dimensional controllable carriage. A personal computer is used to control the carriage movement, pattern presentation, and exposure time control. The setup is depicted in Fig. 13-27.

Several recordings have been performed to demonstrate the multiplexing capability of the POP holographic stereogram technique, including horizontal multiplexed images, horizontal and vertical multiplexed images, and multiple gray scale images. The recorded film is replayed with a reconstruction beam which covers the whole film, as in Fig. 13-28.

Horizontal Multiplexed Stereogram

A series of images are recorded and multiplexed along the horizontal axis. Twenty images from 1 through 20 are encoded in the recording patterns. The resulting stereogram will project them from right to left, as shown in Fig. 13-29.

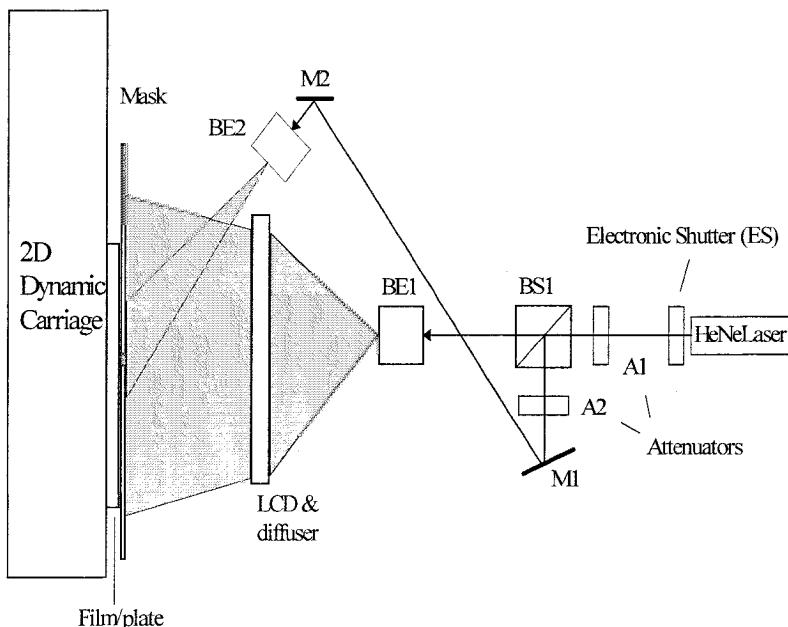


Figure 13-27. The experimental setup for recording the HOE array. The electronic shutter (ES), LCD, and the 2D dynamic carriage are controlled by a computer. The attenuators A₁ and A₂ are used to control the beam intensity ratio for recording.

In the actual experiment, of the twenty images intended to be multiplexed, three (the number 1, 2, and 3) did replay sufficiently well. The rest were clearly visible, and could be picked up by a CCD camera. Shown in Figs. 13-30(a)-(d) are several captured images from the reconstruction.

Horizontal and vertical multiplexed stereogram

One experiment to test the multiplexing capability is to multiplex the images in both vertical and horizontal directions, as shown in Fig. 13-31.

The resulting multiplexed scenes from such a stereogram are shown in Fig. 13-32(a)-(d).

Note that one may apply different gray levels to objects in the same scene using a monochrome LCD which is capable of displaying 16 gray levels.

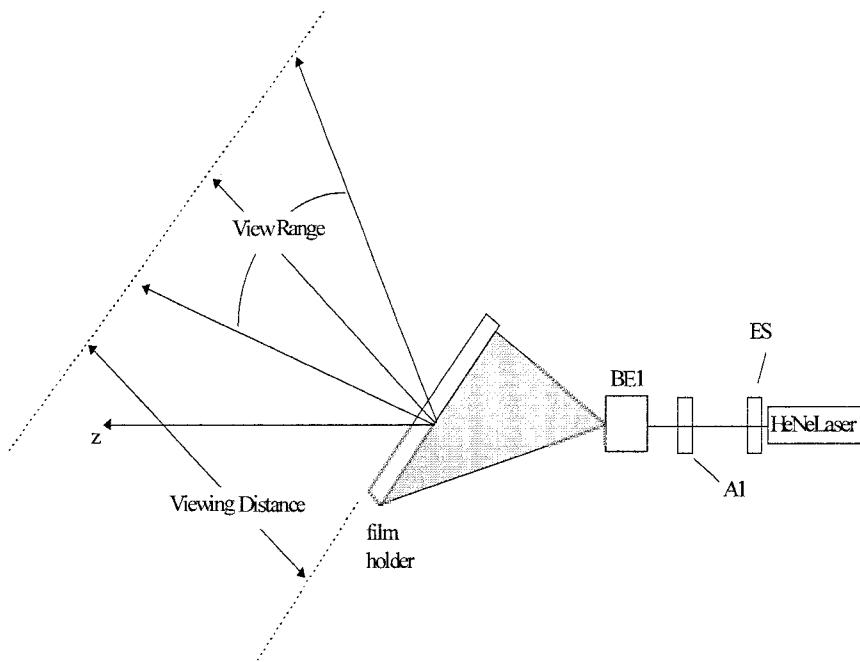


Figure 13-28. The replay setup for holographic stereograms. The film is placed on a film holder. The stereographic images are observed from a viewing distance, over the pre-defined angular range.

4. POST-1995 AND FUTURISTIC PERSPECTIVES ON 3D DISPLAY TECHNOLOGIES

The many concepts being pursued in (auto)stereoscopic display research include 3D generation from 2D scenes, autostereoscopic transformation from stereoscopic systems, and the search for new 3D media. This section briefly outlines several of these concepts; note that only the ideas are presented without any details for implementation. As pointed out by Pastoor and

Wöpking (1997). 3D display development has benefited greatly from recent advances in LCD-technology. Factors such as the emergence of new programmable multimedia signal processors, the rapid growth of high speed digital networks, and the supply of ever more computing power at low costs, will play a leading role in the development of emerging 3D technologies.

4.1 3D effect extraction from 2D scenes using a 2D-3D converter

It is reasonable to speculate that real time digital signal processing (DSP) techniques may be used in conjunction with the stereoscopic techniques described earlier to enhance 2D motion pictures with 3D views.

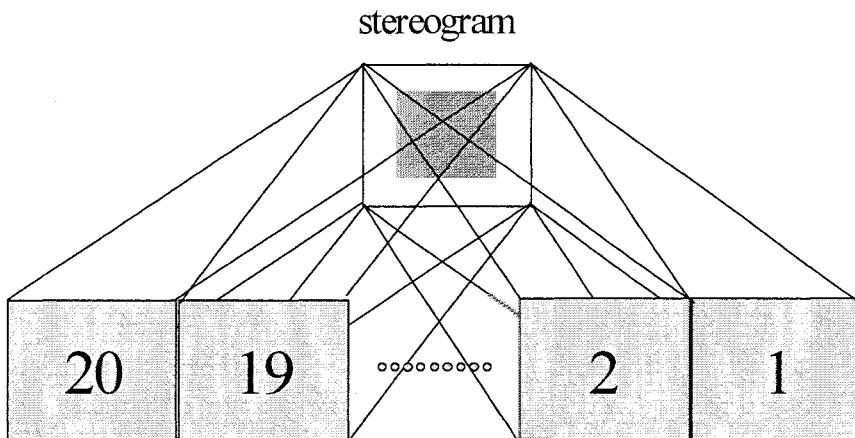


Figure 13-29. Twenty horizontally multiplexed scenes from a POP holographic stereogram.

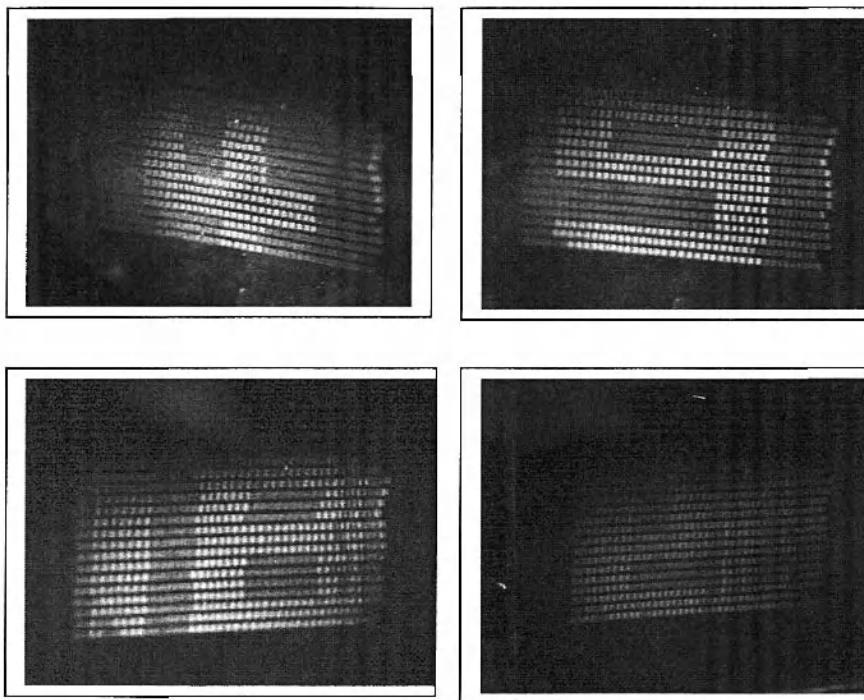


Figure 13-30 (a)-(d). Four of the twenty horizontally multiplexed images viewed from different perspectives.

The system to be developed is a video DSP system which performs image correlation, object analysis, and dynamic image presentation of the stereoscopic display. The input consists of regular 2D video frames that are stored in an image buffer, and whose correlations with previous frames are calculated and stored. The dynamic stereoscopic image presentation uses two video frame buffers. The 2D-3D converter performs a proper choice of paired stereoscopic images from the frames of the motion picture and stores them in the frame buffer.

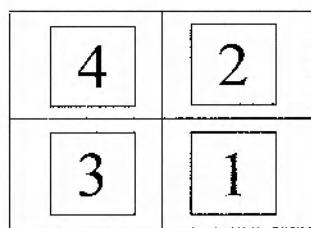


Figure 13-31. Two-dimensional multiplexed holographic stereogram (full parallax).

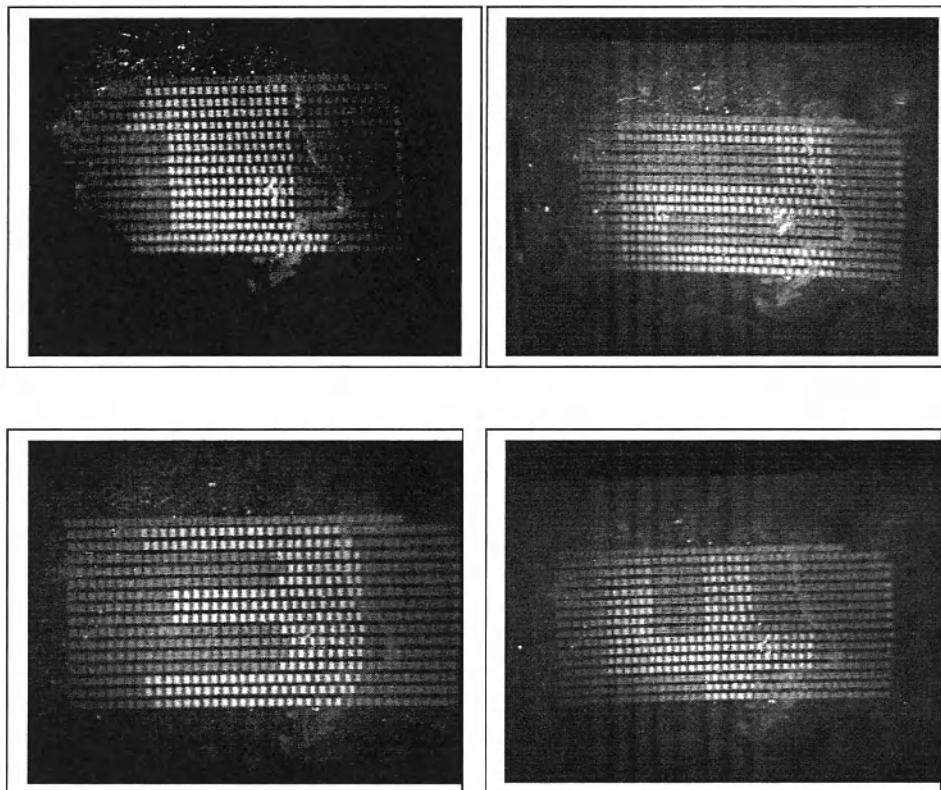


Figure 13-32(a)-(d). The multiplexed scenes from the horizontally and vertically multiplexed holographic stereogram.

In some cases, the DSP unit may generate the pseudo-3D view from the same image frame if enough information is obtained from previous input images.

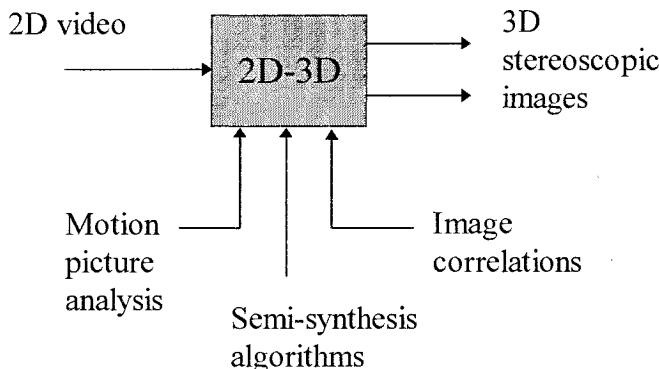


Figure 13-33. The conceptual plot for a 2D to 3D converter, in which 2D video is transformed to 3D stereoscopic images.

In Fig. 13-33, the 2D-3D converter takes a 2D video and converts it into a pair of stereoscopic images to be presented on a stereoscopic display. The converter has internal functions including motion picture analysis, image correlation, semisynthesis algorithms, and a presentation stage. The purpose for each function is listed in Table 13-2.

Table 13-2. The internal functions of a 2D-3D converter.

Motion picture analysis	<ul style="list-style-type: none"> determine which perspective to be presented with current input image determine timings to collect images into buffer determine timing to start 3D animation (no need for small objects)
Semi-synthesis algorithms	<ul style="list-style-type: none"> to prepare artificial assembly for closely related images when no close image may be matched
Image correlator	<ul style="list-style-type: none"> provides the statistical estimations for the relation between input image and buffered images

The converter generates 3D scenes only when the objects in the image frames move left or right, move (near) towards the observer, or rotate horizontally. In some situations, including far-away objects, absence of

previous buffered images (lack of information), direct stereoscopic video input, and text-only display, the converter may cease its operation.

A successful 2D-3D converter allows stereoscopic 3D displays to be used with entertainment motion pictures with real characters, and not be limited to computer generated 3D pictures (such as video games, computer animation, or medical imaging) only.

4.2 Autostereoscopic and simulated autostereoscopic 3D approaches

In general, autostereoscopic display systems have more realistic image realization than standard stereoscopic systems with only two viewing windows. Although more recently developed LCD stereoscopic systems are designed with multiple viewing windows to allow autostereoscopic display, it is achieved at the price of pixel density. There are also stereoscopic systems equipped with head- or eye-tracking capabilities to achieve the desired autostereoscopic effect (Takeda et al., 1995; Mortensen, 1996).

3D displays can generally be categorized by the technique used to channel the left and right images to the appropriate eyes (Pastoor and Wöpking, 1997). Some require optical devices close to the observer's eyes, while others have the eye-addressing techniques completely integrated into the display itself. Displays of the latter category are called autostereoscopic and are technically much more demanding than the type with viewing aids (stereoscopic displays).

A stereoscopic system, such as the shutter stereoscopic display system, may be augmented with autostereoscopic features when the scene changes with viewing angles. The head- or eye-tracking capabilities must be well defined here since the viewing angle may not coincide with the sighting angle.

The solution is to place an eye (or head) angle (location) detector in the stereoscopic display system, as depicted in Fig. 13-34. In a simulated autostereoscopic display system, the detected position is fed to the 3D display engine to generate corresponding stereoscopic images.

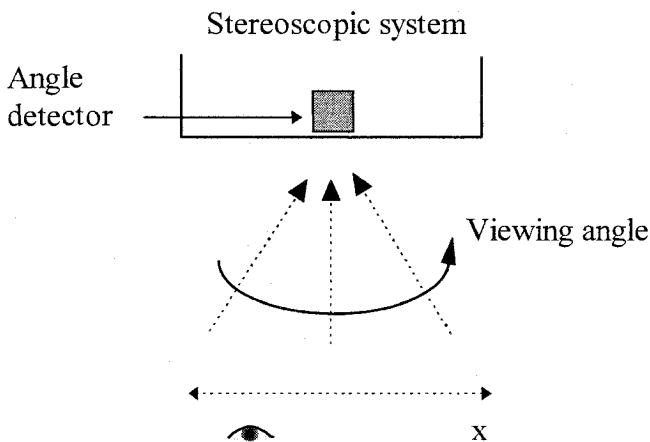


Figure 13-34. Simulated autostereoscopic approach aided by an angle detector.

The angle detection scheme may use a CCD camera or a pair of infrared transceivers as a detector to obtain the actual viewing angle of the viewer. The advantage of using the simulated approach is that it allows a smaller bandwidth system to operate as a truly autostereoscopic system. However, one disadvantage is that, usually, special glasses have to be used in order to use the stereoscopic display system, as is done in a VR system.

4.3 Stereoscopic and autostereoscopic displays

Stereoscopic displays have two different perspective views that are generated (quasi-) simultaneously. The waves of light entering the eyes may have their origin on a common, fixed or movable (gaze-controlled) image plane. Several multiplexing methods have been proposed to carry the optical signals to the appropriate eyes. It is possible to adapt the image content to the current head position (eye-point dependent perspective; e.g., via corresponding camera movements or signal processing). On the other hand, in autostereoscopic systems, the only exploitable constraint for addressing the left and right eye is the fact that they occupy different points in space. The waves of light forming the 3-D image may originate from fixed or gaze-controlled image planes. In both cases, direction-multiplex is the only way to channel the information of the left and right views into the appropriate eyes. Compared to stereoscopic techniques, it is possible to multiplex more than two views at a time. Thus, individual (eye-point depending) perspective views can be delivered to different observers. The volumetric and the electro-holographic approaches produce 3-D images where the effective origin of the

waves entering the observer's eyes match with the apparent spatial position of the corresponding image points. Thus, the fundamental mechanisms of spatial vision are perfectly supported (in principle there is no difference to natural viewing conditions).

4.3.1 Autostereoscopic techniques

The developing field of autostereoscopic displays is marked by a variety of competitive approaches. The following discussion focuses on only a few fundamental trends. The techniques currently recognized in the research domain are: electro-holography, volumetric displays and direction-multiplexed displays.

4.3.1.1 Electro-holography

As is well known, holography involves recording and reproducing the properties of light waves - amplitude (luminance), wavelength (chroma) and phase differences - almost to perfection, thereby enabling a very close approximation of an ideal free viewing 3D technique. Holographic recording requires coherent light to illuminate both the scene and the camera target (used without front lenses). For replay, the recorded interference pattern must again be illuminated with coherent light. Video-based holographic techniques are still in their infancy, although they have received much attention over the past decade or longer. Organised by TAO (Telecommunications Advancement Organization), several Japanese research institutes are working towards adapting the principle of holography to an LCD-based video electronics environment (Honda, 1995). However, the spatial resolution of today's LC-panels is a serious bottleneck. Partitioning the hologram among several LC-panels and reassembling the image with optical beam combiners may alleviate some of the problems. It is yet an open question how to store and transmit the enormous amount of data contained in a hologram. The source rate is estimated to exceed 10^{12} bit/sec (Sato, 1995). Specific data compression methods are required (Yoshikawa, 1993). Currently, the scope of this approach is limited to very small and coarse monochromatic holograms (width of field 1 cm) (Hashimoto and Morokawa, 1995).

4.3.1.2 Volumetric displays

In volumetric displays, image points are projected to definite loci in a physical volume of space where they appear either on a real surface, or in translucent (aerial) images forming a stack of distinct depth planes. With the first type of system, a self-luminous or light reflecting medium is used which either occupies the volume permanently or sweeps it out periodically. Implementation methods range from the utilization of fluorescent gas (with external excitation through intersecting rays of infrared light) over rotating or linearly moved LED-panels to specially shaped rotating projection screens. Rotating screens in the form of a disc, an Archimedian spiral or a helix wound around the vertical axis, have been implemented (Yamada et al., 1993).

Of the types described above (those with a real collecting surface), the helical design has reached maturity. The most elaborate equipment uses a double-helix filling an 91-cm-diameter by 46-cm-high volume at 10 revolutions per second with a maximum of 40,000 color pixels per frame (or 120,000 pixels in the color primaries) (Soltan et al., 1995). Viewers can walk around the display and observe the imaged objects from different angles.

4.3.1.3 Direction-multiplexed displays (including partial pixel, and autostereoscopic)

In such systems, optical effects such as diffraction, refraction, reflection and occlusion are utilized in order to direct the light emitted by pixels of different perspective views exclusively to the appropriate eye. There are four basic sub-categories here: diffraction/DOE/HOE-based, refraction-based, reflection-based, and occlusion-based. These are described briefly next.

(a) *Diffraction-based approaches*

In the diffractive-optical-elements (DOE) approach, corresponding pixels of adjacent perspective views are grouped in arrays of "partial pixels" (such as the ICVision Display due to the Jones group (Jones et al., 1995), and the 3D Grating Image Display due to the Toda group Toda et al., 1995).

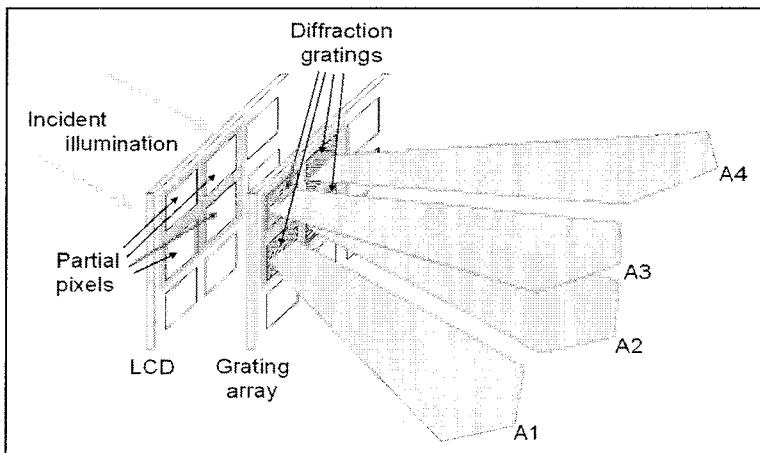


Figure 13-35. Principle of a DOE 3-D display (Mortensen, 1996; Jones et al., 1995).

Small diffraction gratings placed in front of each partial pixel direct the incident light to the respective image's viewing area (first order diffraction; A1 to A4 in Fig. 13-35). Prototypes tested yield images of less than 1.5 inches in diameter. More advanced concepts provide for the integration of image modulation and diffraction of light within a single, high-resolution spatial light modulator (Schulze, 1995). Alternatively, in the holographic optical elements (HOE) model, the properties of conventional optical elements (e.g. lenses) are replaced by holographic methods. Thus, an HOE contains no image information, but serves to diffract rays of light modulated elsewhere.

In a recent prototype display, the HOE is an integral part of the light box of a modified LCD (Trayner and Orr, 1996). It consists of the hologram of an even diffuse plane which is rastered so as to direct the light of alternating lines to specified viewing zones. Outside the zone for stereoscopic viewing, both eyes receive a 2-D view. The stereo zone can be made to follow the observer's head movement by moving the light source. Making the above HOE takes two exposures of a hologram with the same illumination setup. Between the exposures, the object, which in this case is a diffuse plane, is shifted horizontally by its own width. For each exposure, parts of the hologram corresponding to the odd or even image lines are occluded.

(b) *Refraction-based approaches*

Several display concepts have been proposed based on conventional, refractive optical elements (such as picture-sized large lenses or small lenslets) to address the observer's eyes. In integral imaging, the spatial image is composed of multiple tiny 2-D images of the same scene, captured with a very large number of small convex lenslets (fly's-eye lens sheet). Each lenslet captures the scene from a slightly different perspective. A lens sheet of the same kind is used for display (between capture and replay, the image has to be inverted for orthoscopic depth rendition, see Trayner and Orr, 1996). As the image plane is positioned into the focal plane of the lenslets, the light from each image point is emitted into the viewing zone as a beam of parallel rays at a specific direction. Therefore, the observer perceives different compositions of image points at different points of view. The individual lenslets must be very small, since each pixel is spread to the lens diameter at replay, and the image formed behind each lenslet should be as complete and detailed as possible. As a consequence, the display must provide an extremely high spatial resolution. Alternatively, in lenticular imaging, arrays of vertically oriented cylindrical lenslets are used that may be considered a one-dimensional version of integral techniques. The light from each image point is emitted at a specific direction in the horizontal plane, but non-selectively in the vertical plane. Therefore, changes of perspective in accordance with vertical head movements cannot be achieved by optical means (but by headtracking and computational image processing). Direct-view and projection-type 3-D displays have been implemented based on lenticular techniques (Okoshi, 1976).

(c) *Reflection-based approaches*

In this approach, a retro-reflective screen is used for direction multiplexing. Retro-reflective means that the incident rays of light are reflected only into their original direction (Börner, 1993). In a recent prototype (Xenotech, Australia, 1996), dual video projectors are mounted on a laterally movable stage (Fig. 13-36). The screen reflects the two images through a large half mirror to the observer's eyes. The system locates the current head position and adjusts the position of the projectors and the angle of the half-mirror accordingly.

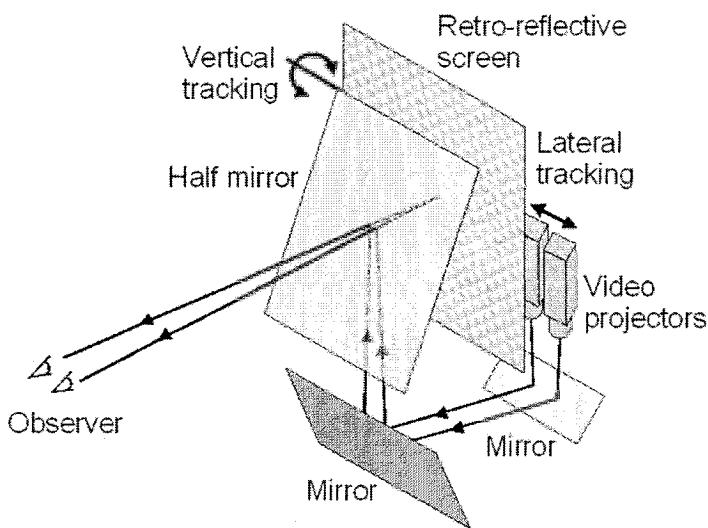


Figure 13-36. Schematic view of the Xenotech display (Pastoor and Wöpking, 1997; Xenotech, Australia, 1996).

A more recent version of a reflection-based 3D display system is one developed and reported by Yan *et al.* in 2004 (Yan *et al.*, 2004). It is known that computer generated 3D graphics or regular camera images provide monocular cues, while stereoscopic systems generate binocular cues. Additionally, head tracking and holography fulfill motion parallax requirements. Following the ICV project (which used a parallel pixel architecture) described earlier, in which monochromatic displays of dimension 3 cm × 2 cm and 16 views, and color displays of the same dimension and 4 views, were demonstrated, the current group (including researchers from the ICV group) has reported on the design, fabrication and testing of 3D display systems based on a micromirror array. In such a system, two spatial light modulators (SLMs) are used, of which the first serves as the image source on which the time-sequential views are painted, while the second consists of a micromirror array which redirects the views into appropriate viewing zones. A crucial requirement is to ensure that the refreshing of the first SLM and the redirection of the viewing zones (scanning the micromirrors) be synchronized. It is suggested that the use of micromirrors provides the following advantages: (a) being reflective, micromirrors are achromatic (i.e., free of wavelength-dependent aberrations); (b) a variety of displays may be used as image sources; and (c) time multiplexing as well as space multiplexing may be applied simultaneously,

allowing the optimization of viewing zones. To accommodate realizable micromirror arrays, the authors have made use of a ray-tracing program called BeamThree, together with microelectromechanical system (MEMS) design and fabrication.

(d) *Occlusion-based approaches*

In all these approaches, parts of the image are hidden from the one eye but visible for the other eye due to parallax effects. The technical solutions differ in the number of viewing slits (ranging from a dense grid to a single vertical slit), in presentation mode (time-sequential versus stationary) and in whether the opaque barriers are placed in front of or behind the image screen (parallax barrier versus parallax illumination techniques). Basically, any of the lenticular display designs can also be implemented with parallax effects, replacing the raster of cylindrical lenslets by a raster of vertical slit openings. Parallax barriers are much easier to manufacture than lenticular screens (e.g. by printing or electro-optical methods). They can be moved electronically for headtracking and time sequential presentation (moving slit technique). Moreover, the dark barriers increase the maximal contrast under ambient illumination (black-matrix effect). In one example, classified under the rubric of barrier-grid displays, Sanyo has optimized the barrier-grid design for LCD-based direct-view displays (Hamagishi et al., 1995). One barrier is in front of the LCD and an additional one is placed between the LCD panel and the backlight case. The additional barrier has a reflective coating facing the backlight. It serves to exploit the light which would otherwise be blocked by the black matrix of the LCD panel, thereby improving the lighting efficiency (reported to be up to a factor of 1.4).

4.4 Optical fiber 3D displays

Optical fibers nowadays are widely used in optical communications, sensors, and medical imaging applications. In principle, the primitive optical fiber for visible wavelengths is considered a light pipe that conducts light waves. A volumetric display was developed using optical fibers to transport light from the source to a 3D fluorescent resin (vixel) array (MacFarlane, 1994). In medical imaging applications, light pipes also transmit images from the inside of a patient, and allow physicians to make visual examinations of the disease. Light pipes are now widely accepted in the medical community for such imaging applications. Note that a single light pipe may only transmit light intensity profile from one end of the pipe to the

other, and not generate a 3D image. However, if a light pipe array arrangement, as depicted in Fig. 13-37(a), is constructed along the lines of the concept of partial object pixel technique, the total construction of the light output from these light pipes may form a 3D image.

Each light pipe may be fed with spatially modulated light from an LCD array. A launching scheme is depicted in Fig. 13-37(b). The number of viewing windows is determined by the number of pixels on the feeding LCD as depicted in Fig. 13-38, since the image on the LCD would be transmitted to the other end of the light pipe. Similarly, the arrangement of pixel locations and shape also defines that of the viewing windows.

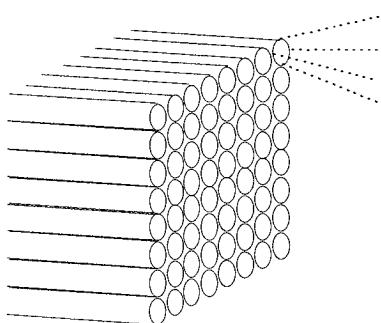


Figure 13-37(a). A light pipe array to deliver 3D pictures.

The difficulties for manufacturing such a device are the alignment of individual light pipes as well as the assembly of a large number of light pipes.

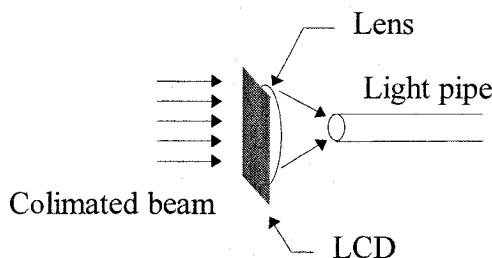


Figure 13-37(b). The launching schematic for a light pipe.

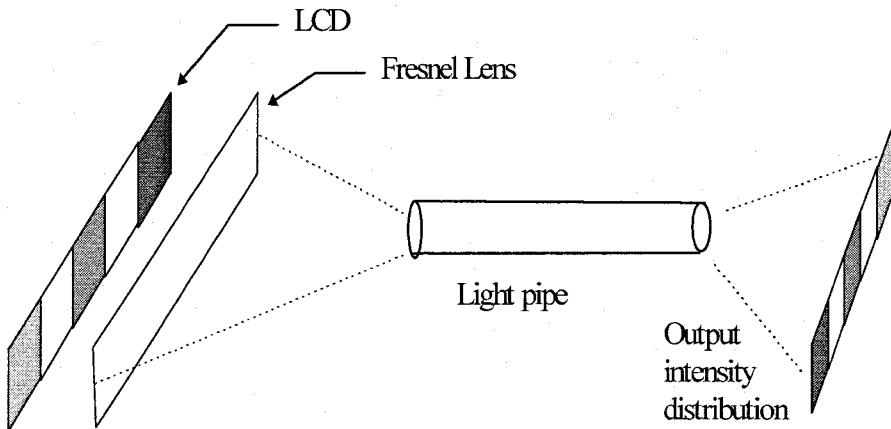


Figure 13-38. The configuration for a horizontal parallax 3D display system.

REFERENCES

- Bains, S., 1996, Low-cost autostereoscopic screen creates 3-D views, *Laser Focus World*, 26-28.
- Boff, K.R., Kaufman, L., and Thomas, J.P., 1986, *Handbook of perception and human performance*, Vol.I, Sensory Processes and Perception, pp.10, 11, 23, Wiley&Sons.
- Börner, R., 1993, Autostereoscopic 3D-imaging by front and rear projection and on flat panel displays, *Displays* 14, 39-46.
- Chen, S.-T., 1996, *Analytical Investigations of Dynamic and Static Optical Gratings and Some Modern Applications*. Ph.D. dissertation, Binghamton University, SUNY.
- Chen, S.-T., and Chatterjee, M.R., 1995, Computer generated spatially multiplexed message display using a holographic optical element array, *Optical Society of America Annual Meeting*, Portland, Oregon.
- Chen, S.-T., and Chatterjee, M.R., 1998, Implementation of a spatially multiplexed pixelated three-dimensional display by use of a holographic optical element array, *Appl. Opt.* 37, 7504-7513.
- Cross, L., 1977, Multiplex Holography, *SPIE Annual Meeting*, San Diego, CA.
- DeMeis, R., 1996, Parallax images provide 3-D color display, *Laser Focus World*, 39-40.
- Dodgson, N.A., 1996, Analysis of the viewing zone of the Cambridge autostereoscopic display, *Appl. Opt.* 35, 1705-1710.
- Ernst, H., Petzold, J., Larice, R., and Breide, S., 1996, Mixing of computer graphics and high quality stereographic video, *IEEE Trans. on Consumer Elect.* 42, 795-799.

- Hamagishi, G., Sakata, M., Yamashita, A., Mashitani, K., Nakayama, E., Kishimoto, S., and Kanatani, K., 1995, New stereoscopic LC displays without special glasses, *Asia Display '95*, 791-794.
- Hashimoto, N., and Morokawa, S., 1995, Motion-picture holography using liquid-crystal television spatial light modulators, *SID 95 Digest*, 847-850.
- Honda, T., 1995, Dynamic holographic 3D display using LCD, *Asia Display '95*, 777-780.
- Iizuka, K., 1985, *Engineering Optics*, Chapter 8, Spring-Verlag, Berlin, Heidelberg.
- Joachimowicz, K.E., and Gold, R.S., 1990, Stereoscopic (3-D) projection display using polarized color multiplexing, *Opt. Eng.* **29**, 838-842.
- Jones, M.W., Nordin, G.P., Kulick, J.H., Lindquist, R.G., and Kowal, S.T., 1995, A liquid crystal display based implementation of a real-time ICVision holographic stereogram display, *Proc. SPIE* **2406**, 154-164.
- Jones, M.W., Nordin, G.P., Kulick, J.H., Lindquist, R.G., and Kowal, S.T., 1995, Real-time three-dimensional display based on the partial pixel architecture, *Opt. Lett.* **20**, 1418-20.
- Khoo, I., 1995, Holographic grating formation in dye- or fullerene-C₆₀-doped liquid crystals, *Opt. Photon. News*, p.29.
- King, M.C., 1968, Multiple exposure hologram recording of a 3-D image with 360° view, *Appl. Opt.* **7**, 1641-1642.
- King, M.C., Noll, A.M., and Berry, D.H., 1969, A new approach to computer generated holography, *Appl. Opt.* **9**, 471-475.
- Kulick, J.H., Nordin, G.P., Parker, A., Kowal, S.T., Lindquist, R.G., Jones, M., and Nasiatka, P., 1995, Partial Pixels: a three-dimensional diffractive display architecture, *J. Opt. Soc. Am. A* **12**, 73-83.
- MacFarlane, D.L., 1994, Volumetric three-dimensional display, *App. Opt.* **33**, 7453-7457.
- Meerholz, K., 1994, A photorefractive polymer with high gain and diffraction near 100%, *Nature* **371**, 497.
- Mortensen, P., 1996, Stereoscopic display watches the viewer, *Laser Focus World*, 61-3.
- Murillo-Mora, L.M., Okada, K., Honda, T., and Tsujiuchi, J., 1995, Color Conical Holographic Stereogram, *Opt. Eng.* **34**, 814.
- Nordin, G.P., Kulick, J.H., Lindquist, R.G., Nasiatka, P.J., Jones, M.W., Friends, M., and Kowal, S.T., 1995, Liquid crystal-on-silicon implementation of the partial pixel three-dimensional display architecture, *Appl. Opt.* **34**, 3756-63.
- Okoshi, T., 1976, in *Three-Dimensional Imaging Techniques*. Academic Press, New York.
- Palais, J.C., and Miller, M.E., 1996, Holographic movies, *Opt. Eng.* **35**, 2578-2582.
- Pastoor, S., and Wöpkink, M., 1997, 3-D Displays: A review of current technologies, *Displays* **17**, 100-110. Also available online at <http://www.dgp.toronto.edu/>.
- Sato, K., 1995, Development of 3-D TV by electro-holography, *Proc. 2nd Int. Display Workshop*, Hamamatsu, 73-76.
- Schulze, E., 1995, Synthesis of moving holographic stereograms with high-resolution spatial light modulators, *Proc. SPIE* **2406**, 124-126.
- Soltan, P., Trias, J., Dahlke, W., Lasher, M., and McDonald, M., 1995, Laser-based 3D volumetric display system - Second generation, in K. Morgan, et.al. (Eds.), *Interactive Technology and the New Paradigm for Healthcare*, IOS Press and Ohmsha, 349-357.
- Spierings W., and van Nuland, E., 1992, Development of an office holoprinter II, *SPIE* **1667**, *Practical Holography VI*, 52-62.
- St.Hilaire, P., Benton, S.A., Lucente, M., and Hubel, P.M., 1992, Color images with the MIT holographic video display, *SPIE* **1667**, *Practical Holography VI*, 73-84.
- St. Hilaire, P., Benton, S.A., and Lucente, M., 1992, Synthetic aperture holography: a novel approach to three-dimensional displays, *J. Opt. Soc. Am. A* **9**, 1969-77.

- Takahashi, S., Toda, T., and Iwata, F., 1991, Three dimensional grating images, in *Practical Holography V*, S.A. Benton, ed., Proc. SPIE **1461**, 199-205.
- Takeda, T., Fukui, Y., Hashimoto, K., and Hiruma, N., 1995, Three-dimensional visual stimulator, Appl. Opt. **34**, 732-38.
- Toda, T., Takahashi, S., and Iwata, F., 1995, 3D video system using grating image, Proc. SPIE **2406**, 191-198.
- Travis, A.R., 1990, Autostereoscopic 3-D display, Appl. Opt. **29**, 4341-43.
- Trayner, D., and Orr, E., 1996, Autostereoscopic display using holographic optical elements, Proc. SPIE **2653**, 65-74.
- Xenotech Australia, 1996, *The Xenotech 3-D auto stereoscopic display system*. Technical Information. Suite 1, 41, Walters Drive, Osborne Park, Western Australia 6017.
- Yamada, H., Akiyama, K., Muraoka, K., and Yamaguchi, Y., 1993, The comparison of three kinds of screens for a volume scanning type 3D display, TAO First Int.Symp. 1993 (Telecom. Advancement Org. of Japan, 2-31-19, Shiba, Minato-ku, Tokyo 105).
- Yan, J., Kowal, S.T., Cho, H.J., Ahn, C.H., Nordin, G.P., and Kulick, J.H., 2004, Autostereoscopic three-dimensional display based on a micromirror array, Appl. Opt. **43**, 3686-3696.
- Yaroslavski, L.P., and Merzlyakov, N.S., 1980, *Methods of digital holography*, Consultants Bureau, New York.
- Yoshikawa, H., 1993, Digital holographic signal processing, TAO First Int.Symp., 1993 (Telecom. Advancement Org. of Japan, 2-31-19, Shiba, Minato-ku, Tokyo 105, Japan).

Index

A

aberration
chromatic 193
spherical 183
acousto-optics 73
anamorphism 179
astigmatism 179

B

birefringence 131

C

Carrier 13
cross-square wave
CGH 1,17,235,257,379
cell-oriented 22
cosine 19
cross-square wave 21
cylindrical 268
Fouier 40
Fraunhofer 41
Fresnel 40
image plane 36
near Fourier 40
near field 39
perfect 18

point-oriented 27
square-wave 19
chirp-Gaussian function 287
Comb function 6
correlator
3D 367
joint-transform 367
convolution 81,118

D

defocus aberration 175
deformation measurement 163
depth extraction 359
diffraction
Bragg 78
theory 101
Dirac delta function 5
digital holography 51
comparative 220
multi-wavelength
phase imaging 64
phase shifting 113,145
digital mirror devices (DMDs)
219, 296
display
3-D 74,274,293,333,379

- 3-D projection** 300
autostereoscopic 301,415
color 48
color-coded stereoscopic 383
direction-multiplexed 417
Fresnel lens-based projection
 322
holographic projection 326
liquid crystal 218,293,382
optical fiber 3D 411
polarized stereoscopic 384
polarized stereoscopic
 projection 302
realtime 49
spatially multiplexed
 390
spatially multiplexed pixilated
 3D message 393
stereoscopic 415
synthetic aperture
 3D 392
time-multiplexed
 385
time-sequential projection 314
time-sequential multiview
 projection 316
virtual reality stereoscopic 391
volumetric 417
- F**
- FFT** 3
Fourier series 3
Fouier optics 57
Fourier transform 2
 Discrete 3
 forward 2
 inverse 2
Fresnel diffraction 53
Fresnel transform 106
fringe printer 252
Fresnel zone plate (FZP) 77,276
 time-depedent 276
- G**
- Gaussian function** 5
Green function 100
- H**
- Helmholtz equation** 101
heterodyne frequency 277
heterodyning 113
HDTV 294,382
holocube 325
hologram 1
 apodization 115
 binary computer-generated 16
 Bragg acoustical 78
 cell-oriented 17
 complex 281
 cosine-Fresnel zone plate 279
 computer-generated
 1,73,235,257
 computer-generated rainbow
 238
 Gabor 16
 Fresnel 77,85
 HPO- 282
 HPO-cosine 82
 HPO-sine 282
 Hybrid 248
 interferometric 43
 interferometric display 45
 off-axis 16
 point-oriented 17
 sampling 116
 sine-Fresnel zone plate 279
 three-color
hologram synthesis 259
holography
 electro- 407
 electronic 274,281
 electronic
 horizontal-parallax-only 282
 computer-generated 15,394

- color 152
digital 51,95,173,214,281
digital interference 67,69
HPO (horizontal-parallax-only)
 274
HPO-electronic 274
in-line 111
multiple wavelength digital
 186
phase-shifting 55,281
phase-shifting digital 145
off-axis 55,111
off-line 11
optical scanning 73,273
single-sideband 16
- I**
- image
 false 15
 modulation 15
 twin 19,112,281
information reduction 47
integral imaging 319,333,338,347
interferogram 224
interferometry
 digital holographic 215
 digital shearing 213
 digital speckle 215
 electronic speckle 99
 electronic speckle pattern
 155,215
 holographic 155
 phase-shifting
 shearing 217
 speckle 146
- K**
- kinoform 16
- M**
- microelectromechanical
systems (MEMS) 96,198,203
- microscopy 51,58,74,95,145
 3-D 74,149
 digital holographic 95,98
 Fourier phase 127
 Hilbert phase 127
 multi-photon 137
 short coherence holographic
 129
modulation
 carrier 9
 cosine-carrier 10
 pulse-code 11
 square-wave carrier 11
modulator
 acousto-optic 113,280
multiplexing 48
 frequency-division 13
 space-division 13
- N**
- numerical aperture 95
- P**
- plasma display panel (PDP), 293
phase clipping 31
phase nonlinearities 29
phase mismatching 31
point spread function 53
procedure
 Gerchberg-Saxton 44
propagation 33
 ping-pong 36
projection display systems 296
- Q**
- quantization 26
- R**
- reconstruction
 acoustooptical 73
 algorithm 121
 numerical 55,183

optical 217
wavefront 108,118
redundancy 42

S

sampling 8, 26, 27
Nyquist 9
theorem 8
scanning 126
simulated annealing 44
spatial light modulator 251,281
stereogram
computer-generated
holographic 390
horizontal multiplexed 401
rainbow holographic 245
surface contouring 159

T

tele-metrology 213

V

vibration analysis 165

W

wave
plane 109,276
spherical 109,276

Z

Zernike phase contrast 62
zero-padding method (ZPM) 185