

Reconstructing Specimens Using DIC Microscope Images

Farhana Kagalwala and Takeo Kanade, *Fellow, IEEE*

Abstract—Differential interference contrast (DIC) microscopy is a powerful visualization tool used to study live biological cells. Its use, however, has been limited to qualitative observations. The inherent nonlinear relationship between the object properties and the image intensity makes quantitative analysis difficult. Toward quantitatively measuring optical properties of objects from DIC images, we develop a method to reconstruct the specimen's optical properties over a three-dimensional (3-D) volume. The method is a nonlinear optimization which uses hierarchical representations of the specimen and data. As a necessary tool, we have developed and validated a computational model for the DIC image formation process. We test our algorithm by reconstructing the optical properties of known specimens.

Index Terms—Differential interference contrast microscopy, hierarchical reconstruction, iterative parameter estimation, nonlinear optimization.

I. INTRODUCTION

THE NOMARSKI differential interference contrast (DIC) microscope is the preferred method for visualizing live biological specimens. The DIC microscope is an interferometer, and therefore, the refractive structure of the specimen is made visible. In biological research, live, transparent cells can be imaged with this microscope modality. Three-dimensional (3-D) structure can be visualized by optically-sectioning¹ through the specimen. To date, however, biologists only qualitatively assess DIC images of cell specimens. Quantitative microscopy methods, such as computational optical sectioning microscopy (COSM), have been restricted to linear microscopy modalities. [5] The inherent nonlinearities in the DIC image formation process have hindered past attempts at quantitative analysis. In this paper, we describe a method to reconstruct specimens imaged with DIC microscopy.

DIC microscopy offers several advantages over other contrast-generating optical systems. In DIC the pupil is unobstructed, and therefore transverse² and axial resolution exceeds that in Zernike phase contrast. Consequently, thick specimens with 3-D features are better resolved. Unlike some fluorescence methods, no dyes are injected and therefore, live specimens are not adversely affected. Finally, unlike confocal scanning methods which have slow rates of acquisition, an entire stack of

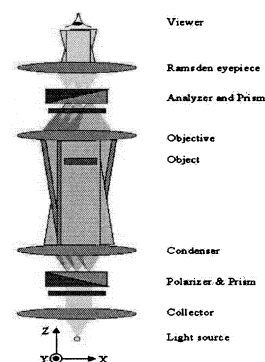


Fig. 1. DIC optical components: Regular brightfield microscope components such as a light source, collector, condenser, objective, and eyepiece are supplemented with a pair of polarizer-prism set. Three light paths are shown to illustrate conjugate planes of reference. Optical elements, spacing, and the incident light angles are not to scale.

optically sectioned images can be acquired within a minute. In the case of mobile cells, the short acquisition time minimizes distortions between optically sectioned slices.

Looking through the eyepiece of the DIC microscope, an observer sees a shadow cast image which deceptively indicates 3-D structure. Actually, the image is the differential of the optical path length introduced by the object into the propagating light wave. The differential is along a particular direction, in the transverse plane, called the shear direction. In addition, each image contains both in-focus and out of focus information. Therefore, the challenge in DIC microscopy remains to interpret image features belonging to in-focus object properties correctly.

II. DIC MICROSCOPY BACKGROUND

The DIC microscope, is essentially a brightfield microscope with a polarizer-analyzer pair and two prisms. [Fig. 1] As in standard brightfield optics, light from a lamp is collimated by a collector and a condenser lens combination. In DIC, however, a polarizer and a Wollaston prism is inserted between the collector and condenser lens. Moreover, the prism is positioned with respect to the back focal plane of the condenser. The described setup produces two mutually coherent, polarized beams. Each electric field is polarized perpendicularly with respect to the other. In addition, the wavefronts impinging on the object are differentially translated with respect to each other. In front of the objective lens, an analyzer and Nomarski prism are inserted and aligned with the front focal plane of the objective. The Wollaston prism behind the condenser introduces a linear phase gradient across the two fields emerging from the condenser. The

Manuscript received May 22, 2001; revised January 5, 2003. This paper was recommended by Guest Editor N. Bourbakis.

The authors are with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213 USA (e-mail: farhana@alummi.cmu.edu, tk@cs.cmu.edu).

Digital Object Identifier 10.1109/TSMCB.2003.816924

¹For each image in an optically sectioned set, the optical elements are configured to focus at a particular object distance.

²Transverse planes are perpendicular to the optical axis of the microscope.

Nomarski prism in front of the objective lens compensates for this linear phase gradient. Therefore, the combined action of the two prisms results in a constant phase bias between the two perpendicularly polarized fields. This combination produces a steady pattern of interference between the two beams which can then be detected by a CCD camera or the human eye.

Mathematically, the DIC imaging process can be summarized by the following set of equations. First, consider a coherent field represented by

$$U_{ill}(\vec{x}) = A \exp[-i\phi_{ill}(\vec{x})]. \quad (1)$$

where $\vec{x} = (x, y, z)$ are spatial variables in the coordinate space of the object with the z -axis coincident with the optical axis. $\phi(\vec{x})$ is the phase function. The action of the two prisms can be represented as an aggregate by $U_{ill}^1(\vec{x})$ and $U_{ill}^2(\vec{x})$, where

$$U_{ill}^1(\vec{x}) = A \exp[-i[\phi_{ill}(\vec{x} + \vec{v}_s) + \Delta\phi]] \quad (2)$$

$$U_{ill}^2(\vec{x}) = A \exp[-i\phi_{ill}(\vec{x} - \vec{v}_s)]. \quad (3)$$

In the above equations, $\Delta\phi$ is the constant phase bias and $2\vec{v}_s$ is the shear vector. After a phase transformation, $\phi_{obj}(\vec{x})$, due to the object, the wavefronts

$$U_{obj}^1(\vec{x}) = A \exp[-i[\phi_{obj}(\vec{x} + \vec{v}_s) + \Delta\phi]] \quad (4)$$

$$U_{obj}^2(\vec{x}) = A \exp[-i\phi_{obj}(\vec{x} - \vec{v}_s)]. \quad (5)$$

contain the object information. The field in the image space is (shown in equation at the bottom of the page) where $\vec{x}_{img} = (x_{img}, y_{img}, z_{img})$ are spatial variables in the coordinate space of the image with the z_{img} -axis coincident with the optical axis. $K_z(x_{img}, y_{img}, x, y)$ is the complex amplitude point spread function of the imaging system (objective and any other auxiliary lens) of the microscope. $K_z(\dots)$ describes the propagation from the object plane at $z = z_1$ to the image plane z_{img} . The image intensity $I(\vec{x}_{img})$ is

$$I(\vec{x}_{img}) = |U_{img}(x_{img}, y_{img}, z_{img})|^2. \quad (6)$$

The nonlinearity in the DIC image has two basic sources. First, since the image is an interference pattern, the detected intensity is the squared magnitude of the light field's complex amplitude. Therefore, a convolution of the light intensity with a lens transfer function does not accurately represent the DIC image, which is a linear superposition of complex amplitude, rather than intensity components, of the light field. In addition, out of focus contributions from the object have to be considered. Therefore a 3-D amplitude point spread (or transmission) function is needed to accurately model the image intensity. Second, the object itself aberrates the light wave as it propagates through. The biological specimens under consideration, though weakly refractive, are thick and scatter

light significantly. Therefore, aberrations due to the object contribute significantly to the image.

Deconvolution methods, such as computational optical sectioning microscopy (COSM), are widely used to recover object information from images acquired by certain optical modalities. COSM methods model the image intensity as a convolution of the object's intensity transmittance with a computed point-spread function. In modalities such as fluorescence and brightfield, a linear function of the intensity from the object provides an accurate, first order approximation of the image. However, in DIC microscopy, the image cannot be represented by merely considering the intensity from the object. Both phase and amplitude information have to be modeled.

Initial work in DIC image analysis algorithms used linear models. [4] In 1996, Feinegle used a contour finding algorithm to locate edges in each image from an optically sectioned stack. [6] In her work, specimen structure is obtained by axial stacking of contours. The most recent work in the analysis of DIC images is by Preza. [12] This work recovers the optical pathlength at each image point due to the object, and therefore does recover some quantitative information. The only other attempt to quantify information from DIC optics has been made by Cogswell, *et al.* using optical techniques referred to as geometric phase methods. [3] Though we are unaware of cases where this method has been applied to recover three dimensional object information. The work by Feinegle produced a 3-D model of the specimen, but the object properties were not quantitated and the recovered specimen model was not validated with ground truth experiments. Preza's work recovers optical path-length but does not actually reconstruct three dimensional object information. So far, no attempt has been made to quantitatively reconstruct the 3-D properties of the object from DIC images.

In contrast, the reconstruction algorithm that we have developed recovers the whole object information. We address the nonlinearities in the image formation process by a precise computational model of DIC microscopy which is used for reconstructing specimens. Using a generalized ray tracing method we have developed a model of light propagation through the microscope and specimen. [8] and [9] The computational model is used to generate simulated images of the current estimated object as part of the hierarchical reconstruction algorithm.

III. RECONSTRUCTION PROBLEM DESCRIPTION

Given a set of DIC images, the goal of the reconstruction is to estimate the refractive index distribution throughout the volume encompassing the object. Due to the nonlinearity of the DIC image-formation process, a direct inversion of the imaging equations is not feasible. Therefore, we use an iterative nonlinear optimization algorithm. The optimization starts with an initial estimate of the distribution. Using our computational

$$U_{img}(\vec{x}_{img}) = \int_{y_0^{obj}}^{y_1^{obj}} \int_{x_0^{obj}}^{x_1^{obj}} K_z(x_{img}, y_{img}; x, y) [U_{obj}^1(x, y, z = z_1) - U_{obj}^2(x, y, z = z_1)] dx dy$$

model, we generate simulated images corresponding to this estimate. The sum of squared differences between the real data and the simulated images is the error function. The estimate is modified at each iteration such that this error function is minimized. Such a reconstruction process requires a representation of the refractive index distribution at every point across the specimen volume. One possible representation is a weighted combination of basis functions where the weights are the parameters. The degrees of freedom in the representation, that is, the number of parameters used, play a critical role in the accuracy with which an object can be represented and the ease with which the optimization converges to the correct answer.

A nonlinear optimization method traverses the parameter space searching for the point where the error function is minimized. To avoid converging to incorrect minima, traditional nonlinear optimization problems require that the initial set of parameters be close to the actual solution. In order to be able to converge to the correct solution, despite the initial estimate being far, one needs a systematic method of traversing the parameter space such that the estimates approach the neighborhood of the global minima without being trapped in local minima. One possible method to achieve this is to first reduce the parameters such that only coarse object properties can be represented. If the volumetric distribution is approximated by a small number of parameters, the consequent error function is also of reduced dimension. In the process, local fluctuations (minima) are smoothed over and the error function retains global shape properties. One can then solve the optimization with respect to the reduced parameter set. When projected back to the original problem, the solution of the coarse optimization is closer to the original global minima. Thus the coarse solution serves as an appropriate initial estimate for the original optimization problem. Depending on the complexity of the problem, this parameter reduction can be done in multiple stages.

In our method, we represent the object and image data hierarchically and optimize at successive levels of the hierarchy by using a multilevel wavelet-like representation for the object. At each level, the original image data is projected onto a basis spanning a space of possible image features at the given object resolution level. This projected data is compared with simulated images in our nonlinear optimization. Thus, at each level of the hierarchical reconstruction we obtain an object estimate that fits the data at that level. This object is used as an initial estimate for the optimization at the next level of the hierarchy.

IV. COMPUTATIONAL MODEL

Our computational algorithm models the DIC image formation process with sufficient accuracy such that it can be used to reconstruct three dimensional optical properties of specimens. We achieve this by developing a generalized ray tracer with virtual models of the optical system and the specimen. We also model the focused and out of focus information present in DIC images by approximating light fields from each point in the object. We assume the fields have small enough aberration that they can be modeled as spherical. For each field, its radius of curvature is determined by the object's refractive structure.

The model consists of a polarized ray tracer and an approximation of the field contributions to the image. We assume that the object has negligible absorption and its refractive index variation significantly impacts the resulting image. In our model, light rays contain the cumulative effect of the object and optics by accumulating phase information. According to laws consistent with geometrical optics and energy conservation principles, we propagate light paths through the object. As each ray propagates through the specimen volume, the ray is deflected and the ray's phase is modified. We also approximate fields which emanate from all points in the object and which are diffracted by the object inhomogeneities. Lastly, we approximate the effect of diffraction by the lens pupil on each of these fields. Thus, we incorporate wavelength-dependent information in addition to the geometrical optics approximations allowing us to model DIC images more accurately than in past models.

V. RECONSTRUCTION ALGORITHM

The hierarchical reconstruction consists of three processes. In the first step, we represent the object at different resolutions. Second, we decompose the image data in order to identify image features corresponding to objects at the different resolutions. Third is the nonlinear optimization algorithm which recovers an estimate of the object at each resolution.

A. Hierarchical Object Representation

Since the DIC image captures the directional gradient of the phase introduced by the object into the light path, we represent the object as a volumetric distribution of refractive index values. At present, we assume that the object is transparent, and therefore the refractive index is a scalar quantity.³ We represent the refractive index by a wavelet-like bases, approximating it at different levels of resolution, where at each level only a subset of spatial frequencies are present. For the purpose of illustration, let us formulate in one dimension. At a particular resolution J , $n_J(x)$ is represented by combining versions of the scaling function, $\Phi(x)$, and wavelet function, $\Psi(x)$

$$n_J(x) = \sum_{k=-2}^{K-2} a_k \Phi(x - k) + \sum_{j=0}^J \sum_{k=-1}^{2^j K - 2} b_{j,k} \Psi(2^j x - k). \quad (7)$$

Both $\Phi(x)$ and $\Psi(x)$ are cubic spline functions, developed by Cai and Wang [2]. The scaling function represents the lowest spatial frequencies. The wavelet function represents the details at different resolutions. Translated versions of $\Phi(x)$, in addition to, scaled and translated versions of $\Psi(x)$ span the domain of an object. To represent a function at a particular level one needs to find the combination of weights up to that level which best approximates the original function. In Fig. 2, at the first level ($J = -1$), an example function is coarsely approximated by a combination of the translated scaling functions. At each subsequent level, the corresponding wavelet functions are added to the representation and the function is better approximated. To represent a two-dimensional (2-D) distribution, $n_J(x, z)$, we have to use a combination of functions that are outer products of

³This may be generalized to include absorption. The refractive index would then be a vector, or complex number.

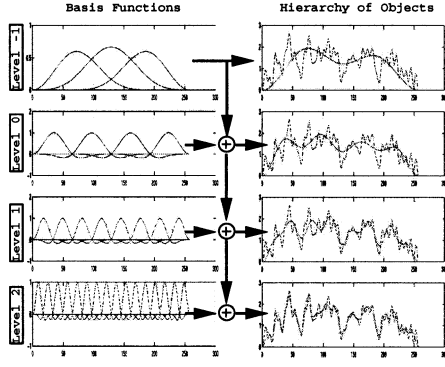


Fig. 2. Object hierarchy. In the right column an original 1-D object is shown superimposed with the approximate object at the level of the hierarchy. The left column shows the basis functions at each level. Each row shows the basis functions that are added to the representation at that level.

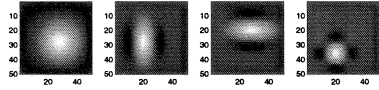


Fig. 3. Sample 2-D scaling and wavelet functions. Each corresponds to a term in (8).

the one-dimensional scaling and wavelet functions, (see equation (8) at the bottom of the page). Each tensor product results in a 2-D function that can represent different properties of the object. [Fig. 3] The extension to three dimensions is similar. The levels of the wavelet functions provide a systematic hierarchy for representing and recovering the object. At each level, of the reconstruction algorithm, the goal is to recover the parameters at that level. The scaling function coefficients ($a_{k,l}$) are recovered in the very first level ($J = -1$). At that level, the distribution is

$$n_{-1}(x, z) = \sum_{k=-2}^{K-2} \sum_{l=-2}^{L-2} a_{k,l} \Phi(x-k) \Phi(z-l). \quad (9)$$

The values of $a_{k,l}$ are initialized to 0. We then proceed to update the parameters $a_{k,l}$ iteratively so that the error between the simulated images and the image data at this level is minimized. The iterations at this level terminate when small changes in the parameter vector do not effect a significant improvement to the error. The algorithm then proceeds to the next level ($J = 0$). The estimated object from the previous level is transferred as the initial object at this level and we proceed to recover $b_{0,k,l}$, $c_{0,k,l}$, $d_{0,k,l}$. Analogous to the previous level, these parameters are initialized to 0 and iteratively updated so that the error between simulated images and the image data at level 0 is minimized. Each subsequent level of the reconstruction proceeds similarly.

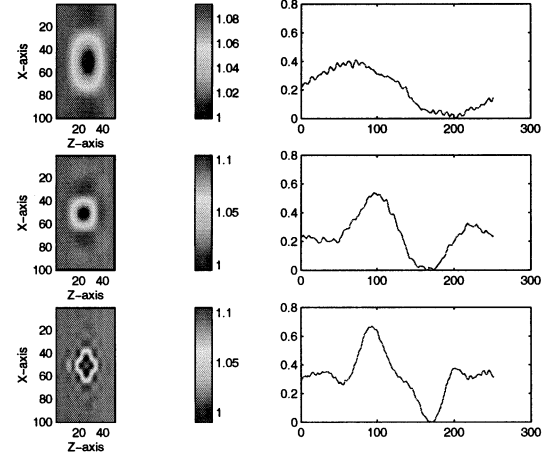


Fig. 4. Hierarchy of objects and images. The two object dimensions are the axial (Z) and transverse (X). For each image, the object plane is at the center of the Z axis. Rows 1–3 show the objects and corresponding images from the coarsest to finest resolution ($J = -1, J = 0, J = 1$).

The wavelet-like basis provides a continuous representation for the refractive index. For computations, we discretize this distribution. In our model, the refractive index is stored at discrete locations of a three-dimensional voxel grid. We typically use grids with resolution $0.2\text{--}1\text{ }\mu\text{m}$ perpendicular to the optical axis and $0.03\text{--}1\text{ }\mu\text{m}$ axially. A typical grid size is $100 \times 100 \times 50$ in the x , y and z dimension, respectively.

B. Hierarchical Image Representation

Unlike microscope modalities for which analytical models of the 3-D optical transfer function are available [13], object structure and the DIC image is related nonlinearly. Due to this non-linearity, an algorithm using an analytical DIC imaging model is not practical. If the relationship between the object and image intensity were linear, then a linear decomposition of the object would result in a linear decomposition of the image. Since this is not possible, we have developed a “matching by synthesis” algorithm that identifies image features which result from the represented object frequencies. The original data consists of image features due to all object frequencies but this data cannot be linearly decomposed into features at all levels. Therefore, at each resolution level we have to determine the image features which are appropriate and match these to the original data. An example of an object at different resolutions and the corresponding images simulated by our computational model is depicted in Fig. 4.

The matching by synthesis algorithm extracts features in the original intensity data which correspond to objects at a particular resolution. At a given resolution level J , we suppose that

$$n_J(x, z) = \sum_{k=-2}^{K-2} \sum_{l=-2}^{L-2} a_{k,l} \Phi(x-k) \Phi(z-l) + \sum_{j=0}^J \left[\sum_{k=-1}^{2^j K-2} \sum_{l=-2}^{2^j L-2} b_{j,k,l} \Psi(2^j x-k) \Phi(2^j z-l) + \sum_{k=-2}^{2^j K-2} \sum_{l=-1}^{2^j L-2} c_{j,k,l} \Phi(2^j x-k) \Psi(2^j z-l) + \sum_{k=-1}^{2^j K-2} \sum_{l=-1}^{2^j L-2} d_{j,k,l} \Psi(2^j x-k) \Psi(2^j z-l) \right] \quad (8)$$

the DIC image of the object can be represented in terms of a basis such that

$$I_J(x_{img}, z_{img}) = \sum_{q=0}^{Q^J} r_q^J B_q^J(x_{img}, z_{img}) \quad (10)$$

where $\{B_q^J\}$ for $q = 0 \dots Q^J$ is a set of basis functions specific to the resolution J . In order to relate the original image data, $I(x_{img}, z_{img})$, to the images at each resolution level, we need to find the features in the original data that correspond to the features in the basis functions. Before we can proceed to match features in the original data, we need to find the basis images. We obtain the basis images by perturbing parameters at a particular resolution and synthesizing an image per perturbed object. Functions selected to be in the basis represent the image features most common to the synthesized images. The original image data is then projected onto this basis, thereby, giving the image features that are present in the original data corresponding to an object at the given resolution. Specifically, our algorithm proceeds in three steps. First, at a particular resolution, we introduce a large set of perturbations to the initial object and generate simulated images of each of the perturbed objects. Second, we select a basis that best characterizes the images corresponding to the perturbed objects. Third, we project the original image data onto the estimated basis.

Object Perturbations: At a given resolution level J , the initial object is perturbed by a large number of random configurations. Each perturbation involves adding or subtracting a random amount from all parameters of the J^{th} level basis functions. Therefore, given random parameter vectors b^{rand} , c^{rand} and d^{rand} , a perturbed object at level J is

$$\begin{aligned} n_J^{pert}(x, z) = & n_{J-1}(x, z) \\ & + \sum_{k=-2}^{2^J K - 2^J L - 2} \sum_{l=-1}^{2^J K - 2^J L - 2} (b_{J,k,l} + b_{k,l}^{rand}) \\ & \Psi(2^J x - k) \Phi(2^J z - l) \\ & + \sum_{k=-1}^{2^J K - 2^J L - 2} \sum_{l=-2}^{2^J K - 2^J L - 2} (c_{J,k,l} + c_{k,l}^{rand}) \\ & \Phi(2^J x - k) \Psi(2^J z - l) \\ & + \sum_{k=-2}^{2^J K - 2^J L - 2} \sum_{l=-2}^{2^J K - 2^J L - 2} (d_{J,k,l} + d_{k,l}^{rand}) \\ & \Psi(2^J x - k) \Psi(2^J z - l). \end{aligned} \quad (11)$$

We use the computational model to generate a simulated image, $I_{sim,J}(x_{img}, z_{img})$, for each perturbed object. The set of simulated images contain a wide range of possible image features corresponding to objects at this resolution. Though, we have only shown perturbations of the wavelet coefficients, the coefficients of the scaling functions are also perturbed at level $J = -1$ to obtain basis images at that level. For a particular resolution, the total degrees of freedom at that resolution is

$$\alpha_J = (2^J K)(2^J L - 1) + (2^J K - 1)(2^J L) + (2^J K)(2^J L). \quad (12)$$

We have experimentally determined that $8\alpha_J$ perturbations (and simulated images $I_{sim,J}(x_{img}, z_{img})$) suffice.

Basis Selection: Next, we implement a basis selection method from the simulated images, $\{I_{sim,J}^i(x_{img}, z_{img})\}$ for $i = 1 \dots 8\alpha_J$. One simple basis selection method is a Karhunen-Louve decomposition applied to 2-D signals. [10], [14] This method results in a set of “eigenimages,” each functioning as a basis vector. We found that the eigenimages based method does not sufficiently abstract image features from the set of images. That is, although each $I_{sim,J}(x_{img}, z_{img})$ can be represented by a combination of the eigenimages, each eigenimage doesn’t necessarily represent isolated image features.

To find basis vectors which explicitly represent image features, we implemented a basis selection method using the “matching pursuit” algorithm. [11], [1], and [7] This algorithm uses a redundant dictionary of functions, where each function is a scaled and translated version of an exponentially modulated window function. In our implementation, the window is Gaussian. Using a redundant dictionary of such functions offers an advantage over decomposition into a pre-established basis, such as a wavelet basis. In a wavelet basis, scale and frequency have a fixed relationship so that only features which have a particular frequency content can be represented at a particular scale. In contrast, exponentially modulated functions can represent several kinds of features at any given scale.

The matching pursuit is an iterative greedy algorithm to identify functions out of the dictionary that best match an image in the set. Note, the iterations of the matching pursuit algorithm are embedded within each resolution level of the reconstruction algorithm. Given an intensity vector $I(x_{img})$, $(I(x_{img}) = I_{sim,J}^i(x_{img}, z_{img} = \text{constant}))$, $i \in [1, 8\alpha_J]$ in our algorithm), the matching pursuit algorithm defines a residual, $R_w\{I\}(x_{img})$ at each iteration w , where $R_0\{I\}(x_{img}) = I(x_{img})$. An iteration consists of finding, the best function $g_b^w(x_{img})$ from candidate functions in the dictionary, such that

$$\langle R_w\{I\}(x_{img}), g_b^w(x_{img}) \rangle > \langle R_w\{I\}(x_{img}), g_i(x_{img}) \rangle \quad (13)$$

where $\langle \rangle$ denotes inner product. $g_b^w(x_{img})$ is then removed from the candidate functions in the dictionary. The residual is updated as

$$\begin{aligned} R_{w+1}\{I\}(x_{img}) = & R_w\{I\}(x_{img}) \\ & - \langle R_w\{I\}(x_{img}), g_b^w(x_{img}) \rangle g_b^w(x_{img}). \end{aligned} \quad (14)$$

The algorithm is terminated when the residual’s norm falls below a preset threshold. More details can be found in the references.

As a result of matching pursuit, each $I_{sim,J}^i$ is decomposed with respect to a unique matched set, $\{g_b^w(x_{img})\}$ for $w = 1 \dots w_{\max}$, selected from the dictionary. Therefore we will have $8\alpha_J$ matched sets. Finally, we extract a set of dissimilar functions that are most prevalent across all the matched sets.⁴ This set of functions forms the basis, $\{B_q^J\}$ for $q = 0 \dots Q^J$, for the subspace of images corresponding to objects at resolution level J . [Fig. 5] The number, Q^J , is set experimentally to ensure that $\{B^J\}$ spans the entire image.

⁴Two functions, $g_1(x_{img})$ and $g_2(x_{img})$, are dissimilar if $\langle g_1, g_2 \rangle / \langle g_1, g_1 \rangle \langle g_2, g_2 \rangle$ is less than an empirically determined threshold.

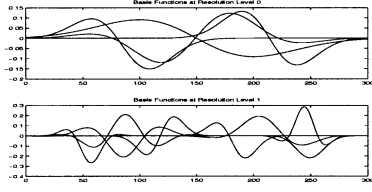


Fig. 5. 1-D basis functions at the levels 0 and 1. (From specimen 1 reconstruction).

Image Projection: Once the appropriate basis hierarchy for the images is selected, the original image data is projected onto the basis at each level. A matrix, M_J , is constructed in which all the columns are the individual functions comprising of the basis at a particular resolution. That is

$$M_J = \begin{bmatrix} B_0^J(x, z=0), & \dots & B_{Q_J}^J(x, z=0) \\ B_0^J(x, z=1), & \dots & B_{Q_J}^J(x, z=1) \\ \vdots & & \\ B_0^J(x, z=z_{\max}), & \dots & B_{Q_J}^J(x, z=z_{\max}) \end{bmatrix}. \quad (15)$$

In the above matrix, each basis image corresponding to a different axial position is stacked vertically so that the entire basis image can be represented as a column vector. Since the matrix will not be square in most cases, the original data is multiplied with the pseudo-inverse to obtain the necessary weights for each of the basis functions. This is the over constrained solution of $(d/d\vec{r})\{M_J\vec{r} - I\}^2 = 0$ where \vec{r} is the vector of weights, I is the original data, formatted into a column vector similar to the basis images above. The least squares solution is $\vec{r} = M_J^\dagger I$, where M_J^\dagger is the Moore–Penrose pseudo-inverse of M_J . The projected image, I_J

$$I_J = P_J\{I\} = M_J\vec{r} \quad (16)$$

corresponding to the J^{th} resolution is the weighted linear combination of the basis functions, using the functions obtained with the matching pursuit algorithm and the weights based on the least squares solution.

C. Optimization at Each Resolution

For each level in the resolution hierarchy, we optimize with respect to the wavelet coefficients at the current resolution. At a particular level, the initial object is the estimated object from the previous level. At the first resolution level, the initial object is a completely blank volume, i.e., all voxels are initialized with the same refractive index as the background. The image data used in the optimization is the original data projected onto the basis selected for this resolution level. (I_J defined previously.) So at each resolution level, the target data is actually the projected images described above. At the final resolution level, we use the original data. The sum of squared differences between the projected data and the simulated data is the error term for the optimization.

Levenberg-Marquardt: At the first resolution level, $J = -1$, the degrees of freedom of the object is quite low. Therefore, a Levenberg-Marquardt type nonlinear optimization produces sufficiently good results. [15] This is basically a Gauss-Newton type gradient-based optimization, where the Hessian matrix is

supplemented with an identity matrix scaled by a parameter. The parameter is decreased at each iteration if the current estimate reduces the error, otherwise it is increased. Implicitly, this optimization method assumes that the neighborhood of the global solution is predominantly convex, barring some minor local undulations, and that the current estimate is within this neighborhood.

Randomized Methods: At subsequent resolution levels, the degrees of freedom in the object increase rapidly. No longer can one assume that the initial estimate is close to the solution at this level and that the neighborhood along the error surface to be traversed is sufficiently convex. Therefore, Gauss-Newton type methods such as Levenberg-Marquardt tend to converge to local minima. Although, the optimization is significantly aided by the resolution hierarchy, we still need to implement a randomized search method within the nonlinear optimization. We have implemented a genetic algorithm that has been very successful in finding the correct solutions.

The genetic algorithm combines parameters from members of a population to produce a new member. The population that we use to initiate the genetic algorithm is the same set of perturbed parameter vectors described above. Each parameter vector, p , is a concatenation of $b_{J,k,l}$, $c_{J,k,l}$ and $d_{J,k,l}$. In addition, each p has an associated error value which is the sum of squared differences between the simulated image due to this object and the projected image at the current level, plus any penalty terms due to constraints. The population is divided into good and bad candidates, represented by p_{good}^k and p_{bad}^k . The good and bad candidates have error values in the top and bottom 50th percentile respectively. There are two ways of combining member parameters. The first method selects two good members (p_{good}^1 and p_{good}^2) and a bad candidate (p_{bad}^1). The new member is $p_{\text{new}}^{\text{comb}} = a(.5p_{\text{good}}^1 + .5p_{\text{good}}^2) - bp_{\text{bad}}^1$ where $a > 0$, $b > 0$ and $a - b = 1$. This method is similar to traditional combination methods in genetic algorithms. In this selection method, good candidates with lower error values have better chances of being selected than ones with higher error values. If the new member is better than either of its “parents,” i.e., p_{good}^1 and p_{good}^2 , according to the above defined error, then the worst parent is replaced by the new member. The second method simply selects a member of the population at random and randomly perturbs some parameter values, i.e., $p_{\text{new}}^{\text{mutate}} = p^k + p_{\text{rand}}$ where p^k is a randomly selected parameter vector from the entire population and p_{rand} is a vector of random values. This method is consistent with traditional mutation. If the error value corresponding to $p_{\text{new}}^{\text{mutate}}$ is better than that of the worst parameter vector, then this mutant replaces the worst member. At each iteration of the algorithm we sort the members according to their error values, and perform one combination and one mutation step.

Constraints: The only strong constraint on the object is that the minimum refractive index be 1.0. Therefore, every estimated distribution is offset by a suitable amount to ensure that the minimum value in the distribution is 1.0. In our reconstruction experiments, the physical specimen has only two distinct refractive index values. So for those experiments we add a term in the error that penalizes for inhomogeneity in the distribution. This additional penalty term is only introduced at the finest resolution level and has considerably sped up convergence. The penalty is

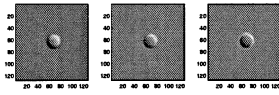


Fig. 6. Physical data of bead specimen.

achieved by segmenting the volume into regions of high and low refractive index values and then measuring the variance within those regions. The variance terms are scaled by parameters, ω and β and added to the error. Therefore, the modified error is

$$E = \sum_{m=1}^M (I_J[m] - I_{sim}[m])^2 + \frac{\omega}{N_{high}} \sum_{k \in \gamma_{high}} (n[k] - n_{high})^2 + \frac{\beta}{N_{low}} \sum_{k \in \gamma_{low}} (n[k] - n_{low})^2 \quad (17)$$

where the first term is the sum of squared differences between the intensity values of the projected image and the simulated image calculated over the total number of pixels, M . In the rest of the error, N is the total number of voxels, n_{high} and n_{low} are the means of the high and low valued $n[x]$ respectively, γ_{high} and γ_{low} are the sets of discrete voxels which are above and below the threshold, respectively.

VI. RESULTS

The first set of reconstruction results we obtain are in two dimensions. That is, instead of estimating the full 3-D volumetric distribution, we chose to reconstruct one slice in the volume. To provide a good proof of concept for this reconstruction algorithm that would be less computationally expensive than a full 3-D reconstruction, we chose an axial slice (rather than a transverse slice). Our decision is based on the fact that reconstructing along the axial dimension presents the greatest challenge. Spatial frequencies which have significant components in the axial direction are attenuated the most in the image formation process. The two transverse dimensions are imaged according to identical principles, thus reconstructing in one of them can serve as a validation for both. We chose two test specimens, each having some unique attributes. The first specimen, consisting of a bead in optical cement, is symmetric and the object structure is embedded in the volume. The second specimen, consisting of an ion-milled wafer, does not have perfect symmetry and the structure extends to the top edge of the volume.

In the first 2-D experiment, we embedded several 10 μm diameter beads in optical cement. The homogeneous beads have a refractive index that is .03 less than the cement. We acquired an optically sectioned data set of this specimen with an axial resolution of .2 μm . A cropped region showing one bead in some of the images is shown in Fig. 6. The data used in the reconstruction experiment consisted of a diagonal cut through the cropped region of 10 images in the set. A cropped specimen is shown in Fig. 7. Some of the intensity plots are shown in Fig. 8. The image data projected onto the basis at different object resolutions is shown in Fig. 8 as well. Fig. 9 shows the estimated objects at the different resolutions and a comparison between the

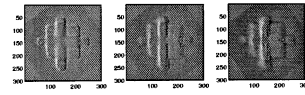


Fig. 7. Real data of wafer specimen.

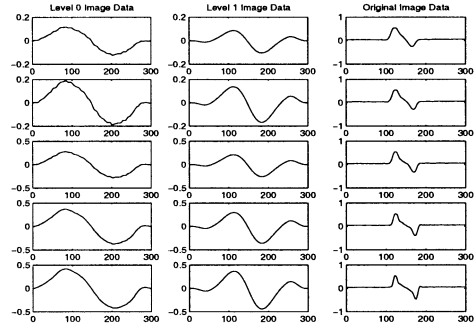


Fig. 8. Intensity plots at different resolutions. Column 1–2: The original data projected onto the levels 0–1 respectively. Column 3: The original data.

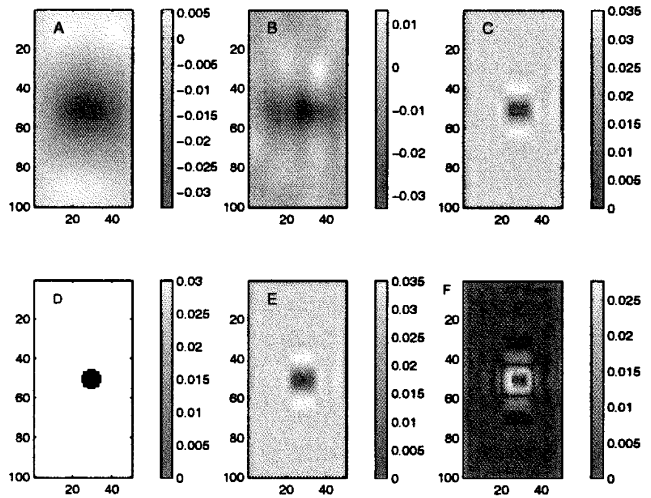


Fig. 9. Results for Specimen 1. A–C are the estimated answers at levels 0–2. D: Real object. E: Final estimate. F: Error between real and estimated object.

expected object structure and the recovered object. The grid resolution in this experiment is 1 μm in both directions. As can be seen from the error image, the object structure is recovered quite accurately. The original object has symmetric structure and the recovered shape is symmetric as well.

The specimen used for the second experiment consisted of an ion-milled glass wafer. Due to error prone multiple millings, the actual milled wafer pattern is defective in that it does not have symmetric walls even though the specifications have symmetric structure. The specimen was prepared by filling the milled cavity with oil that had a refractive index 0.08 less than the glass. The optically sectioned image set of this specimen has a 0.2 μm axial resolution. A subset of the images are shown in Fig. 7. For the reconstruction experiment, we extracted a line of intensity data from each one of ten images in the set. The line (horizontal) was extracted above the midpoint of the shown images at a point such that two of the depressions in the pattern are captured. The original image data is shown in Fig. 10 along with the results of the image data projected onto the basis at different object resolutions. Fig. 11 shows

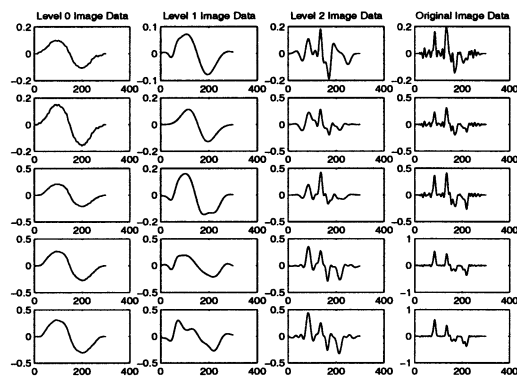


Fig. 10. Intensity plots of wafer data at different resolutions. Columns 1–3: The projected data at levels 0–2, respectively. Column 4: The original data.

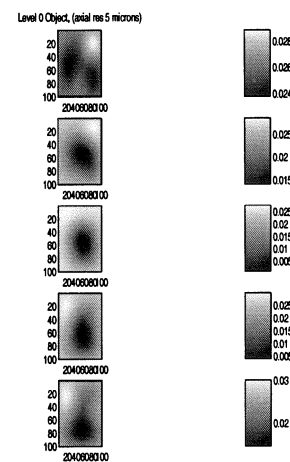


Fig. 12. Bead experiment results: Estimated object at level 0. The object is shown by transverse slices through the volume at a resolution of $5 \mu\text{m}$.

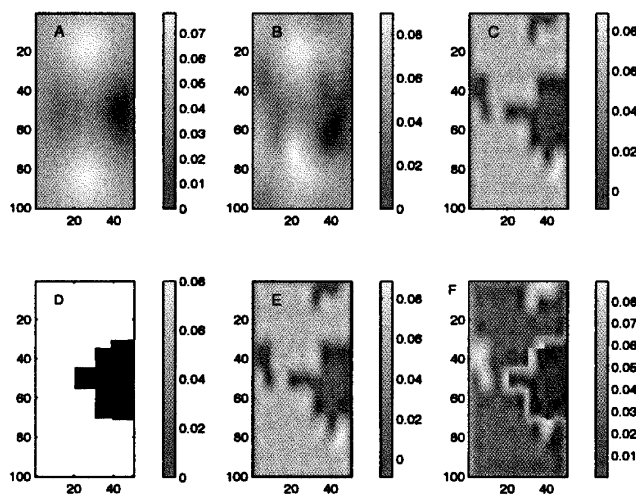


Fig. 11. Results for Specimen 2. A–C are the estimated answers at levels 0–2. D: Real object. E: Final estimate. F: Error between real and estimated object.

the estimated object at different resolutions and a comparison of the final reconstructed object and the expected object. The grid resolution in this experiment is $.3$ in the x -direction and 0.03 in the z -direction. Since the exact nature of the defects introduced by the milling process is not known, the expected object shown is merely a hypothesis. As can be seen, barring some small extraneous patches, the structure of the recovered object is very close to that of the expected object. The lack of symmetry that is apparent from the DIC images appears in the recovered object as well.

For our final experiments, we tried reconstructing a 3-D object. Here we represented the object as a superposition of three-dimensional wavelet basis functions and recovered the respective coefficients. The three-dimensional experiment used the optically-sectioned data set of bead images. A sampling of these images is shown in Fig. 6. We used 25 images from the data set for this reconstruction experiment. The initial object was a blank volume. The estimated objects at resolution level 0 is shown in Figs. 12 and 13. A slice in Fig. 12 represents the volumetric refractive index distribution across a transverse planes cutting the volume. In Fig. 13 we show an axial slice cutting through the center of the volume so the estimated object extents can be seen. Figs. 14–15 show the estimated object at level 1. For visualization purposes, we show the results as

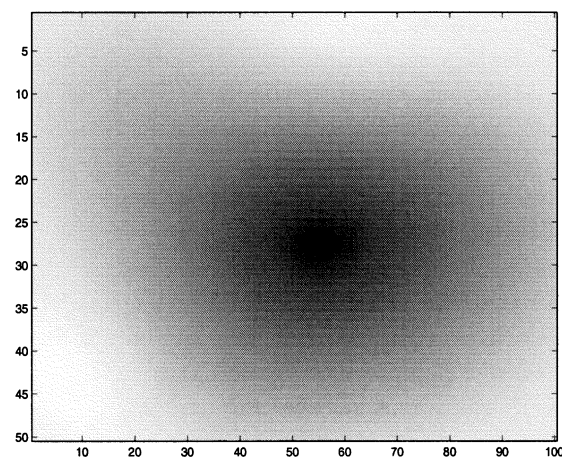


Fig. 13. Axial slice through object at Level 0. The vertical axis is Z and the horizontal axis is X .

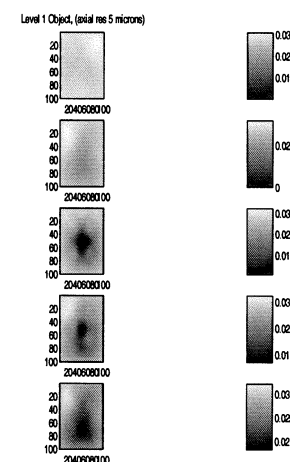


Fig. 14. Bead experiment results: Estimated object at level 1. As in the previous figure, the object is shown using transverse slices through the volume at a resolution of 2.5 mm .

planes through the object. The actual experiment is performed volumetrically. That is, each basis function is 3-D. In Figs. 13 and 15 we show an axial slice cutting through the center of the volume so that the estimated object extents can be seen.

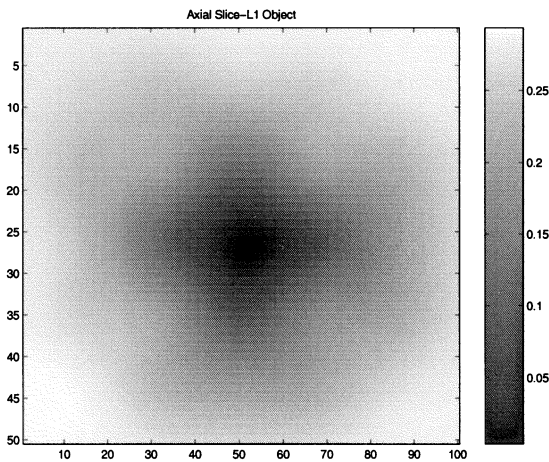


Fig. 15. The object at level 1 is shown as an axial slice through the center of the volume. The vertical axis is Z and the horizontal axis is X .

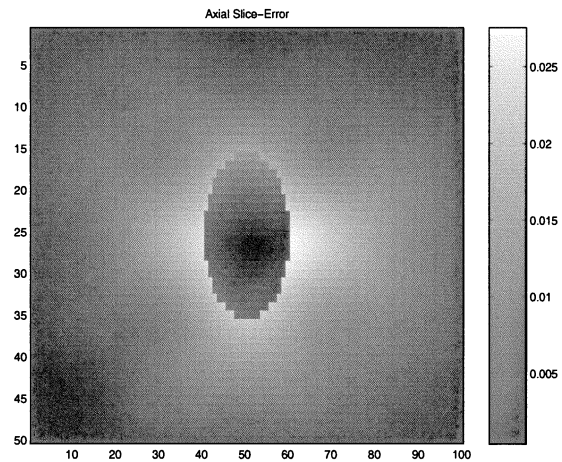


Fig. 17. The error between real object and the estimated object at level 1. The image here shows the error through an axial slice cutting the center of the volume. The vertical axis is Z and the horizontal axis is X .

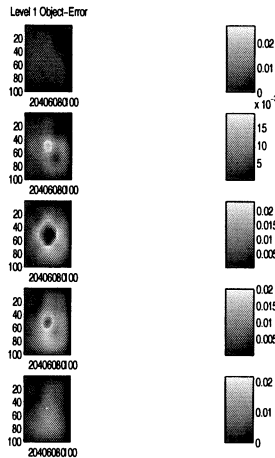


Fig. 16. The error between real object and estimated object at level 1. The error through the volume is shown by transverse slices at a resolution of 2.5 microns.

At level 0, 27 coefficients are estimated which represent all the translated scaling functions. At level 1, the initial object is the level 0 object. The total number of coefficients at level 1 is 316 out of which we estimate only 26. These coefficients correspond to the basis functions spanning the boundary and internal extents of the initial object at this level. By simply ignoring the basis functions which fall outside of the estimated initial object (at this level) we can eliminate the majority of the basis functions. This assumes that new object will fall within the boundary of the old object which is valid since the basis functions at the coarser level have larger support.

Comparison of the real object and the estimated object is shown in Fig. 16. It can be seen that the estimated object is slowly converging to the real object shape. At this resolution level, the basis functions have too wide support to produce an object with sharp boundaries. The comparison does show that the optimization estimates a distribution that approaches the real object. The error is shown axially in Fig. 17. Here it can be seen that the estimated object's extents do approach the real object dimensions in the axial direction as well. One can also note that the refractive index value of the estimated object does approach that of the real object. The centers of both the real

object and estimated object also approximately coincide. These results are meant as a proof of concept to show that three-dimensional reconstruction is possible with this method. This experiment shows that given enough iterations, the estimated object will approach the real object shape and recover the object's optical property.

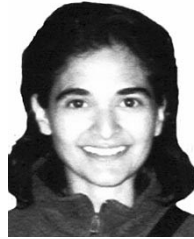
VII. DISCUSSION AND CONCLUSION

The analysis of DIC images presents a substantial challenge due to the nonlinearity of the image formation process and the out-of-focus artifacts. We tackle the problem by developing two tools. First, the computational model has been discussed in detail in previous publications. The second is the hierarchical reconstruction algorithm discussed here. As shown by the results, the nonlinear optimization is powerful enough to recover axial and transverse structure and quantitate the optical properties of the object. Even though, we initialize the optimization far from the actual solution, we are able to converge very close to it. In order to successfully converge to solutions when faced with highly nonlinear error surfaces, we developed a hierarchical method. This method represents the object with respect to a wavelet basis in order to systematically reduce the dimension of the search space and arrive at a chain of object estimates at finer and finer resolutions. The intensity data also has to be decomposed in a manner consistent with the object decomposition. Since the imaging process is nonlinear, we developed a method by which image features present at the different object resolutions are explicitly captured in a particular basis at each level. The real data can then be projected onto the basis at each resolution level to obtain a hierarchical representation of the data. Our results show that such an algorithm is capable of recovering structure along all directions. By recovering the structure of two very different specimens, we illustrate the capabilities of the reconstruction algorithm.

REFERENCES

- [1] F. Bergeaud and S. Mallat, "Matching pursuit: Adaptive representations of images and sounds," *Comput. Appl. Math.*, vol. 15, no. 9, pp. 97–109, 1996.

- [2] W. Cai and J. Wang, "Adaptive multiresolution collocation methods for initial boundary value problems of nonlinear pdes," *SIAM J. Numer. Anal.*, vol. 33, no. 3, pp. 937–970, 1996.
- [3] C. J. Cogswell, N. Smith, K. Larkin, and P. Hariharan, "Quantitative dic microscopy using geometrical phase shifter," in *Proc. SPIE*, vol. 2984, 1997, pp. 72–81.
- [4] K. Dana, "Three Dimensional Reconstruction of the Tectorial Membrane," M.S. dissertation, Mass. Inst. Technol., Cambridge, 1992.
- [5] A. Erhardt, G. Zinser, D. Komitowski, and J. Bille, "Reconstructing 3-D light microscopic images by digital image processing," *Appl. Opt.*, vol. 24, no. 2, pp. 194–200, 1985.
- [6] P. Feineigle, "Motion Analysis and Visualization of Biological Structures Imaged via Nomarski DIC Light Microscopy," Ph.D. dissertation, Carnegie Mellon Univ., Pittsburgh, PA, 1996.
- [7] S. Jaggi *et al.*, "High resolution pursuit for feature extraction," *Appl. Comput. Harm. Anal.*, vol. 5, pp. 428–449, 1998.
- [8] F. Kagalwala, F. Lanni, and T. Kanade, "Simulating DIC microscope images," *Proc. IEEE Wkshp. Photonics Modeling Computer Vision Graph.*, June 1999.
- [9] —, "Computational Model of DIC Microscopy: From Observations to Measurements," CMU-R1 TR, Carnegie Mellon Univ., Pittsburgh, PA, TR-00-10, 2000.
- [10] M. Kirby and L. Sirovich, "Applications of the karhunen-louve procedure for the characterization of human faces," *IEEE Pattern Anal. Machine Intell.*, vol. 12, pp. 103–108, Jan. 1990.
- [11] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. on Signal. Processing*, vol. 41, pp. 3397–3410, Dec. 1993.
- [12] C. Preza, "Phase Estimation Using Rotational Diversity for DIC Microscopy," Ph.D. dissertation, Washington Univ., St. Louis, MO, 1998.
- [13] N. Streibl, "Three-dimensional imaging by a microscope," *J. Opt. Soc. Amer. A*, vol. 2, no. 2, pp. 121–127, 1985.
- [14] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cog. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.
- [15] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C*, 2 ed. Cambridge, U.K.: Cambridge Univ. Press, 1992.



Farhana Kagalwala received the B.S. degree from University of Michigan, Ann Arbor in 1993, and the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, in 2001.

Currently, she is working at the Advanced Imagery Lab at General Atomics, San Diego, CA. As part of her thesis, in addition to the present work, she developed a model of the image formation process in a differential interference contrast (DIC) microscope which went beyond the first Born approximation. Her

approach to extracting volumetric and shape properties from a set of images uses the sensor physics as an integral part of the reconstruction process. In general, her research interests are computer vision, graphical models, and heirarchical methods.

Takeo Kanade (F'92) received the Ph.D. degree in electrical engineering from Kyoto University, Japan, in 1974.

After being on the Faculty of the Department of Information Science, Kyoto University, he joined the Computer Science Department and Robotics Institute at Carnegie Mellon University (CMU), Pittsburgh, PA, in 1980. He became an Associate Professor in 1982, a Full Professor in 1985, the U.A. and Helen Whitaker Professor in 1993, and a University Professor in 1998. He was the Director of the Robotics Institute from 1992 to Spring 2001. He served as the founding Chairman from 1989 to 1993 of the Robotics Ph.D. Program at CMU, probably the first of its kind in the world. He has worked in many areas of robotics, including manipulators, sensors, computer vision, multimedia applications, and autonomous robots, with more than 200 papers on these topics. He has been the founding editor of the *International Journal of Computer Vision*. His research interests are in the areas of computer vision, visual and multimedia technology, and robotics. Common themes that his students and he emphasize in performing research are the formulation of sound theories which use the physical, geometrical, and semantic properties involved in perceptual and control processes in order to create intelligent machines, and the demonstration of the working systems based on these theories. His current projects include basic research and system development in computer vision (motion, stereo and object recognition), recognition of facial expressions, virtual(ized) reality, content-based video and image retrieval, VLSI-based computational sensors, medical robotics, and an autonomous helicopter.

Dr. Kanade's professional honors include: election to the National Academy of Engineering, a Fellow of ACM, a Fellow of American Association of Artificial Intelligence; several awards including C&C Award, the Joseph Engelberger Award, Yokogawa Prize, JARA Award, Otto Franc Award, and Marr Prize Award.