Multilinear Feature Extraction and Classification of Multi-Focal Images, With Applications in Nematode Taxonomy

Min Liu, Amit K. Roy-Chowdhury*
Department of Electrical Engineering, University of California, Riverside

Abstract

In this paper, we present a 3D X-Ray Transform based multilinear feature extraction and classification method for Digital Multi-focal Images (DMI). In such images, morphological information for a transparent specimen can be captured in the form of a stack of high-quality images, representing individual focal planes through the specimen's body. We present a method that can effectively exploit the entire information in the stack using the 3D X-Ray projections at different viewing angles. These DMI stacks represent the effect of different factors - shape, texture, viewpoint, different instances within the same class and different classes of specimens. For this purpose, we embed the 3D X-Ray Transform within a multilinear framework and propose a Multilinear X-Ray Transform (MXRT) feature representation. By combining the tensor texture and shape information we can get better recognition rates than just relying on the original or key frames of DMI stacks. The experimental results on the nematode DMI data show that the 3D X-Ray Transform based multilinear analysis method can effectively give 100% *recognition rate on a real-life database.*

1. Introduction

In the biological field, Digital Multi-focal Images (DMI) is a new way for documentation and communication of specimen data [4]. In this approach, morphological information for a transparent specimen can be captured in the form of a stack of high-quality images, representing individual focal planes through the specimen's body, as shown in Figure 1. With the data from DMI, the viewer can reconstruct a three-dimensional image of the specimen. Thus, the use of DMI overcomes the limitations imposed by traditional methods, which omit three-dimensional topology of structures, especially all topological information along the focal axis [4, 7].

There is little work in the analysis of DMI images, although they are becoming very popular in biological imaging. Figure 1 shows 3 DMI stacks of images taken from a Differential Interference Contrast microscope [1]. Given such stacks of multifocal images, how do we efficiently integrate the information from all layers? To reach this objective, we propose the 3D X-Ray Transform to explore the information through different image layers, so that we can make better feature extraction and classification for DMI images. For example, the first 2 DMI stacks in Figure 1 belong to the same nematode class, but it is possible to make the mistake that they are from different classes by shape or texture based classification, relying only on the original image frames. However, if we project those 2 DMI stacks along different angles with the 3D X-Ray Transform, those projections in the same direction look similar, as shown in the top 2 rows of Figure 5. Moreover, although the original key frames in the 2nd and 3rd stacks look similar, they belong to different species, and in Figure 5 we can see that their projections along the 2nd and 3rd directions are differ-

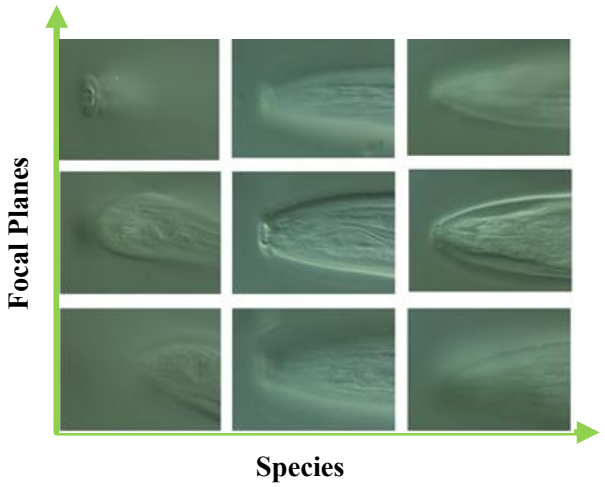

Figure 1. Examples of DMI stacks taken from 3 nematode species (horizontal direction). Each stack contains multiple focal planes (frames) taken from the top to the bottom of the specimen, with only 3 key frames of each shown here (vertical direction).

*This work was supported by NSF Grant IIS-0808770.

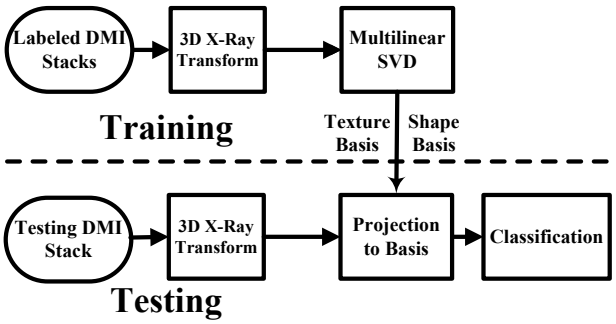

Figure 2. The diagram for training and recognition of the proposed 3D X-Ray Transform based multilinear analysis method.

ent. Based on this we can make the right classification for those 3 DMI stacks. The 3D X-Ray Transform allows us to analyze the entire multi-focal stack of images from different viewing angles - something direct analysis of the images would not allow.

For the classification of DMI stacks, there are multiple factors that need to be analyzed - shape, texture, viewpoint of the X-Ray Transform, different instances within the same class and different classes of specimens. For this purpose, we embed the 3D X-Ray Transform within a multilinear framework. This allows us to model the inter-relationships between the different variables, while at the same time allowing independent analysis along each dimension. We will show the application of this feature extraction and representation technique in classification of nematodes - a species that is very difficult to classify since there are over 80,000 of them. This is possibly the first approach that looks at the automated analysis of DMI images, which is becoming increasingly popular in biology.

1.1. Proposed Approach and Relation to Existing Work

Most of the work on image feature extraction and classification are done through the 2D image processing methods [10, 13, 14]. Since we want to integrate all the 2D feature information across different image layers, we propose the use of 3D X-Ray Transform, which will explore the 3D structural information out of a multifocal image stack from different viewing directions. As shown in Figures 4 B-F, the 3D X-Ray Transform of the 1st DMI stack in Figure 1 gives us different information about this stack at 5 different angles. Recent methods for 3D feature extraction like spatio-temporal interest points are not suited for this purpose because the images in multiple focal planes have different characteristics compared to a space-time volume [8].

In the face recognition field, the multilinear image analysis method plays an effective role for the recognition of natural face images formed by different factors, such as expression, illumination, pose and so on [9, 12]. In multilinear analysis, the facial image ensemble is represented as a high dimensional tensor. By a high-order generalization of Principal Component Analysis (PCA) and high-order singular value decomposition (HSVD), the collection of facial images is represented by separating the different modes underlying the formation of facial images. This idea can be applied to the classification of DMI stacks, since they also represent the combination of various factors - poses (viewpoints) of the X-Ray Transform, different instances within the same class and different classes of specimens. Therefore it is intuitive to use the multilinear method for feature extraction and classification of the DMI data and we do so by combining it with the 3D X-Ray Transform.

1.2. Broad Overview of Solution Strategy

In this paper, we will show how to derive features for DMI data sources that take into account the shape and texture information across the focal planes, as well as the variability when viewed from different directions. We will show how to design classifiers based on this Multilinear X-Ray Transform feature representation and provide experimental results on its usefulness in large real-life databases. The diagram of this proposed MXRT method is shown in Figure 2, and the training and testing steps are listed below.

1. Training
   (a) The inputs are labeled DMI stacks.
   (b) For each training DMI image stack, compute the 3D X-Ray Transform at defined angles (poses).
   (c) For 3D X-Ray Transform output images at each pose (angle), obtain the tensor texture basis and tensor shape basis through multilinear SVD.
2. Testing
   (a) Given an unknown DMI stack, get the 3D X-Ray Transform output images at the same defined poses (angles) as in the training procedure.
   (b) Project the 3D X-Ray Transform output images at each pose to the corresponding tensor texture basis and tensor shape basis, and obtain the distance matrices for texture and shape.
   (c) Get the distance matrix by combining the tensor texture and shape distance matrices, and then find the class label through nearest neighbor criteria for the 3D X-Ray Transform output image at each pose of the given DMI stack.
   (d) Obtain the final recognition result by a majority voting process between the labels for each X-Ray Transform pose from the previous step.

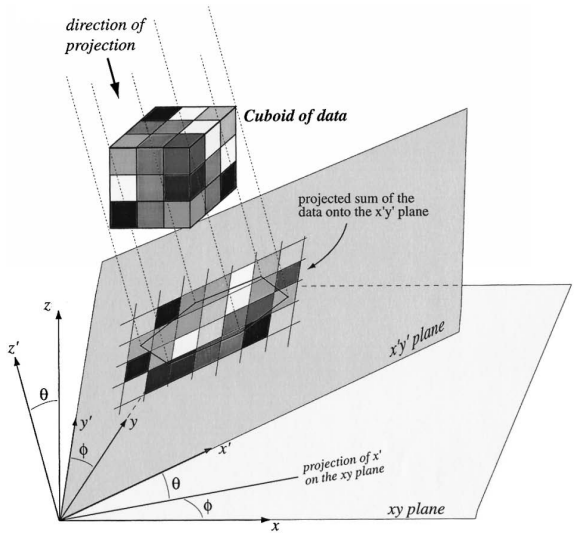

Figure 3. The illustration of 3 dimensional X-Ray Transform [2].

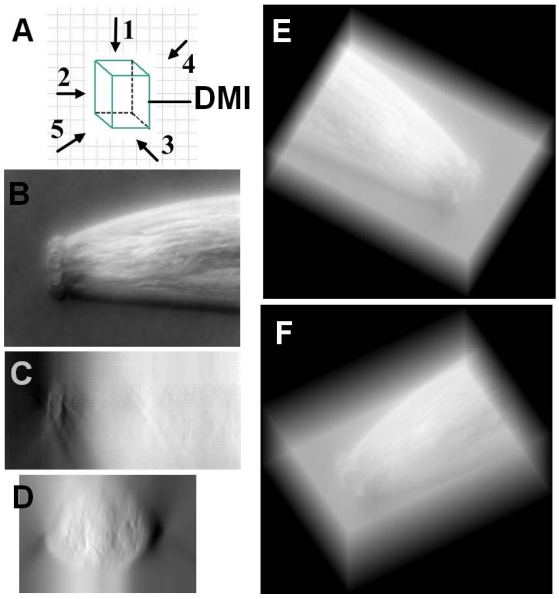

Figure 4. 3D X-Ray Transform output images at 5 poses (angle pairs) (B($\pi/2, \pi/2$), C($0, \pi/2$), D(π, π), E($\pi/5, \pi/3$) and F($\pi/3, \pi/5$)), of the 1st DMI stack shown in Figure 1. A is the illustration of different poses around a given DMI stack for the 3D X-Ray Transform.

2. Detailed Methodology

2.1. 3D X-Ray Transform

The 2D X-Ray Transform of an image is the projection of the image intensity along a certain angle [11], similar as the Radon Transform [5]. In the 3D case, the X-Ray Transform is the object's projection on a plane, whose orientation is identified by a pair of angles (θ, ϕ), as shown in Figure 3. It can be mathematically defined as:

$$R(x', y', \theta, \phi) = \int_{z'} S(x, y, z) dz', \tag{1}$$

where $S(x, y, z)$ denotes the DMI images along the z direction, x, y and x', y' are the coordinates in the original 2D image plane and the transformed image plane respectively, and the transformation of those coordinates is expressed as [2]

$$\begin{cases} x' = x\cos\phi\cos\theta + y\sin\phi\cos\theta + z\sin\theta \\ y' = -x\sin\phi + y\cos\phi \\ z' = -x\cos\phi\sin\theta - y\sin\phi\sin\theta + z\cos\theta \end{cases}. \tag{2}$$

As explained in the Introduction, the biggest advantage of DMI is that it can keep the 3D topology and structure information of the specimen. The topological and structural information can be effectively captured by the 3D X-Ray Transform at different poses (i.e., angles) as in Figure 4A. Figures 4B-F are the projections of 3D X-Ray Transform of the 1st DMI stack in Figure 1, at 5 different poses, $(\pi/2, \pi/2)$, $(0, \pi/2)$, (π, π), $(\pi/5, \pi/3)$ and $(\pi/3, \pi/5)$ respectively. From Figure 4 we can see that different poses of projections capture very different characteristics of the DMI stack. The 3D X-Ray Transform output images at the above 5 poses will be used for multilinear feature extraction and classification of different DMI stacks in the experiments.

Based on the outputs from 3D X-Ray Transform of DMI stacks, we can make very accurate feature extraction and classification from those multifocal image stacks. Figure 5 shows the 3D X-Ray Transform of the 1st, 2nd and 3rd DMI stacks shown in Figure 1 at 5 poses. We can see that although the original 2D image frames in the 1st stack and 2nd stack are totally different, their X-Ray Transform projections are similar. If we do the classification just relying on the 2D image frames, we will wrongly classify the first 2 DMI stacks. But by looking at those 3D X-Ray Transform output images, we can make the correct classification.

2.2. Multilinear Subspace Analysis

A tensor is a multidimensional matrix or we can regard it as high order matrix or n-mode matrix; it is the generalization of a vector (first order tensor) and a matrix (second order tensor). For an N^{th} order tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times \ldots \times I_n \times \ldots \times I_N}$, whose elements can be denoted as $\mathcal{A}_{i_1 \ldots i_n \ldots i_N}$, where $1 \leq i_n \leq I_n$, its mode-n vectors are I_n-dimensional vectors obtained from $\mathcal{A}$ by varying index i_n while keeping the other indices fixed. The mode-n vectors are the column vectors of matrix $\mathcal{A}_{(n)}$ that results by mode-n flattening the tensor $\mathcal{A}$ [12]. As the Singular Value Decomposition (SVD) in matrix analysis orthogonalizes the row space and column space, an N-mode tensor also can be

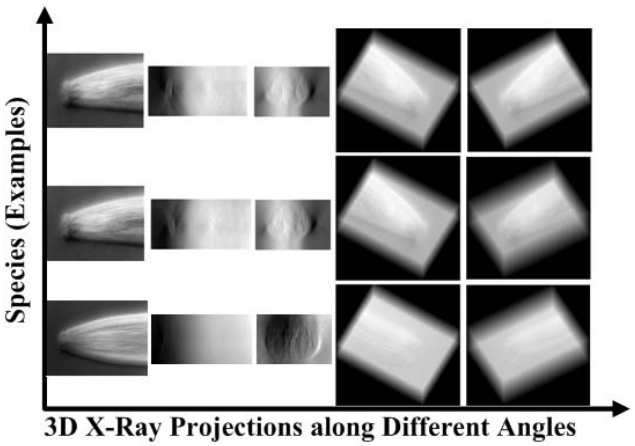

Figure 5. 3D X-Ray projections along 5 directions($(\pi/2, \pi/2)$, $(0, \pi/2)$, (π, π), $(\pi/5, \pi/3)$ and $(\pi/3, \pi/5)$), for the DMI stacks (represented row-wise) shown in Figure 1.

decomposed by N-mode SVD. It is an extension of SVD that orthogonalizes these N spaces and expresses the tensor as the mode-n product [12], denoted $\times_n$, of N-orthogonal spaces as

$$\mathcal{D} = \mathcal{Z} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \cdots \times_n \mathbf{U}_n \cdots \times_N \mathbf{U}_N, \quad (3)$$

where $\mathcal{Z}$ is the core tensor, similar as the diagonal singular value matrix in conventional matrix SVD, and $\mathbf{U}_n$ is the mode-n matrix, containing the orthonormal vectors spanning the column space of matrix $\mathbf{D}_{(n)}$ resulting from the mode-n flattening of $\mathcal{D}$. The core tensor $\mathcal{Z}$ governs the interaction between the mode matrices $\mathbf{U_1}, \cdots, \mathbf{U}_N$.

We will embed the 3D X-Ray Transform of the DMI stacks into the multilinear analysis to learn the shape and texture basis for each of the classes. We will also compare this against using key-frames from each stack replacing the 3D X-Ray Transform output images.

2.3. X-Ray Based Tensor Texture Analysis

In the tensor texture analysis, we train the 3D X-Ray projections (as shown in Figures 4 and 5) in P poses (angle pairs) separately. We introduce the training process and recognition process step by step, as follows.

1. In the training process, the training data are DMI stacks from N classes, and for every class we pick M samples. Let $\mathbf{R}_{p,i}$ denote the 3D X-Ray Transform output image at the p_{th} pose (angle) for the i_{th} DMI stack. Every image is converted to a row vector. The training data of the 3D X-Ray Transform output images at the p_{th} pose is an $N \times M \times L_p$ tensor $\mathcal{D}_{\text{tex},p}$, where L_p is the length of $\mathbf{R}_{p,i}$.

2. The multilinear analysis of $\mathcal{D}_{\text{tex},p}$ can be expressed as

$$\mathcal{D}_{\text{tex},p} = \mathcal{Z}_{\text{tex},p} \times_1 \mathbf{U}_{\text{tex,class},p} \times_2 \mathbf{U}_{\text{tex,sample},p} \times_3 \mathbf{U}_{\text{tex,RT},p}, p = 1, ..., P, \quad (4)$$

where the $N \times M \times L_p$ core tensor $\mathcal{Z}_{\text{tex},p}$ governs the interaction between the factors represented in the 3 mode matrices: the $N \times N$ mode matrix $\mathbf{U}_{\text{tex,class},p}$ spans the space of class parameters (which is composed of the row vectors $\mathbf{c}^T_{\text{tex},p,n}, n = 1, ..., N$), the $M \times M$ mode matrix $\mathbf{U}_{\text{tex,sample},p}$ spans the space of sample parameters, while the $L_p \times (N \cdot M)$ mode matrix $\mathbf{U}_{\text{tex,RT},p}$ orthogonally spans the space of all training X-Ray projections at the p_{th} pose. Here the subscript "RT" denotes the 3D X-Ray Transform.

3. The multilinear method enables us to represent each class by the texture basis of the $N \times M \times L_p$ tensor

$$\mathcal{B}_{\text{tex},p} = \mathcal{Z}_{\text{tex},p} \times_2 \mathbf{U}_{\text{tex,sample},p} \times_3 \mathbf{U}_{\text{tex,RT},p}. \quad (5)$$

Then each class $n (n = 1, ..., N)$ can be represented by a coefficient vector $\mathbf{c}_{\text{tex},p,n}$ (size N), which is the row vector from the class matrix $\mathbf{U}_{\text{tex,class},p}$. For a certain sample m, we can get the basis tensor $\mathcal{B}_{\text{tex},p,m} (m = 1, ..., M)$ of size $N \times 1 \times L_p$. We flatten $\mathcal{B}_{\text{tex},p,m}$ along the class mode to get the tensor texture basis matrix $\mathbf{B}_{\text{tex},p,m}$ for the 3D X-Ray projections at the p_{th} pose.

4. In the recognition process, for a given DMI stack, we get its 3D X-Ray projections at those P defined poses (θ_p, ϕ_p) as $\mathbf{R'}_p$.

5. For each $\mathbf{R'}_p$, we project it to the corresponding tensor texture basis $\mathbf{B}_{\text{tex},p,m}$ to get the coefficient vector

$$\mathbf{c'}_{\text{tex},p,m} = \mathbf{B}^{-T}_{\text{tex},p,m} \mathbf{R'}_p, p = 1, ..., P. \quad (6)$$

6. For each $\mathbf{R'}_p$ we obtain the texture distance matrix

$$\mathbf{D}_{\text{tex},p}(m, n) = \|\mathbf{c'}_{\text{tex},p,m} - \mathbf{c}_{\text{tex},p,n}\| \quad (p = 1, ..., P; n = 1, ..., N; m = 1, ..., M). \quad (7)$$

By the nearest neighboring method, we find the index n_p that yields the smallest value of the texture distance for the p_{th} pose, and we identify $\mathbf{R'}_p$ as class n_p.

7. After we get n_p for all 3D X-Ray Transform output images $\mathbf{R'}_p (p = 1, ..., P)$, we do the majority voting among n_p to get the final recognition result for this testing DMI stack.

2.4. X-Ray Based Tensor Shape Analysis

The analysis of the shape needs to be incorporated into the texture, since the discriminating information is contained in both texture and shape features. We show how to compute the shape feature in the multilinear framework.

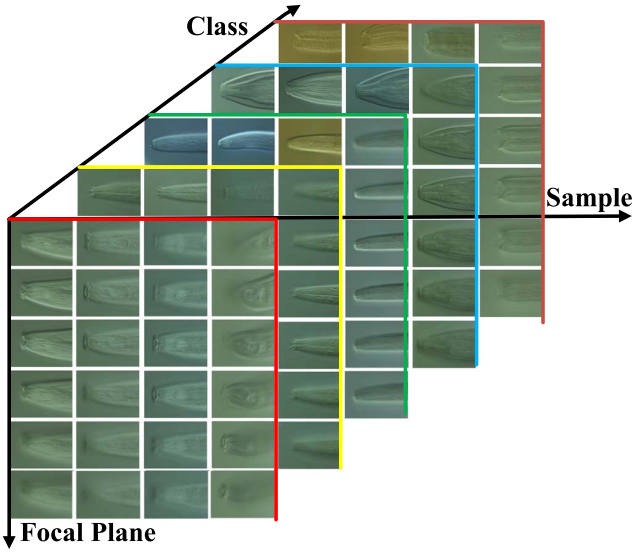

Figure 6. The nematode DMI stack database, including 5 classes of nematodes, and each class includes 10 samples (species). Each sample is represented by a DMI stack containing more than 100 focal planes (frames) taken from the top of the specimen to the bottom. Only 6 key frames from different focal planes of every stack are shown. They are 20 samples from each of the 5 classes (4 samples are picked for each class for illustration purposes).

In the multilinear shape analysis, the training data are the shape vectors, which can be extracted by the method introduced in [6]. There the authors defined the preshape as the geometric information that remains when location and scale effects are filtered out. A centered preshape is obtained by subtracting the mean from the shape vector and then scaling to norm one. Mathematically, if the original shape vector is denoted by X, the centered preshape is given by

$$S = \frac{CX}{\|CX\|}, \text{where} \quad C = I_k - \frac{1}{k} 1_k 1_k^T, \tag{8}$$

where I_k is a $k \times k$ identity matrix and 1_k is a k-dimensional vector of ones. To get shape vector from preshape, we need to remove the rotational component. Since our images are all aligned, we do not need to compensate for the in-plane rotation and hence can work with the preshape.

After we get the preshape vectors $\mathbf{S}_{p,i}$ from the X-Ray Transform output images $\mathbf{R}_{p,i}$, we train the extracted shape tensor in P poses similarly as in the tensor texture case.

1. The training shape tensor, $\mathcal{D}_{\text{shp},p}$, at the p_{th} pose is a $N \times M \times J_p$ tensor (J_p is the length of the preshape vectors $\mathbf{S}_{p,i}$ extracted from the 3D X-Ray projections at the p_{th} angle for the i_{th} DMI stack.)

2. The shape tensor analysis yields

$$\mathcal{D}_{\text{shp},p} = \mathcal{Z}_{\text{shp},p} \times_1 \mathbf{U}_{\text{shp,class},p} \times_2 \mathbf{U}_{\text{shp,sample},p} \times_3 \mathbf{U}_{\text{shp,RT},p}, p = 1, ..., P, \tag{9}$$

where the $N \times M \times J_p$ core tensor $\mathcal{Z}_{\text{shp},p}$ governs the interaction between the factors represented in the 3 mode matrices: $\mathbf{U}_{\text{shp,class},p}$ spans the space of class parameters (which is composed of the row vector $\mathbf{c}^T_{\text{shp},p,n}$), $\mathbf{U}_{\text{shp,sample},p}$ spans the space of sample parameters, while the $J_p \times (N \cdot M)$ mode matrix $\mathbf{U}_{\text{shp,RT},p}$ orthogonally spans the space of all training shape vectors $\mathbf{S}_{p,i}$ extracted from the 3D X-Ray Transform output images $\mathbf{R}_{p,i}$ at the p_{th} pose.

3. The shape basis of the $N \times M \times J_p$ shape tensor is

$$\mathcal{B}_{\text{shp},p} = \mathcal{Z}_{\text{shp},p} \times_2 \mathbf{U}_{\text{shp,sample},p} \times_3 \mathbf{U}_{\text{shp,RT},p}. \tag{10}$$

For a certain sample m, we get the shape basis tensor $\mathcal{B}_{\text{shp},p,m} (m = 1, ..., M)$ of size $N \times 1 \times J_p$. We flatten it along the class mode to get shape basis matrix $\mathbf{B}_{\text{shp},p,m}$ for the 3D X-Ray projections at the p_{th} pose.

4. In the recognition process, for a given DMI stack, we get its 3D X-Ray Transform output images $\mathbf{R}'_p$, $p = 1, ..., P$, and extract the shape vector $\mathbf{S}'_p$ for each $\mathbf{R}'_p$.

5. We project each $\mathbf{S}'_p$ to the corresponding tensor shape basis $\mathbf{B}'_{\text{shp},p,m}$ to get the coefficient vector

$$\mathbf{c}'_{\text{shp},p,m} = \mathbf{B}^{-T}_{\text{shp},p,m} \mathbf{S}'_p, p = 1, ..., P. \tag{11}$$

6. We obtain the shape distance matrix

$$\mathbf{D}_{\text{shp},p}(m, n) = \|\mathbf{c}'_{\text{shp},p,m} - \mathbf{c}'_{\text{shp},p,n}\| \tag{12}$$

for $\mathbf{S}'_p$. By nearest neighboring method we find the index n_p with the smallest value of the shape distance for the p_{th} pose, and we identify $\mathbf{S}'_p$ as class n_p.

7. After we get the recognition results n_p for all the 3D X-Ray Transform output images $\mathbf{R}'_p (p = 1, ..., P)$, we do majority voting among n_p to get the final recognition result for this testing DMI.

3. Experimental Results

The data for our experiment is the nematode DMI stacks, containing 5 classes and each class containing 10 samples (as the DMI stacks shown in Figure 6). Each image stack consists of more than 100 multi-focal frames obtained from a Differential Interference Contrast microscope [1], which is a type of light microscopy where contours are enhanced compared to "plain" bright-field microscopy with little or no loss of resolution (unlike phase contrast) and it does

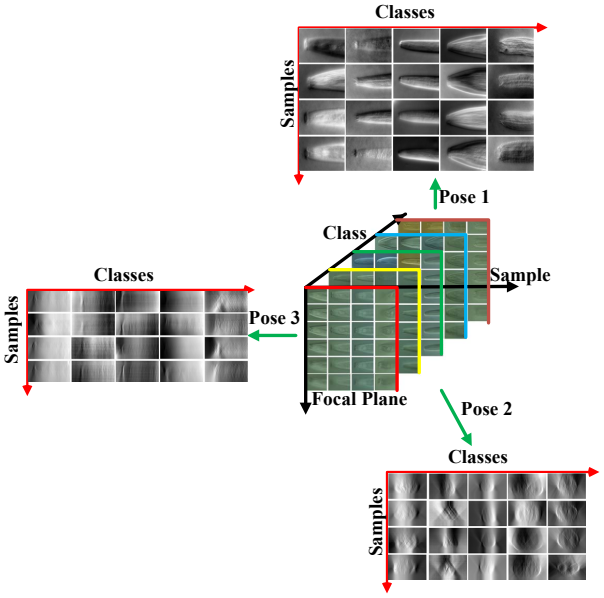

Figure 7. 3D X-Ray Transform projections at 3 poses, $(\pi/2, \pi/2)$, $(0, \pi/2)$ and (π, π), for the training DMI stacks.

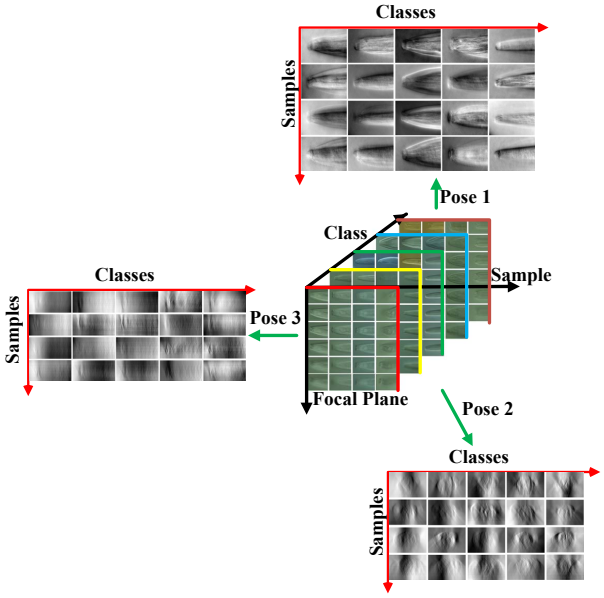

Figure 8. Tensor texture bases of the 3D X-Ray Transform projections at 3 poses, $(\pi/2, \pi/2)$, $(0, \pi/2)$ and (π, π), for the training DMI stacks.

not involve fluorescent excitation from laser sources (unlike confocal). All nematode images are registered to have the same mouth location. In the experiments, we just use one channel data of the original color images for feature extraction and classification, and the image intensity is normalized to remove the affection of illumination. For every class, 4 samples (DMI stacks) are for training and the rest are for testing. We select $(\pi/2, \pi/2)$, $(0, \pi/2)$, (π, π), $(\pi/5, \pi/3)$ and $(\pi/3, \pi/5)$ as the 5 poses (angles) of the 3D X-Ray Transform, and the 3D X-Ray Transform output images at the first 3 poses of those training DMIs are shown in Figure 7. We can see that different angles capture very different characteristics of the image stacks. Since every DMI stack contains different numbers of images, their 3D X-Ray Transform output images are not the same size (even the ratios of image length to width are different). So we have to normalize them to be the same size for the X-Ray projections at the same pose.

3.1. Tensor Texture Analysis

In the tensor texture analysis, we compute the basis for each direction separately, because the images at different angles have different ratios of image length to width. The images at 3 poses, $(\pi/2, \pi/2)$, $(0, \pi/2)$ and (π, π), for the training nematode DMI stacks are shown in Figure 7. The training tensor data at those five poses are of sizes $5 \times 4 \times 97200$, $5 \times 4 \times 48600$, $5 \times 4 \times 64800$, $5 \times 4 \times 196245$ and $5 \times 4 \times 196245$, respectively.

In the tensor texture training step, for the X-Ray Transform output images $\mathbf{R}_{p,i}$ at the $p_{th}(p = 1, ..., 5)$ pose, we get, from Equation (4), the core tensor $\mathcal{Z}_{\text{tex},p}$, the 5×5 mode matrix $\mathbf{U}_{\text{tex,class},p}$ spanning the class space (which is composed of the row vector $\mathbf{c}^T_{\text{tex},p,n}, n = 1, ..., 5$), the 4×4 mode matrix $\mathbf{U}_{\text{tex,sample},p}$ spanning the sample space, and the mode matrix $\mathbf{U}_{\text{tex,RT},p}$ spanning the space of all training X-Ray Transform output images at the p_{th} pose. From Equation (5), we get the texture basis tensor $\mathcal{B}_{\text{tex},p,m}$ for the m_{th} sample $(m = 1, .., 4)$, which we flatten along the class mode to get the texture basis matrix $\mathbf{B}_{\text{tex},p,m}$. The visualization of those bases is shown in Figure 8.

Given a testing DMI stack, first we get its 3D X-Ray Transform output images $\mathbf{R}'_p$ at different poses. Then we project them to the corresponding texture basis to get the coefficient vector and obtain the texture distance matrix $\mathbf{D}_{\text{tex},p}(m, n)(m = 1, ..., 4; n = 1, ..., 5)$. Through the nearest neighboring criteria we find n_p that yields the smallest value of the texture distance for the p_{th} pose. Finally by majority voting among $n_p(p = 1, ..., 5)$ we get the recognition result for this testing DMI stack. This 3D X-Ray Transform based multilinear texture analysis method obtains a recognition rate of 87%. The recognition performance of this multilinear texture analysis is shown in Figure 9.

3.2. Tensor Shape Analysis

For the tensor shape analysis, not all 3D X-Ray Transform images have prominent shape information. From Figures 4-5 and Figures 7, we can see that 3D X-Ray Transform outputs at poses, $(\pi/2, \pi/2)$, (π, π), $(\pi/5, \pi/3)$ and $(\pi/3, \pi/5)$, can be used to extract the shape information,

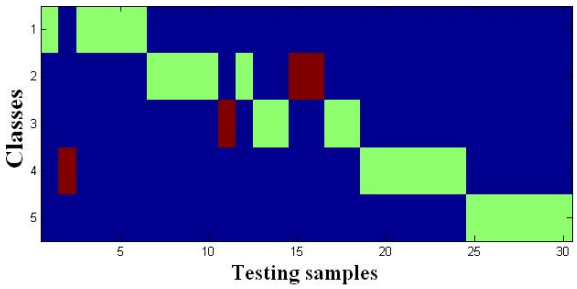

Figure 9. Multilinear texture recognition performance on the 3D X-Ray Transform output images of 30 nematode testing DMI sample stacks. In the vertical direction, numbers 1-5 denote the 5 nematode classes. In the horizontal direction, numbers 1-6 denote the 6 testing samples from class 1, 7-12 from class 2, 13-18 from class 3, 19-24 from class 4 and 25-30 from class 5. The green color denotes correct classification results, and the red color denotes wrong classification results (as viewed on a color monitor).

Figure 10. Extracted shape contours from the 3D X-Ray Transform output images at the pose $(\pi/2, \pi/2)$ for 5 DMI stacks (one DMI stack picked from each class).

while the images from the 2nd pose $(0, \pi/2)$, do not have much shape information. Some extracted shape contours at the pose $(\pi/2, \pi/2)$ are shown in Figure 10. Similar to the above description, we do the multilinear shape analysis and get the recognition performance shown in Figure 11.

3.3. Combination of Tensor Texture and Shape

We combine the multilinear texture and shape analysis to build a more accurate classifier. Specifically, for a given DMI stack, through the 3D X-Ray Transform at the p_{th} pose, we get $\mathbf{R'}_p(p = 1, ..., 5)$ and $\mathbf{S'}_p(p = 1, 3, 4, 5)$. After projecting them to the corresponding tensor texture basis and shape basis, we obtain the texture distance matrix $\mathbf{D}_{\text{tex},p}(m, n)(m = 1, ..., 4; n = 1, ..., 5)$ for $\mathbf{R'}_p$ $(p = 1, ..., 5)$, and the shape distance matrix $\mathbf{D}_{\text{shp},p}(m, n)$ for $\mathbf{S}_p$ $(p = 1, 3, 4, 5)$. When $p = 2$, $\mathbf{D}_{\text{shp},p}(m, n) = 1$, because we are not using the shape information at this pose. We combine the shape and texture distance matrices as

$$\mathbf{D}_p(m, n) = \mathbf{D}_{\text{tex},p}(m, n) \times \mathbf{D}_{\text{shp},p}(m, n), \qquad (13)$$

where $\mathbf{D}_p$ is the combined distance matrix for the 3D X-Ray Transform output image $\mathbf{R'}_p$ at the p_{th} pose. Then we identify $\mathbf{R'}_p$ as class n_p through the nearest neighbor method for all poses, and after the majority voting among $n_p(p = 1, ..., 5)$ we finally get the recognition result for this given DMI stack.

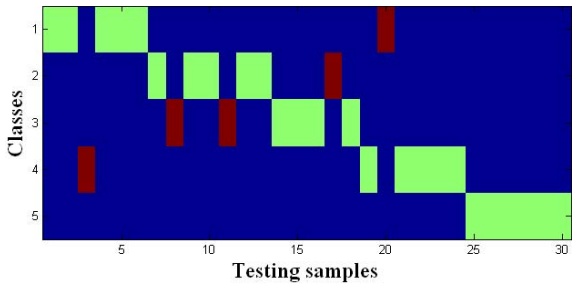

Figure 11. Multilinear shape recognition performance on the 3D X-Ray Transform output images of nematode DMI stacks of 30 testing samples. In the vertical direction, numbers 1-5 denote the 5 class labels. In the horizontal direction, numbers 1-6 denotes the 6 testing samples from class 1, 7-12 from class 2, 13-18 from class 3, 19-24 from class 4 and 25-30 from class 5. The green color denotes correct results, and the red color denotes wrong results.

Classifier	Recognition Rate
PCA	60%
ICA	57%
Gabor Texture	60%
Tensor Texture	87%
Tensor Shape	83%
Key Frames Tensor	73%
Tensor Texture and Shape 1	94%
Tensor Texture and Shape 2	97%
Tensor Texture and Shape 3	100%

Table 1. Recognition rate comparison between different classification methods. The 4th-7th rows are the recognition rates on the training data composed of 4 samples from each class, the 8th row is the recognition rate on the training data of another 4 samples from each class, while the 9th row is the recognition rate on the training data of 5 samples from each class. The training and testing data for the first 3 rows are described in the text.

The combination of the tensor texture and shape information improves the recognition rate, which is shown in the recognition performance comparison table (Table 1) between different classification methods. The training data for the classifiers using PCA and Independent Components Analysis (ICA) [3] is 20 key frames, coming from those 20 DMI stacks used for the MXRT training (one DMI stack, one key frame), and the testing data is 30 key frames from the 30 testing DMI stacks for the MXRT testing. The classification based on Gabor texture is also done for those 50 key frames (from 50 DMI stacks) [10]. The classification performance of those 3 methods is also listed in Table 1.

We have done the classification by changing the training data and testing data, and got similar recognition rate. Also, we increased the amount of the training data, and got slightly better recognition results. These results are summarized in Table 1.

3.4. Tensor Analysis on Key Frames

Given stacks of multifocal image frames, an alternative and intuitive way for classification is to do the tensor analysis on the original image frames. So here we give a short description of the multilinear analysis on the key image frames (i.e., some selected focal planes) from every DMI stack, and the recognition performance is also listed in Table 1.

In the training step, we pick the same number of the most typical frames as the key frames from the same focal plane (for example, 6 frames for each stack), as in Figure 6. Then the training data is a $5 \times 4 \times 6 \times 97200$ tensor. After the multilinear SVD analysis as introduced in the previous section [12], we get

$$\mathcal{D} = \mathcal{Z} \times_1 \mathbf{U}_{\text{class}} \times_2 \mathbf{U}_{\text{sample}} \times_3 \mathbf{U}_{\text{frame}} \times_4 \mathbf{U}_{\text{pixel}}, \tag{14}$$

with the 5×5 mode matrix $\mathbf{U}_{\text{class}}$ spanning the class space (which is composed of the row vector $\mathbf{c}_{\text{tex},n}^T, n = 1, ..., 5$), the 4×4 mode matrix $\mathbf{U}_{\text{sample}}$ spanning the sample space, the 6×6 mode matrix $\mathbf{U}_{\text{frame}}$ spanning the focal plane space, and the mode matrix $\mathbf{U}_{\text{pixel}}$ orthogonally spanning the space of all training key frames.

The basis for different classes of this training tensor is

$$\mathcal{B} = \mathcal{Z} \times_2 \mathbf{U}_{\text{sample}} \times_3 \mathbf{U}_{\text{frame}} \times_4 \mathbf{U}_{\text{pixel}}. \tag{15}$$

Then each class $n (n = 1, ..., 5)$ can be represented by a coefficient vector of size 5, $\mathbf{c}_n$, which is the row vector from the class matrix $\mathbf{U}_{\text{class}}$. For a certain sample m and a certain frame f, we can get the basis tensor $\mathcal{B}_{m,f} (m = 1, ..., 4; f = 1, ..., 6)$ of size $5 \times 1 \times 1 \times 97200$. We flatten $\mathcal{B}_{m,f}$ along the class mode to get the basis matrix $\mathbf{B}_{m,f}$.

In the recognition process, given a testing image $\mathbf{T}$, we project it onto those bases and get the coefficients for every (m, f) pair, i.e., $\mathbf{c}_{m,f} = \mathbf{B}_{m,f}^{-T}\mathbf{T}$. Then we obtain the 3rd order distance matrix $\mathbf{D}(m, f, n) = \|\mathbf{c}_{m,f} - \mathbf{c}_n\|$, and by the nearest neighboring method, we find the index n that yields the smallest value of the distance matrix, and we identify this testing image $\mathbf{T}$ as class n.

4. Conclusions

In this paper, we presented a 3D X-Ray Transform based multilinear feature extraction and classification method for Digital Multi-focal Images. Given a stack of DMI images, different images contain different information. Direct image processing of the single image frames can only provide 2D image plane information, while the 3 dimensional variation information across different image frames will be lost. However, the 3D structure information of a multifocal image stack can be captured through the 3D X-Ray Transform at different angle views. There are multiple factors that need to be analyzed - shape, texture, viewpoint, different instances within the same class and different classes of specimens. For this purpose, we embed the 3D X-Ray Transform within a multilinear framework. By combining the tensor texture and shape information we can make a better recognition rate than just relying on the key frames of DMI stacks. The experimental results on the nematode DMI data show that the proposed 3D X-Ray Transform based multilinear analysis method can effectively improve the recognition rate from 60% (PCA) to 100%.

References

[1] D. G. Allen, R.D. and G. Nomarski. The Zeiss-Nomarski differential interference equipment for transmitted light microscopy. *Z. Wiss. Mikrosk. Mikrosk. Tech.*, 69:193–221, 1969.

[2] P. G. Challenor, P. Cipollini, and D. Cromwell. Use of the 3D Radon Transform to Examine the Properties of Oceanic Rossby Waves. *Journal of Atmospheric and Oceanic Technology*, 18(9):1558–1566, 2001.

[3] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, 1994.

[4] P. De Ley and W. Bert. Video capture and editing as a tool for the storage, distribution, and illustration of morphological characters of nematodes. *Journal of Nematology*, 34:296–302, 2002.

[5] S. Deans. *The Radon Transform and Some of Its Applications*. Krieger Publishing Company, 1983.

[6] I. Dryden and K. Mardia. *Directional Statistics and Shape Analysis*. John Wiley and Sons, 1998.

[7] Eyualem-Abebe, J. G. Baldwin, B. Adams, D. Hope, S. Gardner, R. Huettel, P. Mullin, T. Powers, J. Sharma, W. Ye, and W. K. Thomas. A position paper on the electronic publication of nematode taxonomic manuscripts. *Journal of Nematology*, 38(3):305–311, 2006.

[8] I. Laptev and T. Lindeberg. Space-time interest points. In *IEEE International Conference on Computer Vision*, volume 1, pages 432–439, 2003.

[9] L. D. Lathauwer, B. D. Moor, J. Vandewalle, and J. V. A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1253–1278, 2000.

[10] B. Manjunath and W. Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):837–842, 1996.

[11] F. Natterer. *The Mathematics of Computerized Tomography*. Wiley, 1986.

[12] M. A. O. Vasilescu and D. Terzopoulos. Multilinear subspace analysis of image ensembles. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 93–99, 2003.

[13] A. Veeraraghavan, A. K. Roy-Chowdhury, and R. Chellappa. Matching shape sequences in video with applications in human movement analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(12):1896–1909, 2005.

[14] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003.