CLASSIFICATION OF MULTI-FOCAL NEMATODE IMAGE STACKS USING A PROJECTION BASED MULTILINEAR APPROACH

Min Liu [a], *Xueping Wang* [a], *Xiaoyan Liu* [a], *and Hongzhong Zhang* [b]

[a] College of Electrical and Information Engineering, Hunan University, Hunan, 410082, China
[b] Department of Industrial Engineering & Operations Research, Columbia University, USA

ABSTRACT

In this paper, we propose to use projection methods such as coefficient of variation projection (COV) to exploit the entire information of Digital Multi-focal Images (DMI) using its projection images along different directions. The COV projection takes into account the intensity distribution feature of multi-focal images, so it overcomes the limitation of poor contrast of the projection images from the 3D X-Ray Transform, which is used in a previous work. Because the DMI stacks represent the effect of different factors - texture, projection directions, different instances within the same class and different classes of objects, we embed the projection method within a multilinear classification framework. The experimental results on the nematode data show that the image projection based multilinear classifier can achieve very reliable recognition rate (95.5%), even we only use the texture feature instead of the combination of texture and shape features as in the previous work.

Index Terms—Multilinear analysis, Digital Multi-focal Images, Projection, Image classification

1. INTRODUCTION

In biomedical field, Digital Multi-focal Images (DMI) is a new way for documentation and communication of specimen data [1]. In this approach, morphological information for a specimen can be captured in the form of a stack of high-quality images, representing individual focal planes through the specimen's body. Fig. 1 shows a few nematode image stacks taken from a differential interference contrast microscope. Each stack contains multiple focal planes taken from the top to the bottom of the specimen, with only a few frames of each shown.

Given such image stacks containing so many multi-focal images, how do we efficiently extract effective features from all layers to classify the images is still a problem. Existing 3D feature extraction methods like 3D scale-invariant feature transform (SIFT) and 3D histogram of oriented gradient (HOG) [11, 12] are not suited for this purpose, because images in multiple focal planes have different characteristics compared to a space-time volume [4, 5].

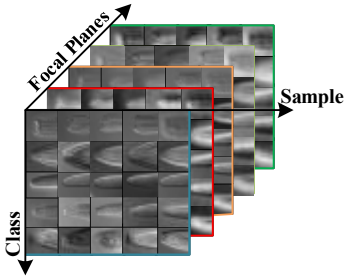

Fig. 1. A few samples of DMI image stacks taken from different nematode species. Each stack contains images of multiple focal planes, with only a few frames of each shown.

In a previous work, a 3D X-Ray projection based classifier is applied in the classification of nematodes - a species which is very difficult to classify since they are one of the most numerous animals on earth [1, 7] and part of them are very similar to each other. On one hand, the projection methods allow us to analyze the multi-focal images from different projection directions - something direct analysis of the images in each DMI stack would not allow. On the other hand, for the classification of DMI stacks, there are multiple factors that need to be analyzed - texture, projection directions, different instances within the same class and different classes of specimens. The multilinear framework allows us to model the inter-relationships between different variables, while at the same time allowing independent analysis along each dimension.

It is confirmed that the 3D X-Ray Transform based multilinear classifier achieves reliable recognition rate for nematode DMI images [7]. However, the nematode images of different focal planes are nearly transparent, resulting very poor contrast of the projection images from 3D X-Ray Transform. In fact, the pixel intensities inside the specimen vary when focus is changing, while the background intensities stay more constant throughout the z-direction, resulting in relatively higher variations inside the specimen, while much lower variations outside [9].

In this paper, instead of using the 3D X-Ray Transform, we propose to use projection methods such as standard deviation projection (STD), which could take into account

This work was supported in part by the National Natural Science Foundation of China under Grants 61301254 and 61374149.

the pixel intensity distribution feature to improve the quality of the projection images. After that the projection images of DMI stacks along multiple directions are embedded into a multilinear classification framework. Because the texture information is more robust than the shape information in most of the nematode images, in the experiment, we demonstrated the strength of the proposed method by using the texture information only, while the previous work used the combination of texture and shape features. Moreover, we gave a comparison of the performance of the multilinear classifier based on different projection methods. Finally, we evaluated the effect of the projection directions on the classification performance.

2. METHODOLOGY

2.1. Projection Methods

In this paper, we mainly introduce 3D X-Ray Transform, STD and COV projection, the description of other projection methods such as maximum intensity projection (MIP) can be found in [9].

2.1.1. 3D X-Ray Transform

The 3D X-Ray Transform is the object's projection on a plane, whose orientation is identified by a pair of angles (θ,ϕ) shown in Fig. 2 (a). It can be 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 2D image plane and the transformation of those coordinates is expressed as below,

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

2.1.2. STD Projection

From the previous work, we know that the contrast of the 3D X-Ray projection images is not very well [7]. The standard deviation projection could enhance the contrast of the projection images by increasing the intensity differences between nematode and background areas [9]. Since there is typically less z intensity variation in the background than in nematodes, these two classes of pixels can be clearly separated. STD projection image is constructed by calculating the standard deviation of intensities in the z'-direction for each pixel of the original stack:

$$I'(x',y')=\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(I_i(x',y')-\mu(x',y')\right)^2} \tag{3}$$

where I_i is the i_{th} slice in the image stack, I' is the output image of STD projection, μ is the mean of all focal plane images, N is the total number of focal plane images in image stacks, x' and y' are the coordinates in the transformed image plane as expressed in Equation (2), similar as in the 3D X-Ray Transform.

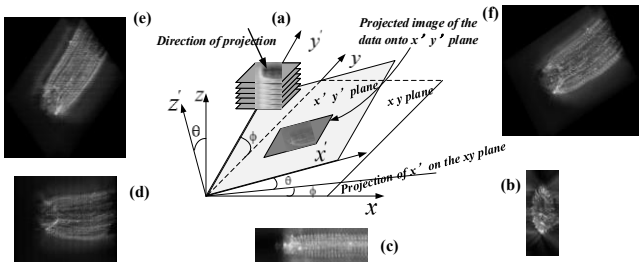

Fig. 2. (a) Image projection along different directions; (b)-(f) COV projection images along different directions, (b) (π/2, 0), (c) (π/2, π/2), (d) (0, 0), (e) (π/3, π/3) and (f) (π/4, π/4).

Because the background pixel intensities at the same position of all focal planes are similar, the nonuniform background is efficiently removed by the STD projection.

2.1.3. COV Projection

COV projection is defined as a transform of STD projection:

$$I''(x',y')=\frac{I'(x',y')}{\mu(x',y')} \tag{4}$$

where I'' is the output image of COV transform. Fig. 2 shows the COV projection images from different angles.

The projection images of the same DMI stack by using different projection methods are shown in Fig. 3, from which we could see that the contrast of the images from the STD and COV projection is better than the images from the other methods. The projection images of DMI stacks are embedded into the multilinear framework for classification.

2.2. Multilinear Analysis

In the recognition of natural face images formed by different factors, such as expression, illumination condition, pose and so on [8, 13, 14], the collection of facial images is represented by separating different modes underlying the formation of facial images by a high-order generalization of principal component analysis and high-order singular value decomposition (HOSVD) [6, 10]. This idea is applied to the classification of DMI image stacks, since they also represent the combination of various factors [7].

Multilinear algebra, the algebra of higher-order tensor, offers a potent mathematical framework for analyzing ensembles of images resulting from the interaction of any number of underlying factors. For and N$^{\text{th}}$ order tensor $\mathcal{T}\in\mathbb{R}^{I_1*\ldots*I_n*\ldots*I_N}$ (its elements are denoted as $\mathcal{T}_{i_1 i_2\ldots i_n\ldots i_N}$), its mode-n vectors are I_n-dimensional vectors obtained from $\mathcal{T}$ by varying index i_n while keeping the other indices fixed. It is an extension of SVD that orthogonalizes

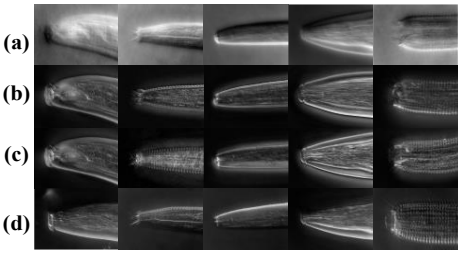

Fig. 3. Projection images of the 5 stacks in Fig.1, by different methods. (a) 3D X-Ray Transform. (b) STD. (c) COV. (d) MIP.

these N spaces and expresses the tensor as the mode-n product, denoted $\times_n$, of N-orthogonal spaces as

$$\mathcal{T} = \mathcal{Z} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \cdots \times_n \mathbf{U}_n \cdots \times_N \mathbf{U}_N \qquad (5)$$

where $\mathcal{Z}$ is the core tensor, and $\mathbf{U}_n$ is the mode-n matrix, containing the orthonormal vectors spanning the column space of matrix $\mathcal{T}_{(n)}$ resulting from the mode-n flattening of $\mathcal{T}$ [10].

The N-mode SVD algorithm for decomposing $\mathcal{T}$ according to Equation (5) is as follows:

1) For $n = 1, \cdots, N$, compute matrix $\mathbf{U}_n$ in Equation (5) by computing the SVD of the flattened matrix $\mathcal{T}_{(n)}$ and setting $\mathbf{U}_n$ to be the left matrix of the SVD.
2) Solve for the core tensor as follows:

$$\mathcal{Z} = \mathcal{T} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \cdots \times_n \mathbf{U}_n^T \cdots \times_N \mathbf{U}_N^T \qquad (6)$$

2.3. Proposed Method

The pipeline of the projection based multilinear classifier using Gabor feature is shown in Fig. 4. Given R projection directions, the training and testing steps are listed below.

1) The training data are DMI stacks from J classes, and for each class we pick K samples. Let $\mathbf{P}_{r,i}$ denote the projection image at the r_{th} angle for i_{th} DMI stack.

2) For each projection image $\mathbf{P}_{r,i}$, it is filtered by a $u \times v$ Gabor filter, where u is the number of scales and v is the number of orientations. After that, each image from the Gabor filter is downsampled by a $d_1 \times d_2$ sized template, and then converted to a row vector. Next, we combine all row vectors into one single row vector whose length is $L_r = \frac{w \times h \times u \times v}{d_1 \times d_2}$, where w and h are the sizes of each COV projection image. The training data of Gabor filtered images along the r_{th} direction is a $J \times K \times L_r$ tensor $\mathcal{T}_{\text{tex},r}$.

3) The multilinear analysis of $\mathcal{T}_{\text{tex},r}$ can be expressed as

$$\mathcal{T}_{\text{tex},r} = \mathcal{Z}_{\text{tex},r} \times_1 \mathbf{U}_{\text{tex,class},r} \times_2 \mathbf{U}_{\text{tex,sample},r} \times_3 \mathbf{U}_{\text{tex,PR},r} \qquad (7)$$

where the $J \times K \times L_r$ core tensor $\mathcal{Z}_{\text{tex},r}$ governs interaction between the factors represented 3 mode matrices: $J \times J$ mode matrix $\mathbf{U}_{\text{tex,class},r}$ spans the space of class parameters (which is composed of the row vectors $\mathbf{c}_{\text{tex},r,j}^T, j = 1, \cdots, J$), the $K \times K$ mode matrix $\mathbf{U}_{\text{tex,sample},r}$ spans the space of sample parameters, while the $L_r \times (J \cdot K)$ mode matrix $\mathbf{U}_{\text{tex,PR},r}$ orthogonally spans the space of all training projection images at r_{th} direction. "PR" denotes the projection methods.

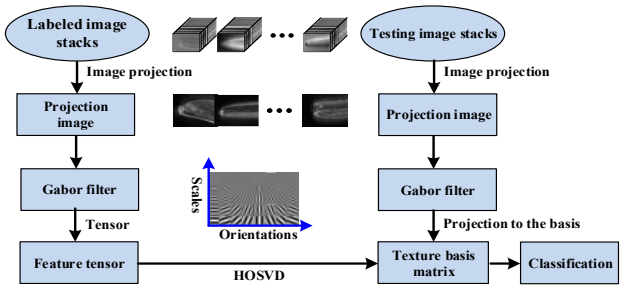

Fig. 4. The diagram of the proposed classification method.

4) The multilinear method enables us to represent each class by the texture basis of the $J \times K \times L_r$ tensor

$$\mathcal{B}_{\text{tex},r} = \mathcal{Z}_{\text{tex},r} \times_2 \mathbf{U}_{\text{tex,sample},r} \times_3 \mathbf{U}_{\text{tex,PR},r}. \qquad (8)$$

Then each class j $(j = 1, \cdots, J)$ can be represented by a coefficient vector $\mathbf{c}_{\text{tex},r,j}$ (size J), which is the row vector from the class matrix $\mathbf{U}_{\text{tex,class},r}$. For a certain sample k, we get the basis tensor $\mathcal{B}_{\text{tex},r,k}$ (size $J \times 1 \times L_r$). We obtain the tensor texture basis matrix $\mathbf{B}_{\text{tex},r,k}$ $(k = 1, \cdots, K)$ for the projections in the r_{th} direction by flattening $\mathcal{B}_{\text{tex},r,k}$ along the class mode.

5) In the recognition process, for a given DMI stack, we extract the Gabor feature $\mathbf{G}_r'$ from its projection images in the r_{th} direction, and then project it to the corresponding tensor texture basis matrix $\mathbf{B}_{\text{tex,r,k}}$ to get the coefficient vector

$$\mathbf{c}'_{\text{tex},r,k} = \mathbf{B}_{\text{tex},r,k}^{-T} \mathbf{G}_r'. \qquad (9)$$

6) For each $\mathbf{P}'_r$, we obtain the texture distance matrix

$$\mathbf{D}_{\text{tex},r}(k, j) = \| \mathbf{c}'_{\text{tex},r,k} - \mathbf{c}_{\text{tex},r,j} \| \qquad (10)$$

where k=1,⋯, K; j=1,⋯, J. Then the distance matrices from different directions are combined as a final distance matrix

$$\mathbf{D}_{\text{tex}}(k, j) = \sum_r \mathbf{D}_{\text{tex},r}(k, j) \qquad (11)$$

By the nearest neighboring method, for the testing image stack, we can find its most similar sample among the training dataset which yields the smallest value of the texture feature distance matrix $\mathbf{D}_{\text{tex}}$, and it is recognized as class j_c, which is the class label of this sample closest to the feature vector of the given testing image stack.

3. EXPERIMENTAL RESULTS AND DISCUSSION

The data for our experiment is a set of 500 nematode image stacks from 10 classes. Each class contains 50 samples and each stack consists of more than 100 multi-focal images with size 270×360 pixels. Based on a cross-validation rule, half of samples (DMI stacks) for each nematode class are randomly chosen for training and the rest for testing. We repeat each experiment 100 times, and the recognition rate is computed as an average of the results.

Previously, the 3D X-Ray Transform based multilinear classifier used the combination of texture and shape features [7]. In this paper we only use Gabor texture feature to demonstrate the strength of the proposed method. The Gabor feature is extracted with 5 scales and 8 orientations, and then downsampled by a factor of 6. From Table 1, we can see that the classifiers with feature extracted from projection images perform better than with feature information directly extracted from the image stacks such as 3D HOG (70.5%) and 3D SIFT (74.2%) [11, 12].

We first demonstrate the strength of the image projection methods, by choosing the projection images along direction (0, 0) as the input of the classifiers. In Table 1, we compared the recognition rates of the same classifiers with projection images and non-projection images as the input. For the projection based classifiers, both the training and testing images are the projection images; while for the non-projection based classifiers, the training and testing images are both key frame images selected manually. The classifiers used in our experiments are multiclass SVM (MSVM) [3], dictionary pair learning classification (DPL) [4], and collaborative representation classification (CRC) [2]. From Table 1 we can see that the same classifiers with projection images perform better than with non-projection images.

Furthermore, we compared the performance of the multilinear classifier with the other classifiers, with the projection images as the input. From Table 1, we can see that the multilinear classifiers reach higher classification rates than those of the non-multilinear classifiers. Table 1 has also compared the recognition rates of the multilinear classifiers based on different projection methods, such as 3D X-Ray Transform, MIP, STD and COV. We can see that the COV projection based method achieved the best recognition rate of 95.5%, much higher than the rate of 88.7% in the previous work [7]. In the later analysis, we choose the COV projection based classification method for further discussion.

In order to evaluate the effect of the projection directions on the classification performance, we have made a comparison of the recognition rates between 7 different projection directions by COV methods in Table 2, from which we can see that the recognition rates using projection images along different angles are different. So we combine the information from the projection images along different directions to make a more reliable classification decision. In Table 3, we have listed the recognition rates of the combination of different projection directions. We can see that the recognition rates of the combination of two or more directions are slight better than that of one direction.

Table 1. Recognition Rate of Different Methods

Classifier	Recognition Rate
MSVM + HOG	63.7%
MSVM + SIFT	66.3%
MSVM + Key Frame	71.0%
MSVM + COV Projection	82.1%
DPL + Key Frame	75.2%
DPL + COV Projection	85.3%
CRC + Key Frame	67.0%
CRC + COV Projection	87.2%
Multilinear + HOG	70.5%
Multilinear + SIFT	74.2%
Multilinear Gabor Texture + Key Frame	91.9%
Multilinear Gabor Texture + 3D X-RAY	88.7%
Multilinear Gabor Texture + MIP	92.5%
Multilinear Gabor Texture + STD	94.7%
Multilinear Gabor Texture + COV	95.5%

Table 2. Recognition Rate of COV Based Classifier from Different Directions

Projection Direction	Recognition Rate
(0,0)	95.5%
($\pi/3$,$\pi/5$)	94.2%
($\pi/4$,$\pi/4$)	93.6%
($\pi/6$,$\pi/6$)	93.9%
($\pi/5$,$\pi/3$)	91.7%
($\pi/2$,$\pi/2$)	88.5%
($\pi/3$,$\pi/3$)	94.9%

Table 3. Recognition Rate of the Combination of COV Projections along Different Directions

Projection Directions	Recognition Rate
(0, 0) and ($\pi/4$, $\pi/4$)	97.0%
(0, 0), ($\pi/5$, $\pi/3$) and ($\pi/3$, $\pi/5$)	97.6%
(0, 0), ($\pi/4$, $\pi/4$) and ($\pi/3$, $\pi/5$)	98.0%
(0, 0), ($\pi/4$, $\pi/4$), ($\pi/3$, $\pi/5$) and ($\pi/3$, $\pi/3$)	98.2%
(0, 0), ($\pi/4$, $\pi/4$), ($\pi/3$, $\pi/5$), ($\pi/3$, $\pi/3$) and ($\pi/5$, $\pi/3$)	97.2%

4. CONCLUSIONS

In this paper, we present a projection based multilinear feature extraction and classification method for multi-focal images. The proposed method can effectively exploit the information in the image stack using projections along different directions, and model the inter-relationships between different variables, such as texture, projection directions, different instances within the same class and different classes of specimens. The experimental results on the nematode data show that the projection based multilinear analysis method can reach very reliable recognition rate.

5. REFERENCE

[1] De Ley, Paul, and Wim Bert, "Video capture and editing as a tool for the storage, distribution, and illustration of morphological characters of nematodes," *Journal of Nematology*, vol. 34, no. 4, pp.296-302, 2002.

[2] P. Zhu, L. Zhang, Q. Hu, and Simon CK Shiu. "Multi-scale patch based collaborative representation for face recognition with margin distribution optimization," in *12th European Conference on Computer Vision, Oct,* 2012, pp. 822-835.

[3] Guo, H., and Wang, W., "An active learning-based SVM multi-class classification model," *Pattern Recognition*, vol. 48, no. 5, pp. 1577-1597, 2015.

[4] S. Gu, L. Zhang, W. Zuo, and X Feng, "Projective dictionary pair learning for pattern classification," in *Advances in Neural Information Processing Systems*, 2014, pp. 793-801.

[5] C. Yuan, X. Li, W. Hu, H. Ling, and Stephen Maybank, "3D R transform on spatio-temporal interest points for action recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 724-730.

[6] Shlens, Jonathon. "A tutorial on principal component analysis," *arXiv preprint arXiv:1404.1100* ,2014.

[7] M. Liu, and Amit K. Roy-Chowdhury, "Multilinear feature extraction and classification of multi-focal images, with applications in nematode taxonomy," in *IEEE Conference Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 2823-2830.

[8] M. A. O. Vasilescu and D. Terzopoulos. "Multilinear subspace analysis of image ensembles," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 2003, pp. 93–99.

[9] J. Selinummi, P. Ruusuvuori, I. Podolsky, A. Ozinsky, E. Gold, O Yli-Harja, A Aderem, and I Shmulevich, "Bright field microscopy as an alternative to whole cell fluorescence in automated analysis of macrophage images," *PloS one*, vol.4, no. 10, e7497, 2009.

[10] D. Luo, H. Huang, and C. Ding, "Discriminative high order svd: Adaptive tensor subspace selection for image classification, clustering, and retrieval," in *IEEE International Conference on Computer Vision (ICCV)*, 2011, pp. 1443-1448.

[11] Klaser, Alexander, Marcin Marszałek, and Cordelia Schmid. "A spatio-temporal descriptor based on 3d-gradients," in *British Machine Vision Conference*, 2008, pp. 275-285.

[12] Scovanner, Paul, Saad Ali, and Mubarak Shah. "A 3-dimensional sift descriptor and its application to action recognition." in *Proceedings of the 15th ACM international conference on Multimedia*, 2007, pp. 357-360.

[13] J. Lu, G. Wang, and Pierre Moulin, "Localized multifeature metric learning for image set based face recognition," *IEEE Transactions on Circuits and Systems for Video Technology* vol. 26, no. 3, pp. 529-540, 2016.

[14] H. Liu, J. Lu, J. Feng, and J. Zhou, "Learning deep sharable and structural detectors for face alignment," *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 1666-1678, 2017.