A MULTI-DIRECTION IMAGE FUSION BASED APPROACH FOR CLASSIFICATION OF MULTI-FOCAL NEMATODE IMAGE STACKS

Min Liu [a], *Xueping Wang* [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, 1255 Amsterdam Ave, New York, NY 10027

ABSTRACT

In this paper, we present to use a multi-direction image fusion based feature extraction approach to classify multi-focal image stacks. The discrete wavelet transform sparse representation (DWTSR) image fusion technique is used to combine relevant information from a given image stack into a single image, which is more informative and complete than any single individual image within the given stack. Besides, multi-focal images within a multi-focal stack are fused along 3 orthogonal directions, and multiple features extracted from the fused images along different directions are combined by using canonical correlation analysis (CCA). The experimental results on the nematode multi-focal images show that our proposed multi-direction image fusion based feature extraction method can improve the recognition rate from 83.8% in the previous work to 96% by using texture feature only.

***Index Terms*—**Multi-focal images, Image fusion, Image classification, Canonical correlation analysis

1. INTRODUCTION

In biological or medical field, morphological information for a transparent specimen can be captured in form of a stack of high-quality multi-focal images, representing individual focal planes through the specimen's body [1]. A few multi-focal image stacks taken from a differential interference contrast microscope are shown in Fig. 1. 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 image stacks is still a problem. Most of the work on image feature extraction and classification are done through the 2D image processing methods [2, 3, 15, 16].

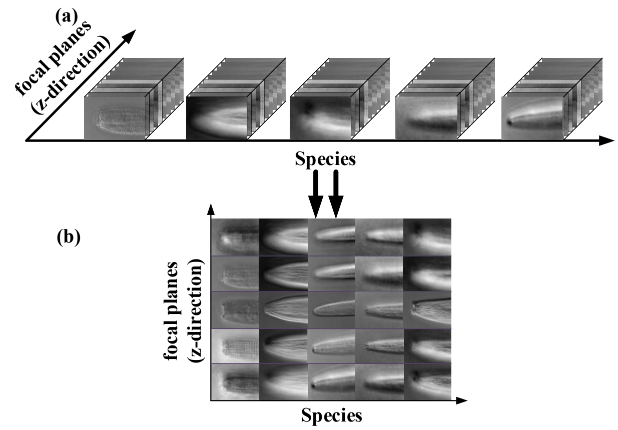

Fig. 1. A few samples of multi-focal image stacks taken from different nematode species (horizontal direction) are shown in (a) and (b). Each stack contains images of multiple focal planes taken from the top to the bottom of the specimen, with only a few frames of each shown (vertical direction).

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]. Previously, a projection based multilinear classifier is presented to classify the nematode multi-focal image stacks [6], where the 3D X-Ray Transform is used for the projection of the multi-focal image stacks. However, in the projection based multilinear classifier, the features extracted from the projection images along oblique directions are not very reliable for classification, especially when the precision along the *z*-direction is not high enough. It poses limitation to the projection based multilinear classifier.

In this paper, we propose to use a multi-direction image fusion based approach to classify multi-focal image stacks. On one hand, the image fusion methods can combine relevant information from multiple images of a given image stack into a single image, and the resultant fused image will be more informative and complete than any of the individual focal plane images. On the other hand, for a given image stack, we use canonical correlation analysis to combine the features extracted from the fused images along different directions.

We will show how to derive features for multi-focal

This work was supported in part by the National Natural Science Foundation of China under Grant 61301254, and the Hunan Provincial Natural Science Foundation of China under Grant 14JJ3069.

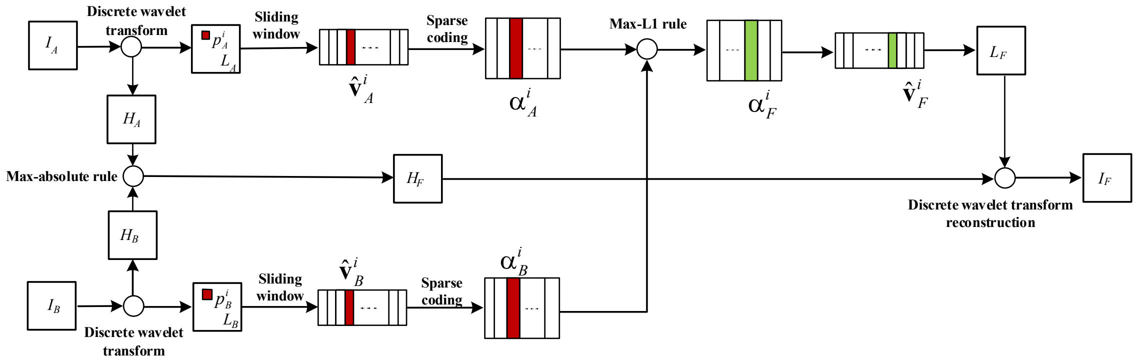

Fig. 2. The diagram for the DWTSR image fusion algorithm. I_A and I_B are the source images, L_A and L_B are the source low-pass bands, p_A^i and p_B^i are the source image patches, $\hat{\mathbf{v}}_A^i$ and $\hat{\mathbf{v}}_B^i$ are the source image vectors, α_A^i and α_B^i are the source sparse vectors, α_F^i is the fused sparse vector, $\hat{\mathbf{v}}_F^i$ is the fused image vector of low-pass band, L_F is the fused low-pass band, H_A and H_B are the source high-pass bands, and I_F is the fused image.

image stacks that take into account the texture information only, instead of the combination of shape feature and texture feature used in a previous work [6]. The proposed fusion based feature extraction method is applied in the classification for multi-focal image stacks of nematodes – a species which is very difficult to classify since they are one of the most numerous animals on earth [1]. The experimental results demonstrated that the fusion based analysis method could reach a higher classification rate (96.0%) than that by the previous projection based approach (83.8%) [6].

2. METHODOLOGY

2.1. Multi-scale Transform Sparse Representation (MSTSR) Image Fusion

In multi-focal image stacks, high-quality images usually contain complementary information which could be integrated. Through image fusion process extended or enhanced information content can be obtained in the composite image, which could be extremely useful for image classification application. In this paper, we employ a multi-scale transform sparse representation image fusion method to fuse the multi-focal images for image classification, which could take the complementary advantages of multi-scale transform (MST) and sparse representation (SR).

The MST-based image fusion methods assume that the underlying salient information of the source images can be extracted from the decomposed coefficients. Its basic idea is illustrated in Fig 2. Given two source images I_A and I_B, it consists of the following three steps in general [7]. First, I_A and I_B are decomposed into low-pass bands $\{L_A, L_B\}$ and high-pass MST bands $\{H_A, H_B\}$; second, the transformed coefficients $\{L_A, L_B, H_A, H_B\}$ are merged with a given fusion rule; finally, the fused image I_F is reconstructed by

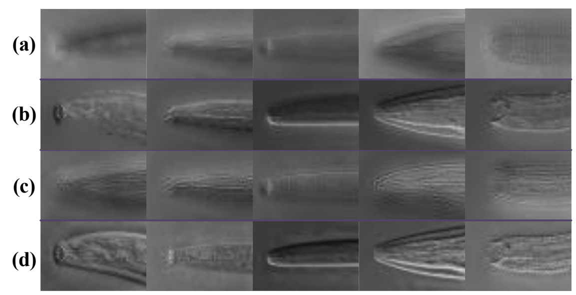

Fig. 3. The fused images of 5 multi-focal stacks along *z*-direction in Fig. 1, using different fusion methods, (a) DWT, (b) SR, (c) NSCT, (d) DWTSR.

performing the corresponding inverse transform over the merged coefficients.

For sparse representation based image fusion algorithm [8], the sparse coefficients are used as the local salient feature and the coefficients are merged with the "choose-max" fusion rule using a l_1- norm. The fused image is then reconstructed using the merged coefficients with a given dictionary $\mathbf{D}$.

Although both the MST- and SR-based image fusion methods have achieved great success, it is worthwhile to notice that both of them have some defects [7].

In fact, an image fusion framework by taking the complementary advantages of MST and SR is presented to overcome the related disadvantages [7]. Here we choose the discrete wavelet transform (DWT) for illustration. The diagram of the discrete wavelet transform sparse representation (DWTSR) image fusion method for two source images I_A and I_B is shown in Fig. 2.

Specifically, the high-pass DWT bands H_A and H_B are fused using the conventional "max absolute" rule shown in Equation (1),

$$H_F = \max\{H_A, H_B\} \tag{1}$$

H_F is the fused high-pass bands.

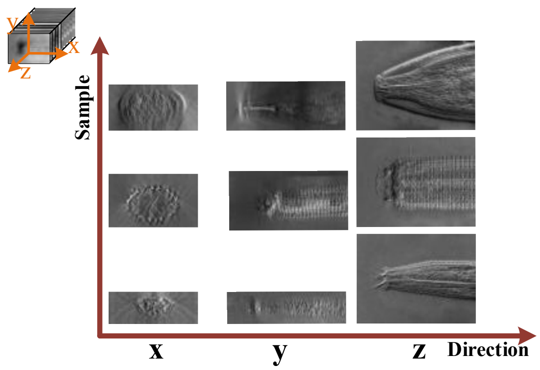

Fig. 4. The fused images of 3 nematode multi-focal stacks along *x, y* and *z* directions, using the DWTSR image fusion algorithm.

For the low-pass DWT bands L_A and L_B, they are first divided into patches p_A^i and p_B^i by a sliding window technique. Let us assume $\hat{\mathbf{v}}_A^i$ and $\hat{\mathbf{v}}_B^i$ are the normalized result of column vectors $\mathbf{v}_A^i$ and $\mathbf{v}_B^i$ from p_A^i and p_B^i, the sparse coefficient vectors α_A^i and α_B^i are then computed using the orthogonal matching pursuit (OMP) algorithm [7] as below,

$$\begin{cases} \alpha_A^i = \arg\min_{\alpha} \| \alpha \|_0 \quad s.t. \ \|\hat{\mathrm{v}}_A^i - \mathbf{D}\alpha \|_2 < \varepsilon \\ \alpha_B^i = \arg\min_{\alpha} \| \alpha \|_0 \quad s.t. \ \|\hat{\mathrm{v}}_B^i - \mathbf{D}\alpha \|_2 < \varepsilon \end{cases}, \tag{2}$$

Then α_A^i and α_B^i are merged with a SR-based fusion approach (max l_1- norm), as shown in Equation (3),

$$\alpha_F^i = \begin{cases} \alpha_A^i & if \quad \| \alpha_A^i \|_1 > \| \alpha_B^i \|_1 \\ \alpha_B^i & \text{otherwise} \end{cases}, \tag{3}$$

where α_F^i is the fused sparse vector.

The fused result of $\mathbf{v}_A^i$ and $\mathbf{v}_B^i$ is calculated by

$$\mathbf{v}_F^i = \mathbf{D}\alpha_F^i + \overline{v}_F^i \cdot \mathbf{1} \tag{4}$$

where the merged mean value $\overline{v}_F^i$ is obtained by

$$\overline{v}_F^i = \begin{cases} \overline{v}_A^i & if \ \alpha_F^i = \alpha_A^i \\ \overline{v}_B^i & \text{otherwise} \end{cases} \tag{5}$$

The low-pass fused result L_F is then generated by reshaping all $\mathbf{v}_F^i$ s and plugged into the original positions.

Once the fused high-pass bands and low-pass bands are calculated, the fused image I_F is finally obtained by performing the inverse DWT on the merged coefficients $\{L_F, H_F\}$.

The fusion results of the nematode multi-focal image stacks along the *z*-direction (as the vertical direction illustrated in Fig. 1) by some typical image fusion methods such as DWT, SR, nonsubsampled contourlet transform (NSCT) and DWTSR are shown in Fig.3. In fact, the image fusion can be done not only along the *z*-direction, but also along the other two orthogonal directions. The image fusion results of the nematode image stacks in Fig.1 by the DWTSR image fusion approach along *x, y* and *z* directions are shown in Fig. 4.

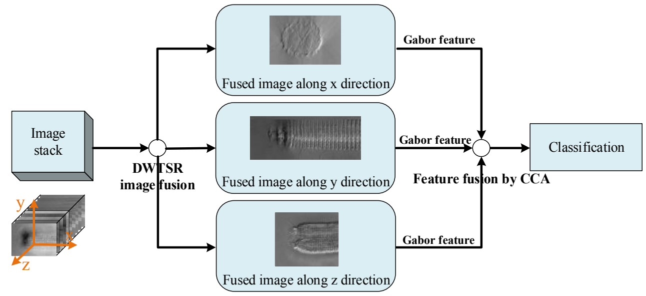

Fig. 5. The diagram of the proposed multi-direction image fusion based feature extraction and classification method.

2.2. Canonical Correlation Analysis

In our proposed classification framework, multi-focal images within a stack are fused along 3 orthogonal directions. Because different feature vectors extracted from the fused images represent different characteristics of multi-focal image stacks, we propose to use the canonical correlation analysis to optimize and combine these different features. The CCA analysis uses correlation feature between two groups of feature vectors as effective discriminant information. It is not only suitable for information fusion, but also eliminates the redundant information within the features [9].

Given a pair of mean-normalized feature vectors $X = [x_1, \dots, x_n]$ and $Y = [y_1, \dots, y_n]$, $x_i \in R^p$, $y_i \in R^q$, $i = 1, \dots, n$, we first establish the correlation criterion function between the two feature vector groups, and then extract their canonical correlation features to form effective discriminant vector for recognition. In general, CCA can be seen as the problem of finding basis vectors w_x and w_y for two sets of variables X and Y such that the correlations of projections of the variables onto corresponding basis vectors are maximized, while the projected variates within each data set are uncorrelated [9]. Mathematically, CCA can be described as

$$\left(w_x, w_y\right) = \arg\max_{w_x, w_y} \frac{w_x^T X Y^T w_y}{\sqrt{w_x^T X X^T w_x} \bullet \sqrt{w_y^T Y Y^T w_y}} \tag{6}$$

The vector pairs $\left(w_{x_i}, w_{y_i}\right)$, $i = 1, \dots, d$ $(d \le \min(p,q))$ can be obtained by solving Equation (6) in an optimization manner. The combined feature Z from X and Y is computed by a linear transformation, as below,

$$Z = \begin{pmatrix} w_x^T X \\ w_y^T Y \end{pmatrix} = \begin{pmatrix} w_x & 0 \\ 0 & w_y \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} X \\ Y \end{pmatrix} \tag{7}$$

2.3. Proposed Classification Method

The diagram of the multi-direction image fusion based classification method is shown in Fig. 5, including steps as below.

1. For each multi-focal image stack, compute the fused images I_x , I_y and I_z along three orthogonal directions using the DWTSR fusion algorithm.
2. For fusion output images along each direction, extract the corresponding Gabor texture features G_x , G_y and G_z .
3. Combine those Gabor features from all directions into one fused feature Z using CCA.
4. Find the class label for the testing multi-focal image stack through nearest neighbor criteria.

3. EXPERIMENTAL RESULTS AND DISCUSSION

In the experiment, we evaluated the proposed approach on a set of 500 nematode image stacks which are from 10 categories. Each category contains about 50 samples and each image stack consists of more than 100 multi-focal images with size 270×360 pixels obtained from a differential interference contrast microscope. Because the shape information is not easy to extract in most of the nematode images, we only use the 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 5.

Based on a cross-validation rule, in each experiment, half of samples for each nematode class are randomly chosen for training and the rest for testing. We repeat such experiment 100 times, and the final recognition rate is computed as an average of the results. The recognition rate comparison is made among the nearest neighbor classifiers with different features, as shown in Table 1.

The dictionary used in the DWTSR image fusion technique is the DCT dictionary used in [14]. The image fusion results of some nematode image stacks by the DWTSR image fusion method along *x, y* and *z* directions are shown in Fig. 4. We can clearly see that the fused images of the same image stack along different directions capture very different characteristics of the image stack. From Table 1, we can see that the classifiers with feature extracted from fused images perform better than with feature information directly extracted from the stacks such as 3D HOG (66.5%) and 3D SIFT (72.2%) [11, 12].

In order to demonstrate the strength of the image fusion algorithms, we compared the recognition rates of the nearest neighbor classifier with fused images and non-fusion images as the input. For the image fusion based classifier, both the training images and the testing images are the fused images along the *z*-direction using the DWTSR, DWT, SR and NSCT [7, 13] image fusion algorithm; while for the other non-fusion based classifiers, the training images and testing images are individual key frame images selected through a conditional entropy based selection method [10], or the projection images using the 3D X-Ray Transform [6].

From Table 1 we can see that the nearest neighbor classifier with fused images along the *z*-direction perform much better than with non-fusion images. For example, the recognition rate of the classification method with the Gabor feature extracted from the fused images along the *z*-direction is 94.5%, much higher than that from the 3D X-Ray Transform projection images. By comparing the performance of the classifier based on fused images using different fusion methods, we confirmed that DWTSR fusion method gets higher recognition rates than others in our experiment.

Table 1. Recognition rate of different classification methods

Classification Method	Recognition Rate
3D HOG	66.5%
3D SIFT	72.2%
3D X-Ray Transform + Gabor	83.8%
Key Frame + Gabor	86.9%
NSCT + Gabor (*z*-direction)	92.9%
SR + Gabor (*z*-direction)	93.2%
DWT + Gabor (*z*-direction)	93.1%
DWTSR + Gabor (*x*-direction)	80.3%
DWTSR + Gabor (*y*-direction)	83.7%
DWTSR + Gabor (*z*-direction)	94.5%
DWTSR + Gabor + CCA ($x + y + z$)	96.0%

Comparing the recognition rate from three different fusion directions, we can see that the fused images from *z* direction is more effective than *x* and *y* direction. The canonical correlation analysis is used to combine multiple Gabor texture features extracted from the fused images of image stacks along 3 orthogonal directions, and the proposed multi-direction image fusion approach using DWTSR with CCA analysis achieves the best recognition rate of 96.0%.

4. CONCLUSIONS

This paper presents to use a multi-direction image fusion based feature extraction approach to classify multi-focal image stacks. The discrete wavelet transform sparse representation image fusion technique is used to fuse the images within a multi-focal stack along 3 orthogonal directions, and then the features extracted from the fused images along different directions are combined by using canonical correlation analysis. The experimental results on the nematode data show that the image fusion based CCA analysis method can reach very reliable recognition rate.

5. ACKNOWLEDGEMENT

We gratefully acknowledge Prof. Paul De Ley from the College of Natural and Agricultural Sciences at the University of California, Riverside for providing us the datasets on which results are shown.

6. REFERENCES

[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, Dec. 2002.

[2] Gu, Shuhang, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. "Projective dictionary pair learning for pattern classification," in *Advances in Neural Information Processing Systems*, 2014, pp. 793-801.

[3] Akata, Zeynep, et al, "Good practice in large-scale learning for image classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 3, pp. 507-520, 2014.

[4] Holte, Michael B., Bhaskar Chakraborty, Jordi Gonzalez, and Thomas B. Moeslund, "A local 3-D motion descriptor for multi-view human action recognition from 4-D spatio-temporal interest points," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 5, pp.553-565, Sep. 2012.

[5] Chakraborty, Bhaskar, Michael B. Holte, Thomas B. Moeslund, Jordi Gonzalez, and F. Xavier Roca. "A selective spatio-temporal interest point detector for human action recognition in complex scenes," in *Computer Vision (ICCV), 2011 IEEE International Conference on*, Nov 6. 2011, pp. 1776-1783.

[6] Liu, Min, and Amit K. Roy-Chowdhury, "Multilinear feature extraction and classification of multi-focal images, with applications in nematode taxonomy," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on,* Jun 13. 2010, pp. 2823-2830.

[7] Liu, Yu, Shuping Liu, and Zengfu Wang. "A general framework for image fusion based on multi-scale transform and sparse representation," *Information Fusion, vol.* 24, pp. 147-164, 2015.

[8] Yang, Bin, and Shutao Li. "Multifocus image fusion and restoration with sparse representation," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 4, pp. 884-892, 2010.

[9] Sun, Quan-Sen, Sheng-Gen Zeng, Yan Liu, Pheng-Ann Heng, and De-Shen Xia. "A new method of feature fusion and its application in image recognition," *Pattern Recognition*, vol. 38, no. 12, pp. 2437-2448, 2005.

[10] Zhuang, Xiahai, Wenjia Bai, Jingjing Song, Songhua Zhan, Xiaohua Qian, Wenzhe Shi, Yanyun Lian, and Daniel Rueckert. "Multiatlas whole heart segmentation of CT data using conditional entropy for atlas ranking and selection," *Medical physics*, vol. 42, no. 7, pp. 3822-3833, 2015.

[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] Sahu, Deepak Kumar, and M. P. Parsai. "Different image fusion techniques–a critical review," *International Journal of Modern Engineering Research (IJMER)*, vol. 2, no. 5, pp. 4298-4301, 2012.

[14] Bo, Liefeng, Xiaofeng Ren, and Dieter Fox, "Hierarchical Matching Pursuit for Image Classification: Architecture and Fast Algorithms," in *Advances in Neural Information Processing Systems,* vol.1, no. 2, 2011.

[15] J. Lu, G. Wang, and P. 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.

[16] J. Lu, V. E. Liong, X. Zhou, and J. Zhou, "Learning compact binary face descriptor for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 10, pp. 2041-2256, 2015.